Deciphering Kunitz proteinase inhibitors in white clover (*Trifolium repens* L.): A transcriptional study

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

Doctor of Philosophy in Plant Molecular Biology

At Massey University, Palmerston North, New Zealand

Afsana Islam, 2013

Abstract

This thesis investigates the *Kunitz Proteinase Inhibitors (KPI)* gene family in white clover (*Trifolium repens* L.) as this family of inhibitors is one of the most abundant among the serine proteinase inhibitor families in legume species. In other studies, these proteins have mainly been shown to serve as storage proteins and to also act as potent defensive factors against insect herbivory. As they are involved in regulating proteolytic activity, the question arises as to how much they are also involved with regulating plant growth and development and how they respond to different stresses other than insect herbivory? Here, in this thesis effort has been under taken to answer these questions using the perennial legume white clover which is a major contributor to pasture productivity in New Zealand. However, as yet, very little is known about the occurrence of *Kunitz proteinase inhibitor (Tr-KPI)* genes or the functions of these genes in white clover. In this study, therefore, the spectrum of *Tr-KPI* genes is characterized, and the regulation of expression at the transcriptional level of different members of the gene family is examined.

To obtain *KPI* genes from white clover, degenerate primers were designed based on known legume *KPI* sequences. Four full length cDNA were obtained using degenerate and later gene-specific primers. Blast searching of the JCVI and NCBI database showed that they encode for proteins fall into the soybean trypsin inhibitor super-family (STI) and were named *Tr-KPI1*, *Tr-KPI2*, *Tr-KPI4* and *Tr-KPI5.* The expression in the transcript level of these four genes showed that *Tr-KPI1*, *Tr-KPI2, Tr-KPI5* are constitutively expressed in vegetative and reproductive parts whereas *Tr-KPI4* is more organ-specific such that it is expressed in the root and mature seed. A leaf and root developmental study showed that *Tr-KPI2* and *Tr-KPI5* are more developmentally regulated and transcript abundance during a germination time course study also suggests the involvement of *Tr-KPI1* and *Tr-KPI5* during seedling establishment.

To explore the function of these genes further, different forms of biotic and abiotic stress were applied to white clover. A mechanical wounding study revealed the possible involvement of *Tr-KPI1*, *Tr-KPI2* and *Tr-KPI5* in plant defense in both local and systemic tissues, and *Tr-KPI4* in the systemic tissue. A shoot herbivore (*Spodoptera litura*) and root herbivore (root knot nematode *Meloidogyne trifoliophila* and cyst nematode *Heterodera trifolii*) were also used to characterize the involvement of the *Tr-KPI* genes in plant defense response. Expression of the *Tr-KPI* genes against the generalist herbivore *S. litura* further supported the view that the *Tr-KPIs* in white clover are involved in plant defense responses where local (leaf), basipetal (root) and acropetal

(apical tissue) tissues were compared. The expression results suggest that *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* are induced by herbivore attack and *Tr-KPI1* was found to be most involved (1600-fold at 24 h) followed by *Tr-KPI2* and *Tr-KPI5*. In the nematode experiment, inoculation by a cyst nematode was able to trigger the expression of *Tr-KPI1, Tr-KPI4* and *Tr-KPI*5 in the root tissue at day 4 and a systemic response of nematode feeding was also observed in the leaf tissue for these genes at day 8. Invasion by the root-knot nematode did not result in any significant up-regulation for *Tr-KPI* genes at day 4 and day 8. This finding suggests that *Tr-KPIs* might be involved in defense against cyst nematode invasion but not by root-knot nematodes. To further elucidate the involvement of *Tr-KPI* genes under cyst nematode attack, a resistant line 17R and a susceptible line 23S were used. In the resistant line 17R, all four *Tr-KPI* genes were significantly expressed by day 4 and day 8, and in the susceptible line 23S, high transcript abundance was observed only at day 4. Therefore, it can be proposed that *Tr-KPIs* in white clover are important in defense against white clover cyst nematode in combination with other defense genes.

For an abiotic stress study, water deficiency and limited phosphorus (Pi) treatments were employed to examine the expression of the *Tr-KPI* genes in white clover. For the water deficiency trials, two treatments were imposed: a pre-stressed (PS) treatment in which plants were subjected to a water deficit for 7 days, followed by watering for a further 7 days before the experimental water deficit was applied, or a non-pre stressed (NPS) treatment in which plants were subjected immediately to a water deficit. The level of *Tr-NCED1* (9-cis-epoxycarotenoid dioxygenase) expression, coding for an enzyme involved in ABA biosynthesis, was also investigated to prove that a water deficit is perceived by the plants. The *Tr-NCED1* level was found to be up-regulated in the NPS treatment when compared with the level observed in fully hydrated tissue. Under the NPS and PS treatments, the transcription level of *Tr-KPI1* and *Tr-KPI5* were induced significantly in the leaf tissue when compared with the control. Interestingly, the pre-stressed treatment triggered the expression of all three genes studied which were significantly higher compared to the expression level under the NPS treatment. To further characterize the role of *Tr-KPIs* under water stress, a drought tolerant ecotype Tienshan and drought susceptible cultivar, Kopu was used. A clear upregulation of *Tr-KPI1* in Tienshan and *Tr-KPI5* in Kopu was observed under the PS treatment when compared with the initial moisture content and NPS treatments indicating some selective pressure on the *Tr-KPIs* under water stress in susceptible and resistant plants.

In a macro-nutrient (Pi) limitation experiment, where the growing root is divided in different developmental regions comprising the elongation zone (EZ), the visible lateral root zone (VL)

and the mature zone, a higher level of *Tr-KPIs* expression was observed in the growing zone rather than mature root zone. Although expression of all four *Tr-KPIs* was up-regulated in the EZ region, only *Tr-KPI2* and *Tr-KPI4* showed an extended level of expression in the visible lateral root zone indicating a possible involvement in lateral root formation as Pi limitation does induce a higher number of lateral root primordia. In leaf tissue, the down regulation of *Tr-KPIs* was observed up to 12 h of the Pi starved treatment and the transcript level started to increase from 24 h onward indicating that *Tr-KPIs* are not early response genes in leaf tissue.

Finally, the *cis*-binding elements in the promoter regions of four *Tr-KPI* genes indicate that this gene family in white clover is controlled by different transcription factors. A number of growth and development-related transcription factor binding sites such as AREF, ASRC, LFY, MADS and biotic and abiotic stress responsive transcription factors binding sites such as EINL, MYBL, MYBS, MYCL, and WNAC have been identified in all the four promoter sequences, although differences in the pattern and frequency were observed across the four *Tr-KPI* genes were observed. This further highlights that this gene family is regulated by a complex network of hormonal and other stress induced cues.

Acknowledgement

First of all, I would like to express my heartfelt respect to my supervisor Professor Michael T McManus for his continuous support, worthy guidance, valuable suggestions, incisive criticisms, encouragement and helpful comments during the research and writing period- I cannot thank you enough. I also thank my co-supervisor Paul Dijkwel for his encouragement, insightful comments, and helpful advice during the course of my study.

I am grateful for the support from Libby Burgess and her team from Plant and Food Research, Auckland during the herbivore experiment. I thank Chris Mercer, AgResearch Grasslands, Palmerston North, for helping me sorting out the nematode experiment. Tons of thanks go to Kim Richardson and his 'working bees' from AgResearch Grasslands, Palmerston North, for white clover transformation. Thanks to Steven Ray for his help in Plant Growth Unit, Massey University. Huge thank to Ann Turter, Cynthia Cresswell and James Connel who were always there to sort out any problems staying behind the scene.

Thank you my previous lab-mates: Aluh and Alvina for supporting me from the very beginning of my study. Thank you Susanna for your kindness and support and making my life easier in the lab. And Jay Jaya, thanks for your multitude help, I will always remember our 5.10 pm meeting. Thanks to you Srishti, will really miss the interesting topics about documentaries and movies. Thanks to the support team-Julia, Ameesha, Sashini, Uttara and Natalie for making the research easier. Thanks to my other lab-mates Mat, Diantha, Caleb, Rubina, Jibran, Faisal, Sam, and Jen.

I convey my special acknowledgement to Massey University for the Massey Doctoral Scholarship and Institute of Molecular BioSciences for funding the final year of my study and financial support for attending conferences.

Last but not the least; I thank my parents for always giving me mental support, my husband, brother and sister for inspiration and encouragement during the period of study.

Contents

Chapter 3 Results

Chapter 4 Discussion

List of Figures

List of Tables

Abbreviations

°C	Degree Celsius
µg	Microgram
µL	Microlitre
µmol	Micromole
A_{260}	Absorbance at 260 λ nm
A_{280}	Absorbance at 280 λ nm
BLAST	Basic Logical Alignment Search Tool
bp	Base-pair
ca.	*circa* (approximately)
cDNA	DNA complementary to a RNA, synthesized from RNA by the reverse transcription *in vitro*
DEPC	Diethyl pyrocarbonate
DMF	*N,N*-dimethyl formamide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxynucleotide 5'triphosphate
DTT	Dithiothreitol
E.coli	*Eschercia coli*
EDTA	Ethylenediaminettetra acetic acid
EST	Expressed sequence tag
g	Acceleration due to gravity (9.8m s^{-2})
g	Gram
h	hour
IPTG	Isopropyl-β-$_D$-thiogalactopyranoside ($C_9H_{18}O_5S$)
Kb	Kilo base-pairs
kDa	Kilo Daltons
L	Litre
LB	Lauria-Bertani (media or broth)
M	Molar (moles per litre)
mg	Milligram
Milli-Q water	Water purified by a Milli-Q ion exchange column
min	Minute
mL	Millilitre
mol	mole (amout of a substrate, Avogadro's number)
MPa	Mega Pascal
NBT	*p*-nitro blue tetrazolium chloride
NCBI	National Centre for Biotechnology Information
ng	Nanogram
nmol	Nanomole
NPS	Non pre-stressed
ORF	Open reading frame

PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline (50 mM sodium phosphate, pH 7.4 containing 250 mM NaCl)
PCR	Polymerase chain reaction
pH	-Log (H^+)
pmol	Picomole
PS	Pre-stressed
PVP-40	Polyvinyl pyrrolidone
qRT-PCR	Quantitative real-time PCR
RH	Relative Humidity
RNA	Ribunucleic acid
RNase	Ribonuclease
RO	reverse osmosis
RT-PCR	Reverse Transcriptase-polymerase chain reaction
sec	second
SEM	Standard error mean
T_m	Melting temperature at which DNA strands separate in preparation for annealing
Tris	Tris (hydroxymethyl) aminomethane
Tween-20	Polyoxyethylenesorbitan monolaurate
U	Unit (commercial enzymes are in U μL^{-1}, where unit is based on enzyme activity)
UTR	Untranslated region
UV	Ultra violet
V	Volt
v/v	Volume per volume
w/v	Weight per volume
w/w	Weight per weight
X-Gal	5-Bromo-4-chloro-3-indolyl-β-$_D$-galactopyranoside

Chapter 1

Introduction

1.1 Peptidases, proteases, proteinases and proteinase inhibitors - A note on nomenclature

The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB, 1992) defines the term peptidase as all enzymes that hydrolyze peptide bonds (Fan and Wu, 2005). These are subdivided into two categories: (i) exopeptidases, which cleave at the N-terminal (aminopeptidases) or the C-terminal (carboxypeptidases), and (ii) endopeptidases, which cleave the internal peptide bonds of the polypeptide chain. According to Barret (1986), the term protease encompasses both exo- and endopeptidases while the term proteinase describes only the endopeptidases. Thus proteinase inhibitors are the proteins that interfere with the ability of proteinases to hydrolyse peptide bonds.

1.2 Proteinase Inhibitors (PI) -The natural antagonists of proteinases

The extra- and intracellular regulation of protein processing and turnover in any living organism is a constant process controlled by specialized enzymes, the peptidases through their interactions with their respective protein inhibitors (Koiwa *et al*., 1997). Thus the proteinase inhibitors also control proteolysis by their respective proteinases within the cell and organelles for proper physiological and biochemical functioning. Proteinase inhibitors interact with their target proteinases by contact with the active (catalytic) site, resulting in the formation of a stable proteinase inhibitor-proteinase complex such that the proteinase is incapable of further enzymatic activity (Norton, 1991).

The existence of proteinase inhibitors in nature has been known for over a century. At the end of eighteenth century, Fermi and Pernossi (1894) first reported that normal human blood serum has the ability to hinder the action of trypsin, although these researches did not understand the mechanism of inhibition (Birk, 2003). After saturation of blood serum collected from different animals with salt (ammonium sulfate), Cathcart (1904) found that the substance responsible for the action of normal serum against trypsin is present in the albumin fraction. Later, Duthie and Lorenz (1949) reported the presence of the inhibitor both in the albumin and globulin fractions after salt precipitation. However, the commencement of prolific biochemical studies with these proteinase inhibitory substances is marked by the pioneering work of Moses Kunitz. Kunitz, in 1936, first isolated and crystallized trypsinogen from beef pancreas and later isolated and crystallized trypsin proteinase inhibitors from soybean (Kunitz, 1946). Since then, researchers in various fields including both animal and plant biology have focused on the study

of these inhibitor proteins as a means of revealing the mechanism of inhibition of proteolytic enzymes and protein-protein associations. Today it is known that these proteins are mostly low molecular weight proteins which occur in all life forms (Fritz, 2000). In mammals, for example, they constitute a significant part of the proteins present in blood serum (about 20 of nearly 200 proteins) and are also found in invertebrates such as Gorgonidae (*Cnidaria, Gorgonuceu;* also called sea fans) and horny corals and in fungi and bacteria (Boigegrain, *et al.,* 1994).

Proteinase inhibitors have also attracted the attention of nutritionists due to their presence in valuable plant foods and their possible involvement as anti-nutritive factors. However, findings of the involvement of plant PIs in the prevention of tumorigenesis may contribute to the nutritional utilization of valuable plant protein sources such as legume seeds. The inhibitors are also being used as valuable tools in medical research because of the unique pharmacological properties that suggest clinical applications. For example, in medical science, HIV protease inhibitors are used to treat HIV patient where PIs interfere with the activity of viral proteases and thereby hinder continuous infection and lower the viral load in AIDS patients (Flexner, 1998).

1.3 The protein 'proteinase inhibitor' in plants

Proteinase inhibitors are distributed throughout the plant kingdom from angiosperms to the chlorophyta, e.g. green algae (Roberts and Heigaard, 2008). As indicated, the study of proteinase inhibitors in plants arose from the earlier work on the isolation and crystallization of a trypsin (serine proteinase) inhibitor from soybean seeds by Kunitz (Kunitz, 1946). As they were first identified in legume seeds, the study of different legume crops occupied the attention of researchers over these early years. However, the endogenous functions of the protein proteinase inhibitors have been questioned for a long time (Ryan, 1989). For decades, different groups of inhibitors were thought to be harmful substances because of their detrimental effect on growth and development of animals that were fed seed or raw seed products (Xavier-Filho, 1992). However, the subsequent detection of these proteins in members of the Graminae, Solanaceae, Brassicaceae and Cucurbitaceae families suggested that these proteins might have a wide range of endo- and exogenous functions in plants.

Plant proteinases in common with all proteinases are classified according to the presence of specific amino acid residues at their active sites and their mechanism of action. According to

IUBMB, there are four major classes of proteinases: the serine proteinases, the cysteine proteinases, the aspartic proteinases and the metallo-proteinases. Proteinase inhibitors are classified into nonspecific and class-specific super-families and the latter is subcategorized into several families according to the type of proteinase they inhibit including serine-, aspartic-, metallo-, and cysteine-proteinase inhibitors (Hibbetts *et al.,* 1999). However, there are examples where a specific class of inhibitor can also interact with other classes of proteinases. For example, *in vitro* studies with a serpin proteinase inhibitor from *Arabidopsis thaliana*, AtSerpin1 (At1g47710) which falls into the serine inhibitor group was found to interact with endogenous metacaspase 9 (AtMC9) which is a cysteine protease (Vercammen *et al.*, 2006). According to the MEROPS database (http://merops.sanger.ac.uk/cgi-bin/family_index?type=I) a family of proteinase inhibitors includes all inhibitors that have significant similarities in amino acid sequence. In the database, when considering both the plant and animal kingdoms, there are 74 inhibitor families from I1 to I91, falling into 73 inhibitor clans listed, depending on the evidence of evolutionary relationships due to similar tertiary structure. However, the PIs from angiosperms have been assigned into 11 families, falling into 9 clans in the database (See Appendix 1).

Plant proteinase inhibitors constitute proteins with a molecular mass range of typically 8 to 20 kDa, although proteins of 85 kDa have been reported (Fan and Wu, 2005). Their general characteristics involve the presence of a number of cysteine residues, tolerance to extreme pH conditions, and resistance to heat and proteolysis (Fan and Wu, 2005). Current knowledge of the interaction mechanism and inhibitory activity of PIs with endogenous proteinases is mainly based on the interaction studied *in vitro* with commercially available proteases such as trypsin, chymotrypsin, elastase, and subtilisin from animal and microbial sources (Brzin and Kidric, 1995). However, these commercial enzymes may not represent the endogenous target proteinases for most of the characterized plant PIs (Fan and Wu, 2005).

1.4 Plant serine proteinase inhibitors

The serine proteinase inhibitors are by far the largest group of PIs in nature and are probably the most studied group of the plant proteinase inhibitors. They have been extensively characterized in different plant species belonging to different families such as the Leguminoceae, Solanaceae, Graminae, Cucurbitaceae, etc. The mammalian digestive serine proteinase, trypsin, was the enzyme first utilized by early investigators to search for proteinase inhibitors in plants (Liener and Kakade, 1980) and a great number of serine proteinase

inhibitors were discovered (Xavier-Filho and Campos, 1989). These inhibitors were identified first in storage organs including seeds and tubers, which are consumed by man and animals as a protein source. Later, the inhibitor proteins were found in non-reserve plant organs such as leaves and flowers (Xavier-Filho and Campos, 1989) suggesting that their presence is not limited to storage organs only.

The proteins of the serine PI family are generally smaller proteins ranging from 29 to 190 amino acids (Ryan, 1990). These proteins from plants have been grouped into families according to their molecular size and primary structures in such a way that inhibitors belonging to the same family are thought to be evolutionarily related (Laskowski and Kato, 1980). The first criteria of this classification are molecular mass, followed by the number of disulfide bridges. Most proteins of this family possess two disulfide bonds, Cys63-Cys110 and Cys160-Cys169. However exceptions occur, such as in the plant *Swartzia,* where the serines PIs have only the first of the two bridges (do Socorro *et al*., 2002). Using these criteria, eight families of serine proteinase inhibitors have been identified currently: the Bowman-Birk, Kunitz, Potato I, Potato II, Cucurbit, Cereal super-family, Ragi AI, and Thaumatin-PR like families (Richardson, 1991). However, the number of known and partially characterized inhibitors of serine proteinases is enormous (Haq *et al*., 2004) as they occur universally in the plant kingdom and have been reported in many different plant species (Mello *et al*., 2002; Haq and Khan, 2003). Seeds from many plant species from several families assayed for the presence of PIs have been found to contain serine proteinase inhibitors (Xavier-Filho and Campos, 1989; Richardson, 1991) supporting their occurrence in seeds.

1.4.1 Localization of plant serine proteinase inhibitors

In terms of tissue localization within plants, Xu *et al*., (2001) used *in situ* hybridization to locate the accumulation of *SaPIN2a* mRNA encoding the proteinase inhibitor II (PIN2) protein from *Solanum americanum* in the companion cells and sieve elements of phloem tissues examined from stems, roots, and leaves. SaPIN2a-specific antibodies used for western blot analysis also supported the protein accumulation in stems, leaf midribs and fruits. Furthermore, an immunohistochemical localization assay revealed that SaPIN2a accumulated in the phloem of the stem. Immunoelectron microscopy of stem, root and leaf sections further characterized the localization of SaPIN2a to the companion cell and sieve elements, particularly the parietal cytoplasm adjacent to the cell wall, the lumen and the sieve-area pores. Xu *et al*. concluded from these findings that other than a possible role in plant defense,

SaPIN2a could also be involved in regulating proteolysis in the sieve elements. Further investigation of the SaPIN2 PIs conducted by Sin and Chye (2004) showed that SaPIN2a and SaPIN2b also accumulate in young floral buds and mature floral tissues as determined by western blot analysis. Immunolocalization of SaPIN2a and SaPIN2b in flowers was carried out using affinity-purified SaPIN2a- and SaPIN2b-specific antibodies which showed their presence in the stylar transmitting tissue, vascular bundles, nucellar cells of the ovule, and the outermost cell layer of the placenta in floral buds.

Apart from these detailed studies with SaPIN2a and SaPIN2b, the exact sub-cellular location of many of the PIs is still unknown (Richardson, 1977; Norton, 1991), although, the intercellular localization of proteinase inhibitors has been studied only in a few plants. The mung bean trypsin inhibitor was found to be localized in the cytosol of cotyledonary cells (Chrispeels and Baumgartner, 1978) which was identified by fractionation of a crude extract using isopycnic sucrose gradients and by immuno-cytochemical localization with rhodamine-labeled IgG antibodies. The soybean trypsin inhibitor (SBTI) was found to be localized in the cell walls, cytoplasm, and the nuclei of cotyledonary and embryonic cells (Horisberger and Tacchini-Vonlanthen, 1983a). The two soybean Bowman-Birk inhibitors (SBBIs) were found to be localized at the ultra-structural level to the nucleus, cytoplasm and intercellular spaces using immunofluorescence and by protein A gold methods on thin sections of the soybean cotyledon and embryonic axis. In the cotyledon and embryonic axis, SBBI was found in all protein bodies, the nucleus and, to a lesser extent, the cytoplasm (Horisberger and Tacchini-Vonlanthen, 1983b).

1.4.2 Mechanism of inhibition

Serine proteinase inhibitors associate with their target enzymes according to what is known as the ‘standard mechanism’ or the 'Laskowski mechanism' (Laskowski and Kato, 1980). The inhibitor consists of a reactive site domain that forms a loop which fits into the active site of the enzyme (Fig 1). The loop is highly flexible but becomes rigid in the complexed state. The length of the reactive site domain varies depending on the size of the inhibitor. For example, in a class of serine PIs from soybean, the Bowman-Birk inhibitor (BBI), the site is nine residues long (Qi *et al.*, 2005). It is thirteen amino acids in the potato type II serine proteinase inhibitor (Schirra *et al.,* 2009) and eighteen in the cucurbit (*Cucurbita maxima*) serpin (Peterson *et al.,* 2005).

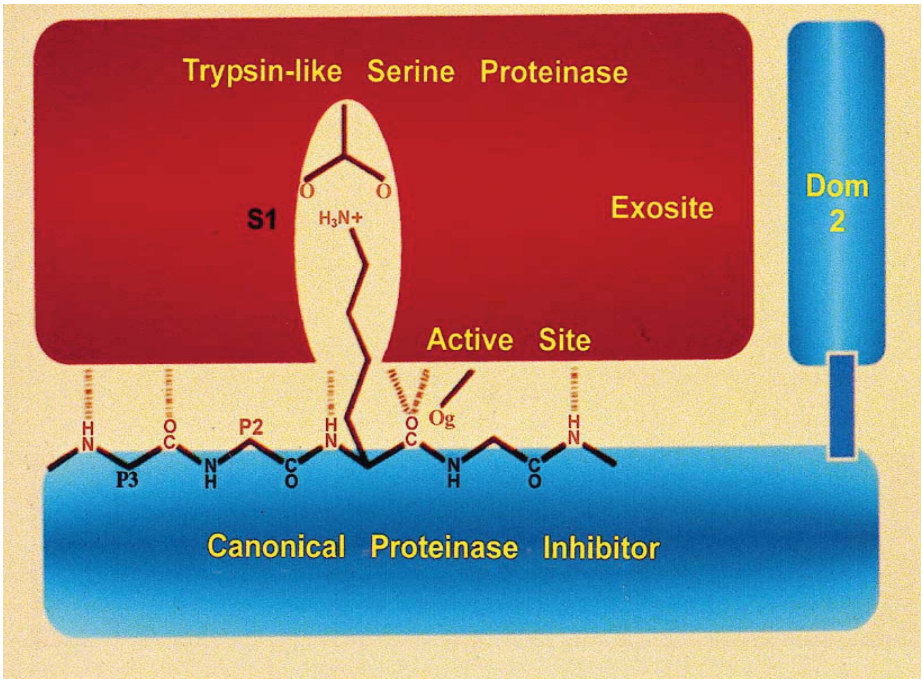


Figure 1.1: Schematic representations of interactions between a proteinase and protein proteinase inhibitor. The possible mechanism of the serine proteinase inhibitor by directly blocking the active site of the proteinase in a substrate-like manner is shown. The inhibitor provides a reactive site loop forming several inter-main chain hydrogen bonds with binding sites of the proteinase, with its scissile peptide bond partially added to the active site, and with the P1 side chain (here that of a Lys residue) as the principal cleavage site determinant projecting into the S1 specificity pocket, making numerous substrate-like interactions. The proteinase is shaded red and the proteinase inhibitor as blue. The superimposed peptide chains (bold chain) represent productively bound peptide substrates. Figure is adopted from Bode and Huber, 2000.

1.5 Plant Kunitz proteinase inhibitors - The diversified inhibitor family

The Kunitz Proteinase inhibitor (KPI) family belongs to the serine proteinase inhibitor group. Numerous Kunitz-type inhibitors have been isolated and characterized from different plants. However, specific information about the gene family in these different plant species is quite rare. Currently, about 230 homologues are listed in the MEROPS database arising from 94 plant species. The Kunitz family of proteinase inhibitors is usually a multi-gene family, although, the occurrence of single members is not uncommon, as observed, for example in *Oryza sativa*. KPI as a widespread family was first described in legumes (Laskowski and Kato, 1980), but the group is also found in a wide range of species falling in the Brassicaceae, Graminae, and Solanaceae.

KPIs are found to be abundant in legume seeds such as soybean, chickpea, cowpea and in the sandalwood tree (*Adenanthera pavonina*) (http://merops.sanger.ac.uk/inhibitors/). Among the legumes, soybean seeds have been thoroughly studied in term of the occurrence of the Kunitz trypsin inhibitor (KTIs) (Laskowski and Kato, 1980; Ryan, 1989). The most well known are the soybean Kunitz trypsin inhibitors (SKTI) which is a seed specific protein that accumulates in high amounts during development (Vodkin, 1981). These proteins are the most prevalent soybean proteinase inhibitor and so far at least 11 Kunitz trypsin inhibitor genes have been identified in the soybean genome (www.merops.sanger.ac.uk). However, it is noted that there are some soybean cultivars whose seeds do not contain KPIs or the protein does not accumulate during development (Orf and Hymowitz, 1979; Jofuku *et al.*, 1989). In addition to seeds, there are also examples of the occurrence of kunitz proteinase inhibitors in vegetative organs such as leaves which have been identified from members of the legume family such as soybean, alfalfa and ladino clover (Kendall, 1951).

Among solanaceous plant species, potato has a complex set of Kunitz-type enzyme inhibitors (Stiekema *et al.,* 1988; Ishikawa *et al.,* 1994) accumulating during tuber development. A protein extraction from vacuoles of potato tubers (cv. Kuras) was shown to contain about 28 distinct Kunitz proteinase inhibitor proteins (Jorgensen *et al.*, 2011). In the potato cv. Provita genome, at least 21 Kunitz proteinase inhibitor genes have been identified by Heibges and co-workers (2003a). The results of that study concluded that the structural and functional diversity observed among members of this complex family varies markedly between potato cultivars. The members show surprising sequence variation, including many non-synonymous substitutions and indels, which appear to translate into functional diversity. It has been hypothesized that wild type potato species that are genealogically distant from cultivated potato harbour KPI isoforms with distinct antiprotease activities and that each potato cultivar harbours several nonallelic genomic loci bearing KPI genes (Speransky *et al.,* 2007).

In 2008, Major and Constabel investigated the functional and biochemical variability of the KTI genes of *Populus trichocarpa* X *Populus deltoides,* a member of the Salicaceae family. Through this study they confirmed that these genes also belong to a large and diverse family with complex expression patterns in *Populus*. Five KTI proteins which were induced by artificial wounding and herbivore damage, represented the diversity of KTI gene family by showing biochemically distinct properties and clear differences in efficacy against trypsin-type, chymotrypsin-type, and elastase-type proteases, suggesting functional specialization of different members of this gene family (Major and Constabel, 2008). The stability of the KTIs *in*

vitro in the presence of reducing agents and elevated temperature also varied widely, further emphasizing the biochemical differences of these proteins (Major and Constabel, 2008).

Although the entire gene family has not been characterized in the model plant *A. thaliana*, there are six full length Kunitz proteinase inhibitor genes in the *Arabidopsis* genome. In the legume model plant *Medicago truncatula*, 13 full length Kunitz proteinase inhibitors genes have been annotated (http://www.jcvi.org/cgi-bin/gb2/gbrowse/mtruncatula/).

1.6 Kunitz Proteinase Inhibitor (KPI)–Are they one unit many functions family?

The Kunitz-type proteinase inhibitors are proteins of approximately 18 to 22 KD with a single reactive site, and the interaction of KPIs with their cognate proteinases is well characterized. Most KPI proteins have four conserved Cys residues that form two disulfide bridges, although examples with only one or no disulfide bridges are also known (do Socorro *et al.,* 2002; Araujo *et al.,* 2005; Macedo *et al*., 2007, Major and Constabel, 2008). Research with KPIs suggests that the first disulfide bridge, which surrounds the reactive loop, is necessary for the inhibitory activity of most KPIs (DiBella and Liener, 1969; do Socorro *et al.,* 2002, Major and Constabel, 2008). KPIs inhibit trypsin or chymotrypsin as well as other serine proteinases such as elastase and subtilisin (Terada *et al.,* 1994; Valueva *et al*., 2000) suggesting that, as a group, they have extremely diverse serine proteinase targets. In addition, potato tuber paralogues have been shown to inhibit the aspartic proteinase cathepsin D, and the cysteine proteinase papain (Ritonja *et al.,* 1990; Heibges *et al.,* 2003b). Moreover, it was reported by McCoy and Kortt (1997) that some proteins of the KPI family do not act as inhibitors, but instead some have lectin like carbohydrate-binding or invertase inhibitory activity (Glaczinski *et al.,* 2002). Some of the plant KPIs also has antimicrobial activity, presumably via inhibition of microbial proteinases (Kim *et al.,* 2005; Park *et al.,* 2005). Therefore, considering the extensive range of inhibitory activity against proteinases as observed for different plant species, the KPIs may have a wide range of functions.

1.6.1 Kunitz Proteinase inhibitors as storage proteins

In the legumes, the amount of seed storage protein can be around 40% while in cereals it is, more typically, up to 10% of dry weight (Shewry *et al.*, 1995). Kunitz proteinase inhibitors were originally proposed to function as storage proteins, as first suggested by Pusztai (1972), on the basis of their occurrence in storage tissues. They are generally considered to be a sink for nitrogen, sulfur and may merely serve as reserve proteins that are mobilized during

germination and sprouting (Norton, 1991). The highest amount of KPI protein has been found in sandalwood constituting about 20% of the soluble protein present in the seeds (Richardson *et al.*, 1986). Shee and Sharma (2008) demonstrated that a trypsin inhibitor in *Murraya koenigii*, from the family Rutaceae, is a major seed storage protein by monitoring the relative concentrations during seed development and germination. These authors found the quantity of protein to be approximately 20% of total protein extracted by simple buffer extraction and the concentration during different stages of seed development was found to be 5.27, 5.5, 8.5, 18.8 and 20% in 7, 19, 25, 37 and 55 days after anthesis, respectively. During seed germination, protein degradation was observed to decline from 20% to 12%, 7% and then 2% at 13, 16 and 22 days post imbibition, respectively. Therefore, it is quite reasonable to assume that KPIs are performing a dual role in dormant seeds serving both as a reserve protein and as inhibitors of any developmentally unscheduled proteinase activity.

As a storage organ, potato contains a large number of proteinase inhibitors classified into different families. It has been reported that after the glycoprotein patatin, which constitutes up to 40% of soluble protein (Bohac 1991; Strickland *et al.,* 1995), Kunitz proteinase inhibitors are the second major group of protein in the tuber (Bauw *et al.,* 2006). Two major proteins, Globulin2a and Globulin2b, which account for about 40% of the total soluble protein in developing taro, belong to the Kunitz trypsin inhibitor family (de Castro *et al*., 1992 and Monte-Neshich *et al*., 1995).

1.6.2 Kunitz proteinase Inhibitors as regulators of proteinases during germination

Kunitz proteinase inhibitors also play an important function in regulating controlled endogenous proteinase activity before and during seed germination as part of storage protein digestion and to control protein turnover. The concentration of inhibitors is reduced during germination, facilitating the hydrolysis of protein for utilization in the germination process. Shain and Mayer (1965) were the first to show that a trypsin inhibitor could suppress its own trypsin-like proteinase in germinating lettuce seeds. Later Tan-Wilson *et al.,* (1985) observed the presence of Kunitz soybean trypsin inhibitors (KSTI) in the protein body, cytosol and cell wall of cotyledon, hypocotyls, root and epicotyl in the germinating Amsoy 71 soybean seedlings. They noted that the level of KSTI declined as the germination process progressed and they concluded that this decline is because of the proteolysis of these storage proteins as a means to provide essential amino acids to the growing embryonic axis.

In view of the transcriptional level of *KPIs* during the seed germination process, Hernandez-Nistal and coworkers (2009) examined two Kunitz trypsin inhibitors: *CaTPI-1* and *CaTPI-2* from chickpea (*Cicer arietinum* L.) during germination. In the study they found that the transcript level of *CaTPI-1* remained high up to 24 h after imbibition, while *CaTPI-2* mRNA appeared later and reached a maximum at 48 h of post imbibition. From this observation, the authors concluded that *CaTPI-1* is involved in protective roles in imbibed seeds to prevent the premature degradation of the proteins stored in the embryonic axes, while Ca*TPI-2* is mainly active following germination especially during elongation of the embryonic axes. Thus both of these transcriptional and translational level studies indicate the involvement of KPIs during the germination process.

1.6.3 Defense against insect herbivores as endogenous insecticides

The defense mechanisms of plants against abiotic and biotic stress can be classified as constitutive or induced. The constitutive mechanisms comprise the existing barriers including, for example, a thick cuticle or the density of trichomes. The induced defense mechanism includes the biosynthesis of different secondary metabolites and proteins that can act as toxins, antifeedents and antinutrients. Among these, one of the most common inducible proteins against herbivore defenses in plants is the proteinase inhibitors. Probably the first study on the effect of proteinase inhibitors on the development of insects was reported by Lipke and coworkers (Lipke *et al.,* 1954), who observed that the larvae of flour beetle (*Tribolium confusum*) failed to develop when fed soybean flour that contains a high proportion of Kunitz trypsin inhibitor. Later, Green and Ryan (1972) provided more evidence that proteinase inhibitors are a part of the natural defensive chemicals in plants. They demonstrated that wounding of tomato and potato leaves by Colorado potato beetles induced rapid accumulation of the proteinase inhibitor I (PIN1) in damaged leaves and also in distal undamaged leaves. Later, Ryan (1990) reported that these proteinase inhibitors are active against insect digestive proteinases which results in reduced insect growth rates or increased mortality as a consequence of either a critical shortage of essential amino acids (Hilder *et al*., 1993; Jongsma and Bolter, 1997) or the hampering of important physiological processes such as moulting (Gutierrez-Campos *et al.,* 1999).

The use of transgenic plants provided possibilities to get more direct evidence for the involvement of proteinase inhibitors in plant defense. Transformation of the cowpea trypsin inhibitor (TI) gene into tobacco to control the generalist herbivore tobacco budworm (*Heliothis*

virescens), was the first example of inhibitors being used for insect pest control (Hilder *et al*., 1987). Johnson and coworkers (1989) showed that the growth of the specialist herbivore *Manduca sexta* larvae, feeding on the leaves of transgenic tobacco containing PI-2, a powerful inhibitor of trypsin and chymotrypsin, was significantly retarded when compared with the growth of these larvae when fed untransformed (control) leaves. In addition to these examples from the Leguminosae and Solanaceae, KPIs are also found to be effective against specialist herbivores specific for different plant species from different families. An example is the over-expression of Kunitz proteinase inhibitor (SKTI) from soybean in rice. Though SKTI acts as a major seed storage protein, in transgenic rice it increased resistance against the brown plant hopper (*Nilaparvata lugens*) when compared with control plants (Lee *et al.*, 1999). Another example is KPIs from *Inga laurina* seeds from the family Leguminosae reduced the growth of coconut palm weevil (*Homalinotus coriaceus*), which is a herbivore of the Palmae family, when incorporate into artificial diet (Macedo *et al.*, 2011).

KPIs have also been showed to act in defense responses against pathogens or parasitic infection using a different approach. Lima and coworkers (2011) demonstrated that this group of serine proteinase inhibitors from corms of *Xanthosoma blandum* (Xb-KTI) has antibacterial activity against *Staphylocccus aureus, Salmonella typhimurium,* and *E. coli* when added directly to the cultures. The antimicrobial screening assay showed 15%, 25%, and 12% inhibition of the bacteria respectively, at a concentration of 100µg/mL of Xb-KTI. Examples of the involvement of KPIs under plant parasitic nematode attack are also available. In more correlative evidence at the transcript level, five out of six transcripts encoding Kunitz trypsin inhibitor genes in soybean were found to be highly up-regulated by soybean cyst nematode (*Heterodera glycines*) (SCN TN8) feeding in soybean cv. Peking root tissue (Rashed *et al.*, 2008) indicating their involvement in defense against nematodes.

Taken together, all of these research approaches together suggest that native or transgenic plants expressing KPI genes have great potential to produce germplasm with enhanced tolerance to pathogens such as nematodes, fungi, bacteria, and viruses as the survival and/or invasion of these organisms requires proteolytic activities.

1.6.4 Kunitz proteinase inhibitors during nodulation

One of the main features of leguminous plants is the ability to fix atmospheric nitrogen by symbioses with the nitrogen fixing bacteria *Rhizobia*. Manen and co-workers (1991) observed

the accumulation of a Kunitz proteinase inhibitor in the senescing nodules of winged bean. An immunocytochemical study further determined that the protein is restricted to the disorganized bacteroids, vacuole membrane and fluid of the infected nodule cell. At the transcriptional level, Lievens *et al.,* (2004) determined that a Kunitz proteinase inhibitor (*SrPI1*) gene of *Sesbania rostrata* is expressed during the early stage of nodule formation in stems. Expression of the gene was not regulated by wounding or by a wide range of pathogens including *Botrytis cinerea* and *Ralstonia solanacearum.* Since the expression of the gene was induced only during the early stages of nodulation, the authors concluded that the gene can act as a nodulation marker and signifies the importance of this particular KPI in guarding against the escape of bacteria which might have pathogenic effects, thus isolating the infected nodule tissue from rest of the plant. Microarray data from the model plant *M. truncatula* provides a clue as to the importance of KPIs during nodulation with differential expression of 10 out of 12 annotated *KPIs* expression in young and mature nodules (Figure 4.1). Though detailed study of the KPIs during nodulation is rare, it can be speculated that rhizobial infection can trigger the expression of members of the KPI family as a part of the plant defense mechanism.

1.6.5 Kunitz proteinase inhibitors as antagonists in programmed cell death (PCD)

Active and controlled cell suicide, essential for development, homeostasis and defense against pathogens in multi-cellular organisms, is known as programmed cell death (PCD) or apoptosis, and involves controlled induction and activation of cellular mechanisms governed by regulted genetic programmes (Li *et al.,* 2008). As reported by Earnshaw (1995), a cysteine proteinase of the Ced-3/ICE family is involved in apoptic cell death in *Caenorhabditis elegans,* but the process can be inhibited by the expression of the serine proteinase inhibitor, CrmA. In plants, proteinase inhibitors have also been shown to play crucial roles in the cellular regulation of proteases during the PCD process by controlling cysteinyl Asp-specific proteases (caspases) and serine proteases (Shi, 2002; Woltering *et al.,* 2002).

Specific inhibitors of PCD in plant cells include cysteine (Solomon *et al*., 1999) and serine (Karrer *et al*., 1998; Park *et al.*, 2001) proteinase inhibitors. It was found in soybean cells by Solomon *et al.,* (1999) that programmed cell death, which initiates during oxidative stress and under pathogen attack, induced a set of cysteine proteinases. The ectopic expression of the endogenous cysteine proteinase inhibitor, cystatin, was able to block PCD by inhibiting these cysteine proteinases. Sin and Chye (2004) showed that the genes encoding two serine

proteinase inhibitors (SaPIN2a and SaPIN2b) showed highest expression and accumulation of protein during early floral development using northern analysis and western blot analyses. Besides this early developmental period, floral tissues destined to undergo developmental PCD (the stylar transmitting tissue, the stigma and the vascular bundles) also showed localization of *SaPIN2a* and *Sapin2b,* as determined using *in situ* hybridization and SaPIN2a and SaPIN2b protein accumulation, as determined using immunolocalization. This result suggests that these serine proteinase inhibitors may function endogenously in impeding PCD during flower development more likely to guard the embryo from proteinases associated with PCD within the surrounding floral parts.

Molecular evidence for the involvement of Kunitz proteinase inhibitors in the complex machinery of PCD is rare. Karrer and co-workers (1998) were the first to hypothesize that KPIs may be involved in PCD when they made a cDNA library from tobacco leaves that were undergoing the hypersensitive response (HR) due to the infection of tobacco mosaic virus (TMV). Park and co-workers (2001) proposed that the soybean Kunitz trypsin inhibitor (SKTI1) is involved in cellular defense responses and thereby might be associated with the hypersensitive response (HR). Li *et al.,* (2008) isolated and characterized a Kunitz trypsin inhibitor gene (*AtKTI1*) encoding a functional KTI protein in *Arabidopsis*. By both over-expression (reduction in lesion formation) and RNAi silencing (increased lesion formation) approaches using PCD-eliciting fungal toxin fumonisin B1 (FB1), they demonstrated that the protein can modulate plant–pathogen related PCD. However, all of these findings concerning the involvement of KPIs in PCD only suggest that they play an antagonistic role to prevent PCD, but the exact mechanism of their involvement remains unclear.

1.6.6 Kunitz proteinase inhibitors in the control of flowering

Recently plant serine proteinase inhibitors have been implicated in the control of flowering. To understand the mechanism of tillering in rice, Yeu *et al*., (2007) applied a proteomics methodology to identify some of the proteins involved. Using two-dimensional gel electrophoresis and mass spectrometry, analysis of the basal nodes from two rice cultivars that differed in the numbers of tillers, identified a rice serine proteinase inhibitor, OsSerpin that accumulates in greater amounts in the high-tillering ‘Hwachung’ rice when compared with the relatively low-tillering ‘Hanmaeum’. Therefore, they concluded that this serine proteinase inhibitor may be involved in the tillering mechanism in high-tillering ‘Hwachung’, but in an, as yet, non-determined way.

Kim and coworkers (2009) suggested that a function of the *Kunitz trypsin inhibitor*, (*AtKTI*) in *Arabidopsis* may be involved with hastening flowering time. The mutant phenotype which flowered earlier than wild type under both long and short day conditions was caused by over expression of *AtKTI1*. No change was observed in the expression of a flowering integrator gene in the vegetative state in *AtKTI1* over-expressing plants. At the 22 day reproductive stage, an increased level of *AtKTI1* expression resulted in a reduced level of *FLC*, a strong floral repressor, and increased levels of *FT* and *AP1* which are downstream floral integrators. Therefore, the authors concluded that the increased *AtKTI1* expression contributed to the reduced level of *FLC* expression, which then causes early flowering by activating *FT* and subsequently *AP1* in an unknown way in the mutant plants. Pereira and co workers (2011) observed an accumulation of Kunitz trypsin inhibitors in developing buds, sepals, and petals and in the sex organs during anthesis in passion fruit flower (*Passiflora edulis* Sims). Though the exact mechanism is not known, it can be hypothesized that *KPIs* play a dual role during flowering. One is to act to protect the flower from herbivore attack thereby ensuring pollinator visitation and fruit setting, while a second is to ensure protein supply to the developing embryo, and so act as a storage protein.

1.6.7 Kunitz proteinase inhibitors in plant defenses against abiotic stressors

Numerous studies have revealed that abiotic stresses such as cold, heavy metal stress, salinity, and drought enhance the expression of serine proteinase inhibitors in plants (Fan and Wu, 2005). Though Kunitz proteinase inhibitors act mainly as storage proteins and are involved in defense responses under insect attack, the possibility of this proteinase inhibitor also being involved in the abiotic stress response is also plausible.

Under a toxic level (50 µm) of Al^{3+}, induction of a Bowman-Birk inhibitor gene after 8 h of treatment has been observed in the roots of *Arabidopsis* plants (Richards *et al.*, 1994) by northern blot analysis. In wheat, the expression of *wali3* and *wali5,* two putative Bowman-Birk proteinase inhibitors increased in root tips after two days of treatment with toxic levels of several metals tested (Cd, Fe, Zn, Cu, Ca, In, and La) (Snowden *et al.*, 1995). The site of expression of *wali-3,* and *wali-5* in root tips was identified, using, *in situ* hybridization as predominantly in the cortical tissue of the root.

Srinivasan and coworkers (2009) generated transgenic tobacco plants that over-expressed (using the *35S* promoter) the serine proteinase inhibitor *NtPI*, which belongs to the serine

proteinase inhibitor family of tobacco. The transgenic plants exhibited tolerance to NaCl (300mM), variable pH (pH 4 to 8) and sorbitol (100 to 300 mM), together with enhanced seed germination, root length and root–shoot ratio, as well as significantly enhanced total chlorophyll content when compared with wild type plants. The probable reason behind this could be the constitutive expression of *NtPI* repressed the rate of protein breakdown by proteases under the elevated stress conditions.

Finally, a 22-kDa (BnD22) protein belonging to the SKTI family has been shown to be induced in leaf tissues of *Brassica napus* when exposed to water deficiency (Downing *et al.,* 1992). A similar induction was observed in hypocotyls while expression of the gene was undetectable in roots. From this study the authors proposed that *BnD22* falls into a new class of genes of the SKTI family which is strictly induced in vegetative tissues during water deficiency. A 27 kDa potato Kunitz proteinase inhibitor gene that is induced by wounding in leaves was also up-regulated by water deficiency (Kang *et al.*, 2001). The exact machinery underlying the activation of these KPIs responses under abiotic stress conditions remains unclear. Therefore, it is important to focus on the involvement of this gene family under abiotic stress conditions *in vivo*.

1.7 Regulation of plant Kunitz proteinase inhibitors

The key hormones abscisic acid (ABA), cytokinin (CA), auxin (IAA), brassinosteroids, strigolactone, salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) are all involved in different aspects of plant growth and development, as well as the responses of plants to stress including both biotic and abiotic. They work through antagonistic and synergistic interactions collectively known as signaling crosstalk (Fujita *et al.*, 2006). The complex hormonal response pathways in plants are involved in many different processes throughout the life cycle. Seed germination, which is a complicated process, is an example of a process that is controlled by a complex interaction of hormones (Kucera *et al*., 2005).

Under normal circumstances, abscisic acid (ABA) controls the dormancy of seeds, as well as the sprouting of buds and tubers whereas gibberellins (GA) release dormancy (Linkies and Leubner-Metzger, 2012). On the other hand, ABA is also widely involved in responses to osmotic stress including drought, salinity stress and low and high temperature stress. The plant hormone ethylene, which is present throughout the life cycle, is involved in embryo, leaf, and flower and fruit development and during abscission and senescence. However, ethylene is also

found to be involved in drought and in regulating responses to pathogen attack (von Dahl and Baldwin, 2007). Though jasmonic acid (JA) and its volatile metabolite methyl jasmonate (MeJA) are mainly known to be involved in the responses of plants to herbivore attack, they also regulate pollen and embryo development and senescence (Linkies and Leubner-Metzger, 2012).

The complex hormonal regulation of proteinase inhibitors under different conditions has been reported. Expression of the *Pin2* gene from tomato and potato has been found to be regulated in response to mechanical wounding, but failed to accumulate in both potato (droopy) and tomato (sitiens) ABA-deficient mutants (Pena-Cortes, 1995). After treating with the JA biosynthesis precursor, linolenic acid, expression of *Pin2* was observed in the mutant plants. This result implicates the involvement of both ABA and JA in controlling the expression of *Pin2.* It has been recently reported by Laluk and Mengiste (2011) that a serine proteinase inhibitor gene (*UPI*) from *Arabidopsis* is considerably up-regulated in leaves by exogenous application of jasmonate (JA), salicylate (SA), and abscisic acid (ABA) but down-regulated by ethylene (ET). They further showed that the loss of function mutants displayed delayed flowering which was not dependent on photoperiod and also the mutant showed increased susceptibility to the nectrotrophic fungi *Botrytis cinerea* and *Alternaria brassicicola* and more damage against the generalist herbivore *Trichoplusia ni*. This detailed study indicates the importance of *UPI*, and highlights that this gene is under the control of complex hormonal regulation.

Proteinase and proteinase inhibitors coexist in all living organisms. Under normal development, and in response to different biotic and abiotic stresses, proteinase inhibitors prevent non-scheduled degradation of proteins by proteinases. As plant Kunitz proteinase inhibitors can act as storage proteins and are also induced in response to herbivory and abiotic stress, it is likely that they are controlled and regulated by a complex set of hormonal pathways. If crosstalk occurs between the biotic and abiotic stress response pathways (Fujita *et al.*, 2006), the same inhibitor can be induced by both biotic and abiotic stresses depending on the transcription factors (TFs) binding to the promoter region. Therefore, proteinase inhibitors belonging to the same family can be involved in diversified functions in response to developmental and environmental cues in the same plant species. As observed by Philippe *et al.,* (2009), the 12 Kunitz proteinase inhibitor genes (*pKPIs*) from six different linkage groups from poplar are constitutively expressed throughout the life cycle, but are also induced at a higher level by larval feeding of forest tent caterpillar (FTC, *Malacosoma disstria*).

Thus a model of induction of *KPI* is suggested (Figure 1.2) based on the biotic and abiotic stress response pathways in plants, and existing crosstalk depending on the molecular convergence of the two pathways (Fujita *et al*., 2006). ROS species and plant hormones such as ABA, ET, JA and SA are upregulated in plants experiencing abiotic or biotic stress. Plant defenses to disease resistance and ABA signaling are also controlled *via* the central role played by reactive oxygen species (Guan *et al.*, 2000). These thus trigger the activation of genes of the MAP Kinase cascade (Yuasa *et al.*, 2001). Later, transcription factors get activated and defense related genes including proteinase inhibitors are induced as a late response (Ryan, 2000).

In summary therefore, it can be speculated that depending on the type of stress and the involvement of key hormones and transcription factors, the *Kunitz proteinase inhibitors* in the same plant can be induced in response to diverse developmental and environmental conditions.

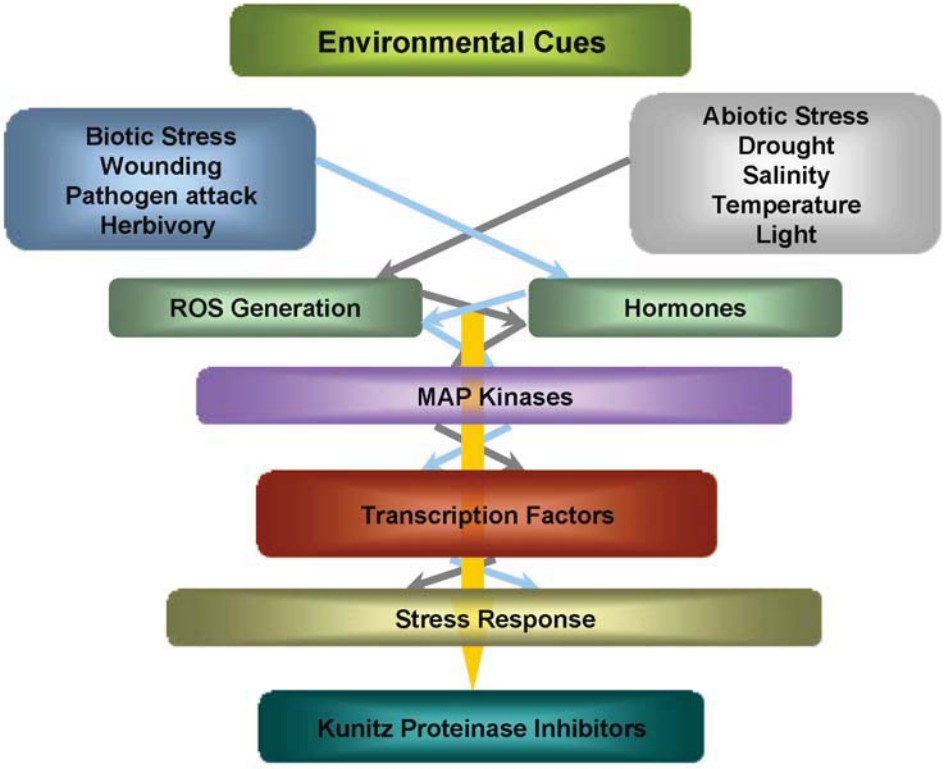


Figure 1.2: Points of convergence between environmental cues and the induction of plant Kunitz proteinase inhibitors (modified from Fujita *et al.*, 2006)

1.8 Concluding statement and research hypothesis

All living organisms experience various kinds of biotic and abiotic stress. In response, plants have evolved a spectacular range of defense systems which may be structural, chemical and/or molecular to tolerate these adverse conditions. In some instances, the cellular levels of compounds or proteins that are indispensable for metabolic processes and survival under normal conditions are increased to act as defensive compounds under adverse conditions (Mosolov and Valueva, 2011). An example of one of these protein groups are the proteinase inhibitors.

1.8.1 Why study Kunitz proteinase inhibitors in white clover?

World wide, white clover (*Trifolium repens* L.) is recognized for its high forage quality. White clover can be considered as one of the premium pasture species as it fixes atmospheric nitrogen in a symbiotic interaction with *Rhizobia*, and thereby can reduce fertilizer cost. It has the potential to regenerate from stolons after grazing and thereby reduces the cost of seeding pastures every year. As the plant species has many such advantages, efforts have been made to incorporate it into the grazing lands of Europe, some parts of west Asia and North America. However, the proportion of this species in the established grazing lands is quite small (Hopkins *et al.,* 1985). The reasons behind this could be high or low soil pH, nutrient deficiency in the soil, choice of the wrong cultivar and/or overexploitation of fertilizers and pesticides. As damage by pests and diseases is regularly observed, they are also partly responsible for the failure or poor growth of clover. The crop is also vulnerable to drought that reduces its growth rate and productivity (Parsons *et al*., 2011).

Biotic and abiotic factors that increase the tolerance of plants to pests, diseases, extreme heat and drought are critical to the future success of white clover cultivation (Rhodes *et al*., 1994). Hence identifying unique genes for achieving enhanced tolerance to biotic and abiotic stresses is a major research focus. Since plant growth and development is influenced by environmental cues, it is perhaps not surprising that exposure to unfavorable conditions causes many physiological and morphological changes, and these can affect plant productivity at every developmental stage. Though many genes involved in the interactions of plants to insect damage, water deficit, and high or low temperature tolerance have been identified over the intervening years, there remains a major gap in terms of precise information because most studies are focused mainly on the analysis and characterization of genes in model species such as *Arabidopsis*. For example, significant differences exist between *Arabidopsis* and cereal root

tissue organization and therefore there is a difference in stress response (Smith and Smet, 2012). Such studies with model plants are fundamental, but it is also important to identify genes that are involved in tolerance mechanisms in crop plants of agricultural importance.

The involvement of proteinase inhibitors in terms of mediating plant tolerance against unfavorable growth conditions is of considerable research interest. As summarized, Kunitz proteinase inhibitors are important in this context as it is well established that KPIs play an important role in plant defense mechanisms against insect herbivory in different plant species. Since the initial achievement of obtaining increased pest tolerance by the introduction of a gene coding for trypsin proteinase inhibitor (Hilder, 1987), researchers have been using different PIs to increase tolerance against insects in different crop plants. But the significance of proteinase inhibitors in the responses of plants to water deficit, high pH and other stress conditions has received less attention.

Until now, no information is available about the occurrence of KPI gene family in white clover. It is necessary, therefore, to initially isolate *KPI* sequences from white clover and then to determine whether the gene family is differentially responsive to a range of developmental and environmental cues.

1.8.2 Hypothesis

Considering this research context, the generalized hypothesis to be tested states that:

The transcription of members of the *Kunitz Proteinase Inhibitor (KPI)* gene family in white clover is not only responsive to endogenously generated cues in response to biotic stress but also those generated during plant growth and development and in response to sub-optimal growth conditions caused by abiotic stresses.

1.9 Research objectives

To test the hypothesis, the following four objectives were set–

- Identify cDNAs which are complementary to genes that encode Kunitz proteinase inhibitor proteins in white clover.
- Investigate the transcriptional regulation of these *KPI* genes at different developmental stages of the plant.
- Determine the transcriptional regulation of *KPIs* under biotic stresses such as wounding or herbivory attack.

- Investigate the involvement of *KPIs* at the transcriptional level under abiotic stresses such as osmotic stress and nutrient deficiency.
-Identification of the regulatory elements in the promoter region of the identified *KPI* genes.

The above objectives will provide vital information on specificity, function and expression of *KPI* genes in white clover at the transcriptional level. These results will provide a foundation for future studies to begin to understand the significance of the functions of the encoded proteins in white clover growth and development including responses to environmental cues.

Chapter 2

Materials and Methods

2.1 Plant Material

2.1.1. Establishment and maintenance of stock plants

The white clover cultivar Grassland Huia was used as the genetic background in different sets of experiments. To establish stock plants, seeds (collected from AgResearch Grasslands, Palmerston North) were planted in soil (Premier Seed Raising Mix, Daltons, Matamata, New Zealand) and the emerged seedlings maintained in a glasshouse. After one month, and to maintain a single genetic line, a plant arising from a single seed was selected randomly and maintained as a stock plant, eventually in 5-L-capacity pots. Vegetative propagation from this stock plant was carried out at least twice a year, and the old stock plants discarded. To propagate stock plants, the apical part of the stolon was excised just proximal to node 4, all leaves except the first emerged leaf were removed and the cuttings planted into soil. Plants were supplemented with nutrients once a month with the addition of Thrive^{R} Soluble All Purpose Plant Food (Yates New Zealand, Auckland, New Zealand).

2.1.2 Vegetative propagation of plant material for experimental use

To provide plant material for different sets of experiments, stolons with four nodes were excised from the stock plants as described (2.1.1). The cuttings were then placed in pots containing vermiculite and watered regularly with half strength (0.5 x) Hoagland's solution (Appendix 8) and rooted for at least one week. The healthy and morphologically similar rooted stolons were then used for different sets of experiment.

2.1.3 Experimental plants and treatments

For different sets of physiological experiments, generally, the plants were grown in a controlled temperature room or Contherm Biosyn Series 6000 Plant Growth Chambers (Model 620RHS, Contherm Scientific Ltd, Wellington, New Zealand). The growth conditions were typically maintained at 22°c during the day and 14^{0}c during the night, with a constant relative humidity (RH) of 65% and with a light intensity of 180 µE over a 14 h photoperiod.

2.1.3.1 Spodoptera litura feeding experiments

The generalist herbivore tobacco cutworm (*Spodoptera. litura*) was used in the study to monitor the expression pattern of *Tr-KPIs* under herbivory attack. For the feeding experiments, the stolons were grown as described (2.1.3). The 24 h insect feeding experiment was conducted at Plant and Food Research, Mt. Albert, Sandringham, Auckland. For this, the larvae were raised on lima bean artificial diet where the following ingredients were used:

Ingredients	**Weight (g)**
Lima bean powder	60 g
Agar	32 g
Brewer's yeast	40 g (roller dried, unsalted)
Wheatgerm	48 g

The above ingredients were stirred in 800 mL of tap water, microwaved on high power for 7 - 8 min or until well boiled, with stirring at 1 to 2 min intervals. The following ingredients were thoroughly mixed into 200 mL of water and added to the microwaved components once they had cooled to 65^0C:

Ingredients	**Measurement**
Flax seed oil	0.8 mL
Wheatgerm oil	1.6 mL
Vanderzant vitamin mixture	8 g
Ascorbic acid	3.2 g
Methylparaben (Nipagin)	1.6 g
Sorbic acid	0.8 g
Penicillin	280 mg
Streptomycin	280 mg
Prochloraz (Octave 50W)	16 mg

2.1.3.2 Nematode feeding experiment

The nematode infestation experiment was performed at AgResearch Grasslands where stolon cuttings of cv. Huia and resistant (17R) and susceptible (23S) lines for cyst nematode were rooted in soil in 300 x 450 mm trays. The cuttings were grown in the green house at 20-25^{0}C on a plant heating mat for 20 days. Well established stolons were selected and planted into 60 mm diameter plastic cups containing sand and were supplied with nutrient solution (0.5 x Hoagland's solution) on a regular basis. The eggs of white clover cyst (*Heterodera trifolii*) and root-knot (*Meloidogyne trifoliophila*) nematodes were supplied by AgResearch Grasslands. To infest the plants, after two days of establishment in sand a hole was made in the middle of each cup and inoculum were added under the plant at a rate of 4000 eggs/3 mL of water for *M. trifoliophila* and 3000 eggs/3 mL of water for *H. trifolii*. The infected plants were allowed to grow for eight days and the plant tissues (leaf and root) were harvested after 4 and 8 days of inoculation. Harvested roots were excised and washed in running tap water to remove sand particles, blot dried and snap frozen in liquid nitrogen and stored at -80°C until required.

For detection of nematode, some of the harvested roots after 8 days of inoculation were stained using the following protocol:

• Infected roots that were washed under running tap water were placed in a beaker containing 1.5% (v/v) NaOCl (45 mL water + 25 mL Janola) and left for 5 min with agitation.
• After that, the roots were rinsed in running water for 30 sec and allowed to stand in tap water for 15 min and blot dried.
• The roots were then incubated in boiling stain [0.05% (w/v) aniline blue in 33% (v/v) glycerol, 33% (v/v) lactic acid] for 1 min, cooled to room temperature, rinsed in running water and then blot dried.
• To count nematodes under a microscope, the roots were placed in acidified glycerol [a few drops of lactic acid added to 50% (v/v) glycerol].

2.1.3.3 Water deficit experiment

A water deficit treatment was imposed by the complete withholding of water to the plants. To do this, four noded stolons (excised as described in 2.1.2) were rooted in a 1:1 vermiculite and perlite mixture in 1.2-L- capacity pots and watered regularly with 0.5 x Hoagland's solution. The plants were grown for four weeks before the experiment started. Two different

treatments were applied: (i) non pre-stressed (NPS) (without a previous period of water withholding) and (ii) pre-stressed (PS) (involving a one week period of water withholding, a week of rehydration and then the water withholding period (outlined in Figure 3.24). For qRT-PCR, samples were collected daily until the moisture content dropped to approximately 32 % for both the PS and NPS treatments. Before the experiment started, excess water was drained away from the media in the pots, and the moisture content of vermiculite and perlite mixture was determined using the gravimetric method of Robinson (1974) and was expressed as a percentage using the following formula:

$\Theta = (Ww-Dw)/Ww \times 100$

Where,
Ww: is the wet weight of vermiculite: perlite mixture and water
Dw: is the dry weight of vermiculite: perlite mixture
At the appropriate harvesting intervals, the moisture content was measured using the same method.

2.1.3.4 Phosphorus limiting experiment

For the phosphorus limiting experiment, stolons were excised and rooted in vermiculite supplemented with 0.5 x Hoagland's medium (as described in 2.1.2). After 5 days, the stolons were transferred to 0.5 x Hoagland's media in 0.6-L-capacity PVC pipes (Figure 3.31). After 14 days, the stolons were either treated with Hoagland's medium containing 5 µM P (Pi deficiency treatment) or 0.5 mM P (Pi sufficient treatment) over the time-course of the experiment.

2.2 Chemicals used

Unless otherwise stated, the chemicals used in this study were obtained from Sigma Aldrich Company (St. Louis, Mo., US), Duchefa Biochemie BV (Haarlem, The Netherlands), Merck Ltd. (Darmstadt, Germany), Promega Corporation (Madison, WI, USA), QIAGEN GmBH (Hilden, Germany), Life Technologies Corp. (Grand Island, NY, USA), Roche Applied Sciences (Roche Diagnostics GmbH, Mannheim, Germany) and Bio-Rad Laboratories (Hercules, CA, USA).

2.3 Molecular Biology Protocols

2.3.1 Nucleic acid isolation

2.3.1.1 Total RNA isolation

Total RNA was extracted from different samples using the Hot Borate method following the protocol used by Hunter and Reid (2001) and Moser *et al*., (2004). All glassware, mortars and pestles, and spatulas used for RNA extraction were wrapped with aluminium foil and baked at 180°C for 16 h. All plastic-ware was treated overnight in 0.3% (v/v) hydrogen peroxide and was rinsed before use with DEPC-treated water. All solutions used were made to required concentrations using DEPC-treated water. Reagents and chemicals used in the protocol are:

- Borate Buffer: 200 mM di-sodium tetraborate decahydrate, pH 9.0, 30 mM EGTA, 1% (w/v) sodium deoxycholate and 1% (w/v) SDS.
- RNA extraction buffer: Borate Buffer (pH 9.0), 2% (w/v) PVP-40, 1% (w/v) IGEPAL CA-630 and 1 M DTT.
- Proteinase-K (20 mg/mL)
- 2 M KCl
- 4 M LiCl
- 3 M sodium acetate, pH 5.2
- Chloroform: isoamyl-alcohol [24:1 (v/v)], isopropanol, and 80% (v/v) ethanol.

Steps followed for total RNA extraction:

- To extract RNA from frozen leaf, stolon, root and seed tissues, typically 20 mg to 200 mg fresh weight was ground to a fine powder in liquid nitrogen using a pre-chilled mortar and pestle or pre-chilled microtube with a micro pestle.
- The powder was transferred into a microtube containing five volumes (w/v) of warm (80°C) extraction buffer, and the ground tissues were vortexed to resuspend in the extraction buffer.
- Proteinase-K (0.75%; v/v) was added and the slurry was incubated at 42°C with shaking for 90 min.
- Immediately after incubation, 2 M KCl was added [0.08 (w/v)] to give a final concentration of 160 mM, the solution mixed and incubated in an ice bath, with shaking, for 30 min.

- The mixture was then centrifuged at 17,000 x *g* for 20 min at 4°C, the supernatant transferred to a fresh tube and an equal volume of cold (4°C) 4 M LiCl (to give a final concentration of 2 M) added to precipitate the RNA for 16 h at 4°C.
- The next day, the precipitate was collected by centrifugation at 17,000 x *g* for 30 min at 4°C, resuspended in 200 μl of DEPC-treated water prior to the addition of 20 μL of 3 M sodium acetate, to give a final concentration of 0.3 M, and 200 μL of chloroform/isoamyl alcohol [24:1 (v/v)] was added.
- The aqueous and organic phases were vortexed for 30 sec and then separated by centrifugation at 17,000 x *g* for 5 min at room temperature. The upper aqueous phase was carefully pipetted and transferred into a fresh sterile microfuge tube.
- One volume (200 μL) of isopropanol was then added, the contents mixed well by inversion and the tubes incubated on ice for 1 h to precipitate the RNA.
- The RNA was collected by centrifugation at 17,000 x *g* for 30 min at 4°C , washed with 500 μL of 80% (v/v) ice-cold ethanol before collection at 17,000 x *g* for 10 min. The pellet was air dried for 10 min and then resuspended in 500 μL of DEPC-water.
- To remove genomic DNA, the RNA was routinely reprecipitated with 4 M LiCl (added to a final concentration of 2 M) and incubated for 16 h at 4°C.
- The next day, the RNA was pelletted by centrifugation at 17,000 X g for 30 min at 4°C, then washed with ice-cold 80% (v/v) ethanol, centrifuged at 17,000 X g for 10 min, air dried for 10 min and then resuspended in 20 μL of DEPC-water.
- RNA was quantified by NanoDrop^{R} ND-1000 spectrophotometry (NanoDrop Technologies, Montchenin, DE, USA) (2.3.1.3) and were stored at -80°C until required.

2.3.1.1.1 DNase treatment

For some protocols, genomic DNA-free RNA samples were prepared using an RNase-free recombinant DNase (Roche) treatment.

- Total RNA (2 to 10 μg), extracted as described in 2.3.1.1, was mixed with 5 μL of 10 x incubation buffer and 1 μL of DNase I (10 U), 1 μL of Protector RNase Inhibitor (10 U)before water was added to give a final volume of 48.4 μL.
- The mixture was then incubated at 37^{0}C for 20 minutes.
- The reaction was stopped by the addition of 1.6 μL of 0.25 M EDTA pH 8.0 (to give a final concentration of 8 mM) and the reaction was heated at 75^{0}C for 10 min.

2.3.1.2 Isolation of genomic DNA (gDNA) using the CTAB method

Extraction of genomic DNA from white clover was performed following the modified protocol of van der Biezen *et al.*, (1996).

Reagents used for the protocol were:

- Extraction buffer: containing 0.1 M Tris-HCl, pH7.5, 0.35 M sorbitol, 5 mM EDTA and 20 mM $Na_2S_2O_5$.
- Nucleus lysis buffer: 0.2 M Tris.HCl, pH 7.5, 50 mM EDTA, 2 M NaCl, and 2% (w/v) CTAB
- 5% (v/v) Sarkosyl
- 24:1 (v/v) chloroform/ isoamylalcohol
- Isopropanol
- 70% (v/v) Ethanol
- 10 mg/mL RNase A

Steps followed were:

- Leaf tissue (100 mg) was ground in liquid N_2 using a chilled mortar and pestle and vortexed in 0.5 mL of chilled extraction buffer.
- The slurry was then centrifuged at 17,000 x *g* for an hour at 4°C.
- After discarding the supernatant, the pellet was dissolved in 0.2 mL of extraction buffer, 0.3 mL of nucleus lysis buffer and 0.1 mL of 5% (v/v) sarkosyl.
- After mixing by inversion, the mixture was incubated at 65^0C for 1h.
- Chloroform/ isoamylalcohol [0.5 mL of 24:1 (v/v)] was added to the mixture and centrifuged at 17,000 x *g* for 15 min to separate the aqueous and organic phase and the aqueous phase was transferred to a clean tube
- An equal volume of ice-cold isopropanol was added to precipitate the DNA for 20 min, followed by centrifugation at 17,000 x *g* for 15min.
- After decanting the supernatant, the pellet was washed with 70% (v/v) ethanol for 30 min and air dried for 10 min.
- Finally the pellet was dissolved in TE buffer and gDNA was quantified by NanoDrop^R ND-1000 spectrophotometer (NanoDrop Technologies, Montchenin, DE, USA) (2.3.1.3).

2.3.1.3 Quantification of nucleic acid

The RNA/DNA concentration was determined by measuring the absorbance at 260 nm (A_{260}) using an NanoDrop ND-1000 spectrophotometer V3.6 (Thermo Scientific, USA). The purity of the nucleic acid was determined using the A260/A280 ratio (measured against a water blank

for RNA and TE buffer for DNA), with relatively pure solutions having an A260/A280 ratio of >1.8 (Sambrook *et al*., 1989).

2.3.2 Synthesis of cDNA

To synthesize the first single strand DNA, the Transcriptor First Strand cDNA synthesis kit (Roche) was employed using oligo (dT) primer.

Reagents used:

- Transcriptor Reverse Transcriptase (20 U/µL)
- Transcriptor RT reaction buffer (5X): 250 mM Tris-HCl, pH 8.5, 150 mM KCl and 40 mM $MgCl_2$.
- Oligo $(dt)_{15}$ primer (final concentration 2.5 µM)
- Protector RNase Inhibitor (40 U/µL)
- dNTP mixture (final concentration 1 mM)

To obtain the sequence of the 3' UTR region, a 3'RACE OligoT Adapter primer was used (Appendix 5).

Programme used for cDNA synthesis:

Steps	**Temperature**	**Time**
Denaturation	65°C	10 min
Incubation	55^0C	30 min
Inactivation	85^0C	5 min

Steps followed in the protocol were:

- Total RNA (1 µg) was combined with the Oligo $(dT)_{15}$ primer in 0.2 mL capacity tubes and the volume was adjusted to 13 µL with DEPC-treated water.
- The RNA and primers were denatured at 65°C for 10 min and placed on ice immediately.
- Seven µL of master reaction mixture containing 5X Transcriptor RT Reaction Buffer, Protector RNase Inhibitor (40 U/µL), 10 mM dNTP-Mix and Transcriptor Reverse Transcriptase (20 U/µL) were then added.
- The tubes were placed in the Palm-Cycler (Corbett Life Science, Corbett Research Pty Ltd, Australia) and, typically, cDNA synthesis was carried out at 55°C for 30 min
- The Transcriptor Reverse Transcriptase was then heat inactivated at 85°C for 5 min.

- The cDNA was either used immediately or stored at -20°C until required.

2.3.3 Polymerase chain reaction (PCR)

2.3.3.1 Primer design

For isolation of the gene of interest from white clover, degenerate primers were designed using *Medicago truncatula*, *Glycine ma*x and *Cicer arietinum* conserved region sequences (Appendix 3, chapter 3.1.2). Each primer (Sigma) was dissolved in sterile water to give a concentration of 1 mM and stored at -20°C until required. To prepare working stock, 1 µL of the stock (1 mM) was added to 99 µL of sterile water to give a concentration of 10 µM. To obtain the 3'UTR of each of the *Tr-KPI* genes, two primers were designed in the 3' region of the known sequences to provide 60 to 100 bp of overlap (Appendix 4 and 5).

For qRT-PCR analysis, specific primers were designed according to the general requirements of qRT-PCR primers [T_m = 60°C (± 1°C)], a minimal secondary structure, and an inability to form stable dimers] and based on the cDNA sequence for the representative target genes (Appendix 7). All primers annealed such that the size of the amplified products was as similar as possible and within the range of 150- 170 nucleotides. The efficiency of all the primers was determined using the standard curve method (Ruijter *et al.,* 2009).

2.3.3.2 General PCR protocol for amplification of cDNA

General PCR setup:

Forward primer (10 µM)	1 µL
Reverse primer (10 µM)	1 µL
2X PCR Master Mix containing *Taq* DNA polymerase, dNTPs, $MgCl_2$, and reaction buffer	10 µL
cDNA	1 µL
Sterile water	7 µL
Total volume	20 µL

PCR optimization was undertaken to find the ideal annealing temperature for the amplification of each transcript and this was achieved using gradient PCR. The typical PCR programme used was:

Steps	Temperature	Time	Cycle
Initialization	94 to 95^{0}C	10 min	1
Denaturation	94 to 95°C	30-40 sec	
Annealing	4 to 5^{0}C lower than the primer melting temperature	30-40 sec	30
Extension/Elongation	72 to 75^{0}C	Depending on product size (e.g. 1 kb for 1 min)	
Final extension	72 to 75^{0}C	10 min	

To obtain the sequence of the 3'UTR region, PCR was performed using a 3'RACE adapter primer and specific primers designed for each sequences (Appendix 5). The product of the first round was then used for a second round PCR using the 3'RACE adapter primer and a nested primer.

Similar PCR conditions and programmes were used for PCR screening of *E. coli* colonies containing the anticipated fragment but the initial incubation was performed at 95°C for 5 min to promote cell lysis of *E .coli* before amplification of the DNA.

2.3.3.3 Agarose gel electrophoresis

DNA fragments, generated by PCR, were separated using agarose gel electrophoresis.

Reagents used for this purpose were:

- UltraPURE™ agarose (Life Technologies)
- 50 X TAE Buffer
- Loading dye [10 X SUDS (0.1 M EDTA, pH 8.0, 50% (v/v) Glycerol, 1% (w/v) SDS, 0.025% (w/v) bromophenol blue)]
- Ethidium bromide (10 mg/ml)
- HyperLadder™ 1 DNA ladder (Bioline, London, UK)

To prepare a 1% (w/v) gel, 1 g of agarose powder was dissolved in 100 mL of 1 X TAE buffer, and the cooled agarose poured into the appropriate gel tray with a sample-well-forming comb inserted. After polymerization of the agarose gel, running buffer (1 X TAE) was added to gel electrophoresis apparatus and the comb was removed carefully. PCR products were mixed with 2 µL of 0.1% (v/v) SUDS, mixed properly and loaded into the wells. Generally, PCR

products were separated at 90 V for 1 h. For products smaller than 0.2 kb, a 2% (w/v) 1X TBE gel was used and the fragments were separated at 100 Vs for 30 min for better resolution.

After electrophoresis, the gel was stained with 0.1 μg mL^{-1} ethidium bromide for 10 to 15 min, destained with water for 5 to 10 min and then the DNA fragments were visualized using a Gel Doc 2000 Gel Documentation System from Bio-Rad Laboratories (US). For in-gel quantification, the quantity of DNA was estimated by comparing the relative intensity of the DNA band to the DNA ladder.

2.3.3.4 Quantitative RT-PCR (qRT-PCR)

qRT-PCR was performed using the LightCycler® 480 Real-Time PCR (Roche) and system series software 1.7, with three technical replicates of each cDNA sample (20-fold dilution). SYBR green I was used to monitor efficient DNA synthesis.

Reaction setup for qRT-PCR:

Forward primer (10 μM)	0.5 μL
Reverse primer (10 μM)	0.5 μL
2X LightCycler® 480 SYBR Green I Master Mix	5 μL
cDNA	2.5 μL
Sterile water	1.5 μL
Total volume	10 μL

Master mixture and cDNA templates were dispensed into 96 well plates. The following programme was used:

Steps		**Temperature**	**Time**	**Cycle**
Preincubation		95^0C	5 min	1
Amplification	Denaturation	95°C	10 sec	45
	Annealing	60°C	10 sec	
	Extension	72°C	10 sec	
Melting curve		95°C	5 min	1
Cooling		40°C		1

Relative transcript abundance was determined by comparative quantification to the geometric mean of the two reference genes, *β-actin* and *GAPDH* (Pffafl, 2001). Fluorescence measurements were performed at 72^0C for each cycle and continuously during final melting.

2.3.4 Isolation of the promoter region using genome walking methodology

2.3.4.1 Digestion of genomic DNA

Genomic DNA (gDNA) was digested with three blunt cutting restriction enzymes (*EcoRV*, *ScaI*, *DraI*). For each digestion, 2.5 µg of gDNA, 8 µL of restriction enzyme (10 U/µL) and 10 µL of restriction enzyme buffer were mixed, made to 100 µL with water, mixed gently and incubated at 37^0C for overnight. The completion of digestion was checked by running 5 µL of the reaction on a 0.5% (w/v) agarose gel and staining with ethidium bromide.

2.3.4.2 Purification of genomic DNA

The procedure for purification of genomic DNA after digestion was as follows:

- Equal volume of phenol (95 µL) was added to each reaction tube, the mixture vortexed at slow speed and then centrifuged briefly to separate the aqueous and organic phases.
- The aqueous phase was transferred to a fresh tube and an equal volume of chloroform was added, the mixture vortexed and then centrifuged briefly.
- The clear supernatant was transferred to a fresh tube and 2 volumes of 100% ice cold ethanol, 1/10 volume of 3 M NaOAc, and 20 µg of glycogen were added. The mixture was then vortexed for 10 sec and centrifuged at 17,000 x *g* for 10 min.
- After decanting the supernatant, the pellet was washed with 100 µL of ice-cold 70% (v/v) ethanol and centrifuged at 17,000 x *g* for 5 min.
- After decanting the supernatant, the pellet was air dried for 15 min and finally dissolved in 20 µL of water.

2.3.4.3 Adaptor sequence, primers and ligation of genomic DNA to adaptors

The adaptor for genome walking has been specially designed, so that amplification of non-target products from the adaptor primers alone is inhibited, and so reducing the background. To prevent amplification of the adaptor, the lower strand has the 3' end blocked with a terminal amine moiety.

For ligation with the adaptor (Appendix 6) for each library, the reaction was set up in the following way:

Digested and purified gDNA	4 µL
Adaptor (25 µM)	1.9 µL
10X ligation buffer	1.6 µL
T4 DNA ligase (6 U/µL)	0.5 µL
Total volume	8 µL

The solutions were mixed together, incubated at 16°C overnight in a thermocycler and then to deactivate the ligase, incubated for 5 min at 70°C. Water (72 µL) was then added to each tube and the contents then mixed well.

Two primers for each gene were designed to amplify the 5'UTR and the promoter region in the antisense orientation. The primers were designed to anneal at least 100 bp downstream to the end of the known sequence to facilitate the alignment of genome walker products with the known sequences by creating an overlapping region. Each gene-specific primer was designed to have a very high Tm (at least 72^0C or higher) compared to the adaptor primers for efficient and specific binding of the gene specific primers since the PCR cycling conditions require efficient binding/annealing at 72°C. The adaptor primers themselves had a much lower Tm than the temperature the PCR was performed at, which means that annealing was less frequent from this strand but the gene specific primers will bind very efficiently.

2.3.4.4 PCR protocol to amplify upstream region of the selected genes by genome walking

Two sets of PCR programmes were used for the genome walking protocol: primary PCR and secondary PCR.

Primary PCR setup:

10 µM Adaptor primer (NA46/AP1, Appendix 6)	1 µL
GSP1 primer (10 µM)	1 µL
10X FastStart High Fidelity Reaction Buffer	5 µL
$MgCl_2$	
10 mM dNTPs (200 µM each)	1 µL
gDNA/adaptor ligation	1 µL

Faststart High Fidelity Enzyme Blend	0.5 µL
Sterile water	40.5 µL
Total volume	50µL

The primary PCR cycling conditions were:

Steps	Temperature	Time	Cycle
Denaturation	94^0 C	25 sec	
Primer annealing and extension	72°C	3 min	7
Denaturation	94^0 C	25 sec	
Primer annealing and extension	67^0 C	3 min	32
Final extension	67^0C	7 min	1
Hold	12^0C		1

The PCR products were examined by separating 5 µL on a 1% (w/v) agarose gel.

The same reaction mixture was used for the secondary (2^0) PCR except for the template. For the secondary PCR, 1 µL of the primary PCR product was diluted in 49 µL of H_2O and this diluted product (1 µL) was used as template in the 2^0 PCR.

Secondary PCR setup:

10 µM Adaptor primer (NA47/AP2; Appendix 6)	1 µL
GSP2 primer (10 µM)	1 µL
10X FastStart High Fidelity Reaction Buffer	5 µL
$MgCl_2$	
10 mM dNTPs (200 µM each)	1 µL
gDNA	1 µL
Faststart High Fidelity Enzyme Blend	0.5 µL
Sterile water	40.5 µL
Total volume	50 µL

The secondary PCR cycling conditions were:

Steps	Temperature	Time	Cycle
Denaturation	94^0C	25 sec	

Primer annealing and extension	72^0C	3 min	5
Denaturation	94^0C	25 sec	
Primer annealing and extension	67^0C	3 min	20
Final extension	67^0C	7 min	1
Hold	12^0C		1

Five µL of product was separated on a 1% (w/v) agarose gel. Discrete bands were excised from the agarose gel and purified.

2.3.5 DNA recovery

For column purification of the DNA fragments after completion of PCR, a QIAquick^R PCR purification kit (QIAGEN) was used. The reagents were:

- Buffer PB (binding buffer)
- Buffer PE (washing buffer)
- Elution buffer (10 mM Tris-HCl, pH 8.5)

The purification steps undertaken were:

- Five volumes of buffer PB was added to 1 volume of the PCR product.
- After placing a QIAquick column in a provided 2 mL collection tube, the samples were added to the column and centrifuged for 1 min at 17,000 x *g*.
- The flow through was discarded and the column was placed in the same tube.
- For washing, 0.75 mL of buffer PE was added to the column and the assembly centrifuged for 1 min at 17,000 x *g*.
- After discarding the flowthrough, the column was centrifuged again to remove any residual wash buffer.
- After that, the column was placed in a 1.7 µL microcentrifuge tube and 50 µL of elution buffer was added to the centre of the column and the assembly centrifuged for 1 min at 17,000 x *g*.
- The purified DNA in the eluate was quantified by agarose gel electrophoresis.

For agarose gel recovery of the DNA fragment, the QIAquick^R Gel Extraction Kit composed of Buffer QC (pH $\leq$7.5), Buffer PE (washing buffer), and Buffer EB (10 mM Tris-HCl, pH 8.5) was used. The purification steps were:

- The DNA fragment was excised from agarose gel using a clean sharp scalpel, the weight of the gel slice recorded and 3 volumes of buffer QG was added per 1 volume of gel.
- The gel with buffer QG was incubated at 50^0C until the gel slice was completely dissolved.
- One gel volume of isopropanol was added and the solution mixed well.
- A QIAquick spin column was placed in a provided 2 mL collection tube and the sample was placed in the column and centrifuged for 1 min at 17,000 x *g*.
- After discarding the flowthrough from the collection tube, 0.75 mL of buffer PE was added to the column and the assembly centrifuged for 1 min, followed by discarding the flowthrough, and then an additional centrifugation for 1 minute at 17,000 x *g*.
- To elute DNA, 30 μL of elution buffer was added after placing the column in a clean 1.7 mL microcentrifuge tube and the centrifuged for 1 min at 17,000 x *g*.
- The purified DNA in the eluate was quantified by agarose gel electrophoresis.

2.3.6 Plasmid cloning and transformation

2.3.6.1 Ligation of DNA into the pGEM® T Easy vector

DNA sequences were ligated into the pGEM®-T Easy vector (Promega) according to the protocol supplied with the cloning kit. Ligation of DNA into the vector was performed using a 1:3 to 1:10 molar ratio of vector to insert. The quantity of PCR products (insert) to be included in the ligation reaction was calculated according to the following equation:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{Insert: Vector molar ratio} = \text{ng of insert}$$

The appropriate amount of PCR product was ligated in a reaction mix that contained

- 50 ng of pGEM®-T Easy vector
- 5 μL of 2X Rapid Ligation Buffer
- Required amount of DNA
- 3 Weiss units of T4 DNA ligase (1 μL)
- Sterile water to make the final volume of 10 μL.

The ligation reaction was incubated at 25°C overnight.

2.3.6.2 Preparation of competent cells for plasmid transformation

E. coli cells used for transformation were prepared from strain DH5α (GIBCO BRL) according to the method of Inoue *et al.,* (1990) with minor modifications. The media and reagents used were:

- LB broth: containing 1% (w/v) bacto-tryptone (DIFCO Laboratories, Detroit, MI, US), 0.5% (w/v) bacto-yeast extract (DIFCO Laboratories), 1% (w/v) NaCl, pH 7.0
- 60 mM $CaCl_2$
- Glycerol

The steps used to prepare competent cells were:

- Bacterial cells, from a single colony or a glycerol stock, were cultured in 10 mL of LB broth at 37°C overnight, with shaking.
- Forty ml of LB broth was inoculated with 400 µL of overnight culture and incubated at 37°C until cell growth reached an optical density of 0.4 at 600 nm.
- The cells were then centrifuged at 3000 x *g* for 5 min at 4°C, the cell pellet resuspended gently in 10 mL of ice-cold 60 mM $CaCL_2$ followed by addition of 10 mL of ice-cold 60 mM $CaCl_2$.
- After incubation on ice for 30 min, the cell suspension was centrifuged at 3000 x *g* for 5 min at 4°C and the cell pellet was resuspended gently in 4 mL of 60 mM $CaCl_2$ containing 15% (v/v) glycerol.
- Aliquots (300 µL) of the cell suspension were transferred to microfuge tubes, snap frozen in liquid nitrogen and stored at -80°C until required for transformation.

2.3.6.3 Transformation of E. coli with pGEM®-T Easy vector

The putatively ligated pGEM®-T Easy Vector was transformed into competent cells of *E. coli* strain DH5α using the heat-shock method.
Reagents used for this purpose were:

- LB agar plate: LB broth with 1.5% (w/v) bacteriological agar supplemented with 100 µg mL^{-1} Ampicillin (LB-Amp^{100} medium)
- 0.5 mM IPTG
- 80 µg mL^{-1} X-Gal (LB Amp^{100}/IPTG/X-Gal)

- SOC media: containing 2% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, 1% (w/v) NaCl, and 2.5 mM KCl. Following sterilization by autoclaving, the media was left to cool and 2 M Mg^{2+} (consisting of 2 M $MgCl_2$ and 2 M $MgSO_4$) and 2 M glucose were added to give a final concentration of 200 mM.

For transformation, the following steps were undertaken:

- Competent *E. coli* cells, stored at -80°C, were placed on an ice bath until just thawed. Half of the ligation mixture (5 μL) was transferred to a new microcentrifuge tube on ice and 50 μL of the competent *E. coli* cells were added and the contents mixed gently.
- The cell mixture was maintained on an ice bath for 20 min and then heat-shocked at 42°C for 45 to 50 sec before being placed back immediately on the ice bath for 2 min, after which 950 μL of SOC medium was added.
- The transformation mixture was incubated at 37°C for 1.5 h, with shaking.
- An aliquot (200 μL) of cells was then plated onto a LB Amp^{100}/IPTG/X-Gal plate, air dried in the laminar air flow for 5 min, then incubated at 37°C overnight.
- Initially all blue and white colonies were selected and cultured in LB Amp^{100} broth and the presence of inserts was confirmed by colony PCR, using the appropriate programme for a specific product and specific primer sets.

2.3.6.4 Isolation of plasmid DNA from E. coli

Plasmid DNA was isolated using the High Pure™ Plasmid Isolation Kit (Roche) according to the manufacturer's instructions. To isolate plasmid DNA, the following steps were followed:

- Eight mL of overnight *E. coli* cultures were harvested by centrifugation at 17,000 x *g* for 1 min.
- The cell pellet was resuspended in 250 μL of suspension buffer containing RNase and then 250 μL of lysis buffer was added and the cells were mixed gently by inverting 3 to 6 times.
- The lysed solutions were treated with 350 μL of chilled binding buffer, mixed by inverting 3 to 4 times and incubated on ice for 5 minutes and then centrifuged at 17,000 x *g* for 10 min to clarify the lysate.
- The supernatant was transferred onto the provided High Pure Filter tubes inserted into a 2 mL collection tube and centrifuged at 17,000 x *g* for 1 min at room temperature to bind the DNA to the column.

- After discarding the flowthrough, 700 µL of wash buffer was added to the column and the column was centrifuged at 17,000 x *g* for 1 min.
- After discarding the flow-through, the column was again centrifuged at 17,000 x *g* for 1 min.
- The column was then placed into a clean microcentrifuge tube, 100 µL of elution buffer was added and after incubating for 1 min at room temperature, the DNA was eluted by centrifugation at 17,000 x *g* for 1 min.
- To quantity the eluted DNA, an aliquot was seperated by 1% (w/v) agarose gel electrophoresis, and the quantity of DNA was estimated by comparing the relative intensity of the DNA band to the HyperLadder^{R} I DNA ladder (Bioline)

2.3.7 Automatic sequencing of DNA

For sequencing, DNA/plasmid samples were submitted to the DNA Analysis Facility, Massey Genome Service, Massey University for sequencing based on the standard protocol for automated capillary analysis on the ABI3730 DNA Analyzer (Applied Biosystem). Reagents used for sequencing reactions were:

- M13 Forward primer with a concentration of 3.3 pmol (5' CCC AGT CAC GAC GTT GTA AAA CG 3')
- M13 Reverse Primer with a concentration of 3.3 pmol (5' AGC GGA TAA CAA TTT CAC ACA GG 3')
- A gene specific forward or reverse primer, as appropriate (3.2 pmol)
- ABI PRISM^{R} BigDye^{TM} Terminator Version 3.1 (Applied Biosystems, Life Technologies)
- Template (plasmid; 300-500 ng)
- Sequencing buffer

Reaction setup:

Reagents	**Concentration/volume**
M13/Gene specific Forward/Reverse primer	3.2 pmol
Plasmid DNA	300-500 ng
Sequencing buffer	5 µL
Big dye terminator	2 µL
Sterile water	Upto 20 µL
Total volume	20 µL

PCR programme:

Steps	Temperature	Time	Cycle
1	96^0C	10 sec	
2	50^0C	5 sec	25
3	60^0C	2 min	
Hold	4^0C		

For cleaning up of the PCR reactions, the *X-terminator® s*ystem was used to remove unincorporated fluorescent ddNTPs and salts from the sequencing reactions following the protocol provided with the Big Dye® X-Terminator™ protocol.

2.3.7.1 Sequence analysis

After sequencing, the DNA sequences were first compared to test their similarity using BioEdit and ClustalW programmes. The sequences were then analysed by blast using NCBI, MEROPS and JCVI databases to find homologs from different plant species. The deduced amino acid sequence (obtained using BioEdit Sequence Alignment Editor v5.0.9) for genes of interest were then aligned with their homologs from different species of interest using the ClustalW programme to generate the multiple sequence alignment data.

2.4 Generation of transgenic plants

2.4.1 Developing RNAi knockdown lines in white clover

The non-overlapping sequences in the coding region of the *Tr-KPI* genes were selected for making the RNAi constructs. RNAi constructs were assembled in the pRNA69 vector and then cloned into the binary vector phZBar which were supplied by Dr. Kim Richardson, AgResearch Grasslands. The primers designed to amplify the gene specific fragment from four *Tr-KPI* genes were tailed with *XhoI* and *BamHI* at the 5' end of the forward primer for the sense and antisense strand, respectively, and the reverse primers were tailed with *EcoRI* and *XbaI* to the 5' end for the the sense and anti sense strand, respectively (Appendix 5). The amplified PCR fragments were then digested with the corresponding digestive enzymes using standard procedures and cloned into the sense (*XhoI/EcoRI*) and anti-sense (*BamHI/XbaI*) sites of the pRNA69 vector. After confirming by sequencing, the sense and anti-sense cassette that includes the CaMV 35S promoter and OCS terminator were digested by the *NotI* enzyme and cloned in the backbone of phZBar binary vector harbouring the spectinomycin resistant gene. The cassette was transformed in *Agrobacterium* strain GV3101 by electroporation using standard protocols:

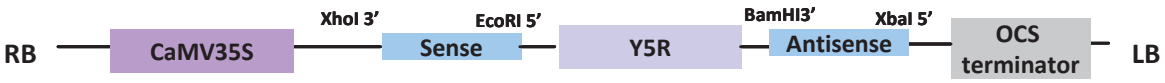


Figure 2.1: Schematic representation of *Tr-KPI* RNai construct in pRNA69 vector

2.4.1.1 Growing Agrobacterium for transformation of RNAi constructs in white clover

Genetic constructs were transformed into white clover at AgResearch, Grassland following the procedure of Voisey *et al.,* (1994) by Kim Richardson and his team where the following steps were used:

- Seeds were weighed to provide approximately 400-500 cotyledons (i.e. 200- 250 seeds) for dissection (0.06 g = 100 seeds). For sterilization, the seeds were rinsed with 70% (v/v) ethanol for 1 min, followed by surface sterilization in bleach (5% (v/v) available chlorine) by shaking on a circular mixer for 15 min followed by four washes in sterile water. The seeds were imbibed overnight at 4°C.
- The genetic constructs, harboured in *Agrobacterium* strain GV3101, were inoculated into 25 mL of MGL broth [mannitol (5.0 g/L), L-glutamic acid (1.0 g/L), KH_2PO_4 (250 mg/L), $MgSO_4$ (100 mg/L), NaCl (100 mg/L), biotin (100 mg/L), bactotryptone (5.0 g/L), and yeast extract (2.5 g/L), pH7.0] containing spectinomycin as a selection marker at a concentration of 100 mg/L.
- Cultures were grown overnight (16 h) on a shaker at 28°C, harvested by centrifugation (3000 x *g*, 10 min), the supernatant removed and the cells were resuspended in 5 mL of 10 mM $MgSO_4$.

2.4.1.2 Dissection and transformation of cotyledonary explants

After removing the seed coat and endosperm, the cotyledons are dissected from seeds using a dissecting microscope and separated from the radical with the scalpel by placing the blade between the cotyledons and slicing through the remaining embryonic axis. Cotyledonary explants are harvested onto a sterile filter disk on CR7 media (Appendix 9). For transformation, a 3 µl aliquot of *Agrobacterium* suspension is dispensed onto each dissected cotyledon. Plates are sealed and cultured at 25°C under a 16 h photoperiod. Following a 72 h period of co-cultivation, transformed cotyledons are transferred to plates containing CR7 medium

supplemented with ammonium glufosinate (2.5 mg/L) and timentin (300 mg/L) and returned to the culture room.

Following the regeneration of shoots, explants were transferred to CR5 medium supplemented with ammonium glufosinate (2.5 mg/L) and timentin (300 mg/L). Regenerating shoots were subcultured at three weekly intervals to fresh CR5 media containing selection. As root formation occurs, plantlets were transferred into tubs containing CR0 medium containing ammonium glufosinate selection. Large clumps of regenerants are divided to individual plantlets at this stage. Media composition used for tissue culture is provided in Appendix 9. Whole, rooted plants growing under selection were then potted into sterile peat plugs. Once established in peat plug, plants were transferred in soil (Daltons Potting Mix) and were grown in the glasshouse. After establishment in soil, leaf and root materials were harvested for positive RNAi construct where primers were designed to OCS3' to give a 439 bp fragment (iYAB5-1 and ocs3'-1, Appendix 5). RNA was extracted from the positive transformed lines and the level of expression of each *Tr-KPIs* were checked by qRT-PCR (see 2.3.4.4).

2.5 Statistical analysis

Statistical analysis was performed using SPSS 11.5 for Windows to perform student's T-Tests. All the graphs presented in the thesis were prepared using Windows Excel. Multiple sequence alignment was generated using ClustalW. A Neighbour joining phylogenetic tree was built using Geneious Pro 5.4.3. The putative transcription factor binding sites (TFBS) in the promoter sequences were analyzed with MatInspector (Cartharius *et al.,* 2005). The relative transcript abundance for qRT-PCR data were analysed following the formula of Pfaffl (2001).

Chapter 3
Results

3.1 Identification of Kunitz proteinase inhibitor genes in white clover

White clover (*Trifolium repens*) is an allotetraploid (2*n* = 4*x* = 32) outcrossing plant species (Atwood and Hill 1940) and as reported by Ellison *et al*., (2006), the two diploid ancestors are proposed to be *T. occidentale* (Schreb.) and *T. pallescens* (Schreb.) which were recognized using phylogenetic tools based on nuclear ribosomal DNA ITS (internal transcribed spacer) and chloroplast *trnL* intron sequences. Therefore, a single genotype (a plant coming from a single germinated seed) was selected randomly and used for the identification of *Kunitz proteinase inhibitor* (*KPI*) genes, and this same genotype was used in most experiments by means of vegetative propagation.

To search for *KPI* genes in white clover, two approaches were undertaken. The first involved screening the AgResearch Grasslands, Palmerston North, New Zealand white clover EST database, while the second involved designing degenerate primers, based on *KPI* sequences expressed in other legume species, and then obtaining putative *KPI* sequences using PCR.

3.1.1 Approach I: *Searching the AgResearch EST database*

To identify *KPI* genes, the AgResearch white clover EST database was screened using a *Medicago truncatula KPI* sequence (AF526372.1) by tblastn. Among legumes*, M. truncatula,* which has a small genome of approximately 500 million base pairs, has been widely adopted as a model for genomic research. It was found previously by Donald Hunter (1999) that *ACC* oxidase genes of white clover share 90 to 95 % identity with the 3' UTR region of *M. truncatula ACC* oxidase genes. Therefore, initially a *M. truncatula KPI* sequence (AF526372.1) was used to virtually probe for KPI genes in the white clover EST database.

The AgResearch database gave hits to about 27 sequences (Appendix 2). The first group comprised seven full length sequences and showed 80% identity with the *M. truncatula* sequence, with minor variations at the translational level. A ClustalW alignment of the white clover sequences showed 100% identity (Appendix 2; Group A) indicating that they are the same protein. The second group (Appendix 2; Group B) comprised five partial sequences and showed 50% identity with the *M. truncatula* protein. Alignment of these sequences also showed that they are the same gene. The other hits were either small partial sequences or the reverse complement strand of the first group with minor identities with the *M truncatula KPI* sequence. Therefore, CTR0034079347, from the first group of the seven full length sequence

was selected to be the typical match for the *M. truncatula* sequence as it was full length and this was subsequently designated as *Tr-KPI1* (Figure 3 .1).

3.1.2 Approach II: Using different sets of degenerate primers

To get additional *KPI* members, degenerate primers were designed based on conserved regions of *M. truncatula* (AF526372.1)*, Glycine max* (EU444603.1) and *Cicer arietinum* (AJ276263.3) Kunitz genes found in the NCBI database (Appendix 3; Group 1) and one degenerate primer was designed. The *M. truncatula* database was also searched [http://www.jcvi.org/cgi-bin/gb2/gbrowse/mtruncatula/ *M. truncatula* genome release version 3.0 (Mt3.0)] with '*Kunitz'* as the keyword. The *Medicago* GBrowse gave 45 hits and it was found that *Kunitz* genes are clustered on chromosome 1, 3, 5 and 6 of this plant species. In the database only the genes sitting in chromosome 6 were annotated as '*Kunitz Proteinase Inhibitor'.* Therefore, the genes on chromosome 6 were further investigated, aligned and three further different sets of degenerate primers were designed in the conserved region for fishing out the genes from white clover (Appendix 3, Group 2,3 ,4).

The highlighted regions (Appendix 3) were selected as the best region for all four degenerate primers to avoid primer self-complementarities (i.e., ability to form secondary structures such as hairpins and the tendency to form homo-dimers with itself or hetero-dimers in the reaction). Though the selected regions are not completely conserved in the sense that there are some single nucleotide differences, only this region showed a similar melting temperature for the forward and reverse primers (64.1^0C and 64.3^0C) and a reasonable product size (more than 300kb product).

A single PCR product ranging in size from 336 to 450bp was obtained from cDNA combined from different tissues of white clover such as dry seed, imbibed seed (6 h, 24 h and 48 h), leaf, petiole, stolon and root when using each of the four primers. The purified PCR products were cloned into the pGEM® T Easy vector and transformed into *E. coli* strain DH5α and the inserts were sequenced. At least 25 inserts were sequenced for each group of sequences amplified with each degenerate primer. The ORFs of these cDNA sequences yielded amino acid products of 112 to 140 aa which were selected from six frame translations and these were proposed to be putative *Kunitz Proteinase Inhibitor (KPI)* genes and were named as *Tr-KPI2, Tr-KPI3, Tr-KPI4, Tr-KPI5 Tr-KPI6, Tr-KPI7 and Tr-KPI8* (Figure 3.1A).

(A)

```
Tr-KPI1             DINGNAIFPGGEYYILPALRGPG-GGGVRIGKTGD--LKCPVTVLQDRREVKNGLPVKFT 57
Tr-KPI8             DINGNAIFPGGEYYILPALRGPG-GGGVRIGKTGD--LKCPVTVLQDRREVKNGLPVKFT 57
Tr-KPI6             DIHGTPIFPGGEYYILPALRGPG-GGGVRIGKTGD--LKCPVTILQDRREVKNGLPVKFT 57
Tr-KPI3             DLHGTPIFPGGKYYIFPVSHDDTYGGGLRLAKTGD--SKCEVTALQDDNIVIDNIPVKFS 58
Tr-KPI7             DKHGIPIFPGRKYYIFPVSHDDTYGGGLRLAKTGD--SKCEVTALQDDNIVIDNIPVKFS 58
Tr-KPI2             GYKWQPPYSGGKYYIFPVSHDETYGGGLRLAKTGD--SKCDVTALQDDNIVIDNIPVKFS 58
Tr-KPI4             DLNGNPIFYSTRFYIMPSIFGAA-GGGLKLGETGK--LTCPLTVLQDYSEVINGLQLKFT 57
Tr-KPI5             DKNGNPVVSGKQYFIFPATDNPK-KGGLTLNNVGDDDSKCPVTVLQNN--AITGLPVKFT 57
                    . :  .   . .::*:*   .    **: : :.*.  .* :* **:   .  .: :**:

Tr-KPI1             IPD-ISTGIIFTGTPVE-IEFFKKPNCAKSSKWLVFVDNVIKKACVG 102
Tr-KPI8             IPD-ISTGIIFTGTPVE-IEFFKKPDCAKSSKWLVFVDNVIKKACVG 102
Tr-KPI6             IPD-ISTGIIFTGTPIE-IEFFKKPNCAKSSKWLVFVDNVIKKACVG 102
Tr-KPI3             IPG-ISPGIIFTGTPIE-IEFTKKPSCVESSKWLIFVDDVIQKACVG 103
Tr-KPI7             IPG-ISPGIIFTGTPIE-IEFTKKPSCVESSKWLIFVDDVIQKACVG 103
Tr-KPI2             IPG-ISPGIIFTGTPIE-IEFTKKPSCVESSKWLIFVDDVIQKACVG 103
Tr-KPI4             PPGEIFVDLISTDQPLKGIEFVEKPECAESSKWVVVEDDDFPRPYVG 104
Tr-KPI5             IPQ-TTTDNIVTGTDLD-IEFTEKPDCAESSKWLLVTDDNTQQSYVG 102
                    *     . * *.  :. *** :**.*.:****::. *:   :. **
```

(B)

	Tr-KPI2	Tr-KPI3	Tr-KPI4	Tr-KPI5	Tr-KPI6	Tr-KPI7	Tr-KPI8
Tr-KPI1	63	64	46	45	95	63	99
Tr-KPI2		94	40	44	61	92	74
Tr-KPI3			45	50	68	97	76
Tr-KPI4				40	48	44	60
Tr-KPI5					47	50	65
Tr-KPI6						66	94
Tr-KPI7							63

Figure 3.1: (A) Alignment of Tr-KPI partial amino acid sequences using ClustalW. (B) Identity of partial sequence of theTr-KPIs at the amino acid level determined using the BioEdit Sequence Alignment Editor for Windows (v5.0.9) where the sequences were trimmed to the same length as shown in A.

When using pair-wise sequence alignment of the identified partial amino acid sequences of Tr-KPIs in the present study, the sequence similarity between two entry genes would be notably different if the length of the two sequences were different or covered different regions of the homologue. Therefore, the result of similarities and positive identities over the sequences would presumably be somewhat biased if the reported sequences were included in the analysis without considering their differences in length and covering the complementary regions of the genes in question. For this reason, all sequences isolated in the present study, were trimmed to the same length and to cover the same region of the *KPI* genes before being used for sequence comparison. Pair wise alignment of the deduced amino acid sequence showed that Tr-KPI1 and Tr-KPI6 share 95% and Tr-KPI1 and Tr-KPI8 share 98% identity, Tr-KPI2 and Tr-KPI3 share 94% identity, Tr-KPI2 and Tr-KPI7 share 92% identity and Tr-KPI3 and Tr-KPI7 share 97% identity. Therefore focus was given to obtaining the full length coding sequence, 3' UTR and promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* (Figure 3.1B).

3.1.3 Identification of full length Tr-KPI genes

To obtain the 3'UTR of each of the *Tr-KPI* genes, two primers were designed in the 3' region of the known sequence to provide 60 to 100 bp overlap (Appendix 4). The 5'UTR region and the promoter region were obtained by genome walking using gDNA. The whole sequence for each gene was linked *in silico* based on the overlapping regions of the sequences. The start and the stop codons (highlighted) were identified using the BioEdit (Biological sequence alignment Editor) programme which yielded an ORF of 657, 663, 633 and 678 bp for *TrKPI1, TrKPI2, TrKPI4* and *TrKPI5* respectively (Figure 3.2). To confirm the *in silico* sequence, gene-specific primers starting from the start codon and ending at the stop codon were then designed to amplify the full length sequence. The templates were the cDNA from different tissues such as 6hr imbibed seed, mature leaf, root, mature leaf petiole and mature stolon. A single PCR product of the expected size was obtained from all the tissues screened and the purified PCR products were cloned into the pGEM® T Easy vector, transformed into the *E. coli* strain DH5α and the inserts sequenced.

```
TrKPI2          ATGAAGCCTATGTTATCACTCATCTTTTCTTTCTTCCTCTTTATTTTAATCACCAATCTT 60
TrKPI1          ATGAAGCATGTTTCATCACTCACCCTTTCCATTCTCTTCTTTGTTTCCATCACCAATCTT 60
TrKPI5          ATGAATCCTACTTTATCCATTACCCTTTCCTTCCTCCTATTTGCATTTATCTCCAATCTT 60
TrKPI4          ATGAAACCTACAATGCTTACCACCCTTTCTTTACTCCTCTTTGCCTTAACCACCTACTTT 60
                ***** * *           * * **** *  ** * ***   *  * * ** *  **

TrKPI2          TCACTTGCTTTATCAAATGAAGCTGTTGAGCAAGTATTGGACTCTCTTGGTAACCCCATT 120
TrKPI1          TCACTAGCATTCTCAAATGAAGATGTTGAGCAAGTGCTGGATATAAATGGTAACGCCATC 120
TrKPI5          TCACT---------AAACAATGCAGTTCAGCAAGTATTTGACACAAATGGTAACCCTGTT 111
TrKPI4          CCATTAGCTTT------TAGTTCTAACGAACAACTAGCAGACTTGAATGGAAACCCCATC 114
                 ** *                     * *** *    **      *** *** *  *

Tr-KPI2         TTTCCTGGTGGAAAATACTACATTTTTCCTGTATCTCATGATGAGACATATGGTGGAGGA 180
Tr-KPI1         TTCCCAGGTGGCGAATACTACATTTTACCAGCACTTCGCGGCC---CAGGAGGCGGAGGA 177
Tr-KPI5         GTCTCTGGTAAACAATACTTCATTTTTCCAGCAACTGATAACC---CTAAAAAGGGAGGA 168
Tr-KPI4         TTTTATTCTACTCATTTCTATATTATGCCATCTATCTTTGGAG---CTGCAGGTGGTGGA 171
                 *      *    * * **  *** * **                 *       ** ***

Tr-KPI2         TTAAG-ACTAGCAAAAAC-----CGGCGATTCAAAGTGTGAAGTTACTGCCTTACAAGAT 234
Tr-KPI1         GTAAGGATTGGCAAAA-C-----CGGAGACTTAAAGTGTCCGGTTACTGTCTTACAAGAT 231
Tr-KPI5         CTAACCTTAAACAATGTCGGCGACGATGATTTAAAATGTCCAGTAACTGTCCTACAAAAC 228
Tr-KPI4         CTCAAGCTTGGTGAAACT------GGAAAATTGACATGTCCACTTACTGTACTTCAAGAT 225
                 * *         *          *   * *  *  ***    * ****   * *** *

Tr-KPI2         GACAATATAGTTACTGAAAGTATACCAGTGAAATTCAGTATACCAGGAATAAGTCCTGGT 294
Tr-KPI1         CGTAGAGAAGTCAAAAATGGTTTACCAGTGAAGTTTACCATACCAGATATAAGTACTGGT 291
Tr-KPI5         AAT------GCCATAACAGGTTTACCAGTTAAATTCACCATCCCACAAACCACCACCGAT 282
Tr-KPI4         TATTCTGAAGTTATCAATGGTCTGCAACTAAAATTTACCCCTCCAGGTGAAATTTTCGTT 285
                         *  *      ** * * * * ** ** *     ***      *     * *

Tr-KPI2         ATAATTTTCACTGGTACG--CCGAT---TGAGA----TTGAGTTTACAAAGAAGCCTAGT 345
Tr-KPI1         ATAATCTTCACTGGTACA--CCAGT---TGAGA----TCGAGTTCTTTAAGAAACCTAAT 342
Tr-KPI5         AATATCGTAACAGGTAC----CGATC-TTGACA----TCGAGTTCACCGAGAAACCTGAT 333
Tr-KPI4         GA---TTTGATAAGTACAGACCAACCACTGAAAGGTATTGAATTTGTAGAGAAGCCAGAG 342
                       * *   ****    *      *** *    * ** **     **** **

Tr-KPI2         TGTGTTGAATCATCGAAATGGTTGATATTTGTTGACGATGTTATTCAAAAAGCTTGTGTT 405
Tr-KPI1         TGTGCTAAATCATCGAAATGGTTGGTATTCGTTGATAATGTTGTAAAAAAAGCTTGTGTT 402
Tr-KPI5         TGTGCTGAATCATCAAAATGGTTACTAGTTACTGATGATAATACTCAACAAAGCTATGTT 393
Tr-KPI4         TGTGCTGAATCCTCCAAGTGGGTGGTGGTCGAAGACGATGATTTCCCTCGACCATATGTC 402
                **** * **** ** ** *** *  *  *    **  **  *        *   * ***

Tr-KPI2         GGTATAGGTGGTCCTGAAAATTATCCTCATTTTAAAACATTGAATGATGGTAGATTTTAT 465
Tr-KPI1         GGTATTGGTGGTCCTGAAAATTATCCTGGTGTGCAAACATTGAGTG---GAACATTTAAT 459
Tr-KPI5         GGTATTGGTGGACCTGCAAATTATCCTGGTGTAGAATTAATCAGTG---GAAAATTTTTA 450
Tr-KPI4         GGAATTGGTGGTATTGAAGACAATAAAGGTGAGAGGATCATAAATG---GTAGCTTTAAA 459
                ** ** *****   ** * *  **     *          * * **   * *  ***

Tr-KPI2         ATTGAGAAACATGAATCTGGATTTGGTTATAAGCTTGGATATTGT-----GTGAAAGATT 520
Tr-KPI1         ATTCATAAACATGAATCTGGATTTGGTTATAAACTTGGATTTTGC-----ATTAAGGGTT 514
Tr-KPI5         GTTGTGAAACATGGAACTGGTGGTAGTTATAAGATTGGATTTTGT-----TTGGATAGTA 505
Tr-KPI4         ATTGTGAAACATGGTTTTGGA------TACAAGATTGTGTTTTGTCCTCGATTCACTGCA 513
                 **   *******    ***       ** **  ***  * ***       *  *

Tr-KPI2         C----TCCTACTTGT-------TTGGATATTGGGAG-ATCTGGTAATGTGACGGAGGAAG 568
Tr-KPI1         C----ACCTACTTGT-------TTGGATATTGGGAG-GTATGACAATG------ATGAAG 556
Tr-KPI5         C----TGGTGATTGTGGTTATCTTGGAT--TACAAGTGTTTAATTCCG------GAGAAG 553
Tr-KPI4         CCACCTGGTCTTTGT-------TTTGATATTGGAAG-GCATGATGATG-------AGAA- 557
                *       *  ****       ** ***  *   **    *      *       ***
```

Continued

```
Tr-KPI2         GTGGGTTTCGTTTGAATTTGACTCATCAAGTGGCTTATGCGGTTGAGTTTATTGATGTTA 628
Tr-KPI1         CTGGAAAACGTTTGAATTTGACTGAACATGAGTCTTATCATGTTATCTTTGTTGATGCTG 616
Tr-KPI5         GTGGTTCGCGTTTGATTTTAACTATAACTGATGCTTATTCTGTTGTGTTTGTTGATTCTG 613
Tr-KPI4         -TGGAAGGCGTCTCATCCTCACTGAAAATGATCCTTTTGAAATTGTCTTTGT-GATTC-- 613
                ***    *** * *   * ***      *   *** *    **   *** * ***

Tr-KPI2         CTCCTTTTGAAGCTAGA------ATTAAGTCTGTTG---------------------- 658
Tr-KPI1         CTTCTCATGAAGCTGATCAGTATATTAAGTCTGTTG---------------------- 652
Tr-KPI5         CTTCTGTTAAATCTGAA---AATATTGTGTCTGCTGAAGGTAATCCAAATATTGCATTGC 670
Tr-KPI4         ------------CTAGA---------AGATCTGTTG---------------------- 628
                           **               **** **

Tr-KPI2         ---TTTGA 663
Tr-KPI1         ---TTTAA 657
Tr-KPI5         CTATTTAA 678
Tr-KPI4         ---CTTGA 633
                   ** *
```

Figure 3. 2: Alignment of identified full length *Tr-KPI* genes. '*' denotes that the nucleotides in that column are identical in all sequences in the alignment.

The cDNA of the identified *Tr-KPIs* were 633 to 678 bp long. The translation of the ORFs of these cDNA sequences yielded amino acid sequences of 210 to 225 aa (Figure 3. 3 A). The predicted signal peptide is 23 aa long for Tr-KPI1 and Tr-KPI2 and 25 aa long for Tr-KPI4 and Tr-KPI5 which was identified using the WoLF PSORT Prediction as described by Nakai and Kanehisha (1991) and SignalIP 4.0 prediction Server (Peterson *et al.,* 2011). Without the signal peptide at the N terminus, the mature proteins comprised 195, 197, 181 and 200 amino acids for Tr-KPI1, Tr-KPI2, Tr-KPI4 and Tr-KPI5, respectively, and the theoretical masses are presented in Figure 3.3 B and C.

(A)

```
TrKPI1          MKHVSSLTLSILFFVSITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPG-GGG 59
TrKPI2          MKPMLSLIFSFFLFILITNLSLALSNEAVEQVLDSLGNPIFPGGKYYIFPVSHDETYGGG 60
TrKPI5          MNPTLSITLSFLLFAFISNLSL---NNAVQQVFDTNGNPVVSGKQYFIFPATDNPK-KGG 56
TrKPI4          MKPTMLTTLSLLLFALTTYFPLAFSSN--EQLADLNGNPIFYSTHFYIMPSIFGAA-GGG 57
                *:      :*:::*   : :.*   .:  :*: *  **.:. . .::*:*   .    **

TrKPI1          VRIGKTGD--LKCPVTVLQDRREVKNGLPVKFTIP-DISTGIIFTGTPVE-IEFFKKPNC 115
TrKPI2          LRLAKTGD--SKCEVTALQDDNIVTESIPVKFSIP-GISPGIIFTGTPIE-IEFTKKPSC 116
TrKPI5          LTLNNVGDDDLKCPVTVLQNN--AITGLPVKFTIP-QTTTDNIVTGTDLD-IEFTEKPDC 112
TrKPI4          LKLGETGK--LTCPLTVLQDYSEVINGLQLKFTPPGEIFVDLISTDQPLKGIEFVEKPEC 115
                : : :.*.   .* :*.**:   .  .: :**: *     . * *.  :. *** :**.*

TrKPI1          AKSSKWLVFADNVVKKAWVGIGGPENYPGVQTLS-GTFNIHKHESGFGYKLGFCIKGSPT 174
TrKPI2          VESSKWLIFVDDVIQKAWVGIGGPENYPHFKTLNDGRFYIEKHESGFGYKLGYCVKDSPT 176
TrKPI5          AESSKWLLVTDDNTQQSYVGIGGPANYPGVELIS-GKFLVVKHGTGGSYKIGFCLDSTGD 171
TrKPI4          AESSKWVVVEDDDFPRPYVGIGGIEDNKGERIIN-GSFKIVKHG--FGYKIVFCPRFTAP 172
                .:****::. *:   :. ***** :    . :. * * : **    .**: :*   :

TrKPI1          ---CLDIGRYDN--DEAGKRLNLTEHESYHVIFVDAASHEADQYIKSVV-------- 218
TrKPI2          ---CLDIGRSGNVTEEGGFRLNLTHQVAYAVEFIDVTPFEAR--IKSVV-------- 220
TrKPI5          ---CGYLGLQVFNSGEGGSRLILTITDAYSVVFVDSASVKSENIVSAEGNPNIALPI 225
TrKPI4          PGLCFDIGRHDD---ENGRRLILTENDPFEIVFVIPRRSVA---------------- 210
                   *  :*       * * ** **   .: : *:     :
```

(B)

Nucleotide Level ↓ / Amino Acid Level →	TrKPI1	TrKPI2	TrKPI4	TrKPI5
TrKPI1		64	45	44
TrKPI2	*74*		41	43
TrKPI4	*57*	*56*		40
TrKPI5	*66*	*63*	*54*	

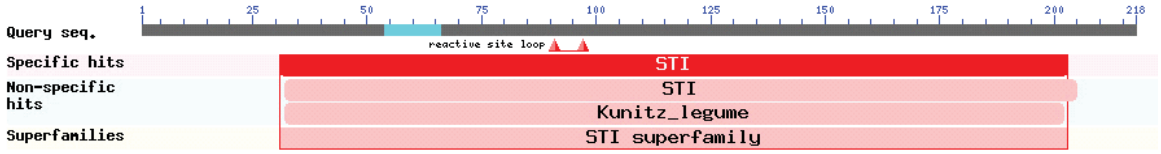
C)

	pI	Mw
TrKPI1	6.23	21.22
TrKPI2	5.01	21.56
TrKPI4	4.90	20.49
TrKPI5	4.35	21.12

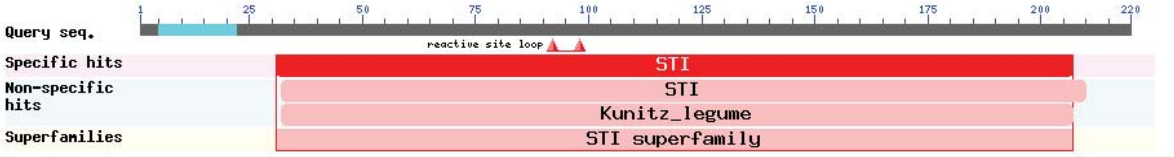
Figure 3.3: (A) Alignment of Tr-KPIs deduced amino acid sequences. The yellow box indicates conserved cystine residues, the purple indicates additional cystine residues, green indicates the signal peptide and grey denotes the reactive site, '*' denotes that the amino acid in that column are identical in all sequences in the alignment. ':' denotes that conserved substitutions have been observed and '.' indicating semi-conserved substitutions. **(B)** Identity of Tr-KPI protein at the nucleotide and amino acid level. **(C)** Theoretical pI/Mw (averages) obtained using the ExPasy Compute MW tool (http://web.expasy.org/compute_pi/v)

The highest sequence similarity of the identified *Tr-KPI*s was 74% between *Tr-KPI1* and *Tr-KPI2* and the lowest is 54% between *Tr-KPI4* and *Tr-KPI5* at the nucleotide level and 64% and 40%, respectively, at the amino acid level along the full length sequence. BLAST searching of the translated amino acid sequence against the GenBank database and MEROPs database gave specific hits to the Soybean Trypsin Inhibitor family and the sequences showed that these full length genes, when translated, comprises sequence characteristics typical of other KPI proteins from different plant species (Figure 3.4).

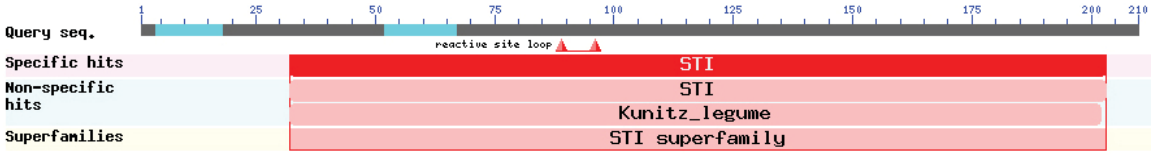
(A) Tr-KPI1



(B) Tr-KPI2



(C) Tr-KPI4



(D) Tr-KPI5

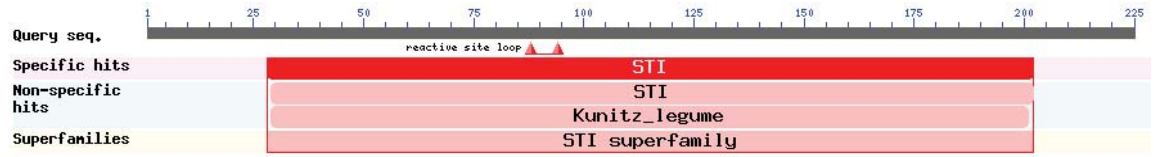
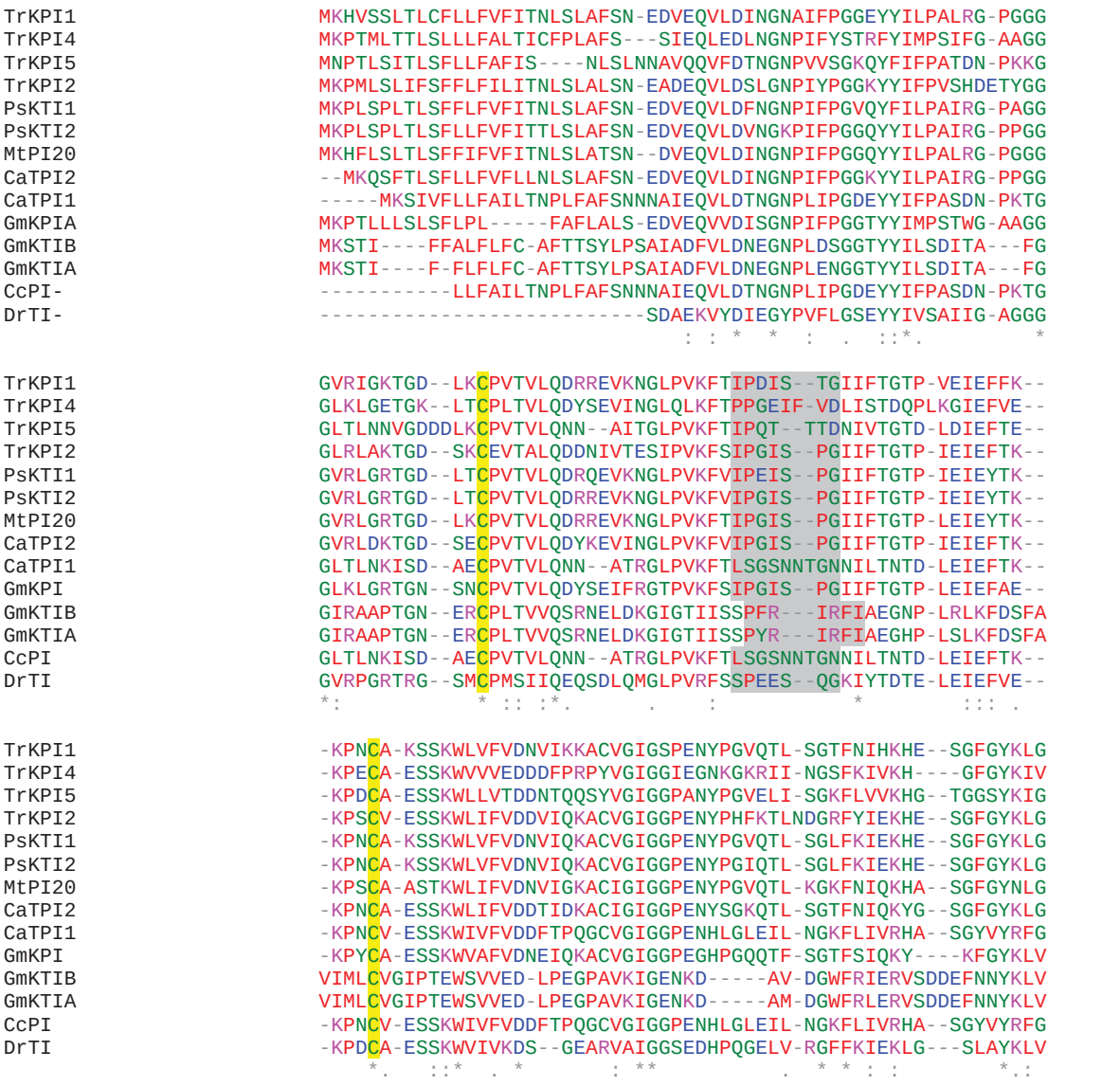


Figure 3. 4: Reactive site loop for Tr-KPI1 (A), Tr-KPI2 (2), Tr-KPI4 (C) and Tr-KPI5 (D) as determined using the BLAST search tool available on the NCBI database (http://blast.ncbi.nlm.nih.gov/)

3.1.3 Tr-KPI genes share sequence similarity with KPI genes from other plant species

Multiple sequence alignment of the Tr-KPI proteins with some representative Kunitz proteinase inhibitors from the legume family showed that sequence variation occurred randomly along the 210 to 225aa sequences (Figure 3.5). The first 25 to 26 sequences may have varied because of the presence of the signal peptide that directs the transport of the protein to different cellular organelles. Therefore the sub-cellular localization predictor programme (Emanuelson *et al.,* 2000) predicted that the Tr-KPI proteins are targeted to the secretory pathway in common with MT, CaTPI1, CaTPI2, GM1, PsKTI1, PsKTI2 and CJ (Table 3.1). Closer observation of the aligned region of the signal peptide from different species shows that the TrKPIs, MT, PSKTI1, PSKTI2, CaTPI2 and GM1 are reasonably conserved (Figure 3.5). The ClustalW alignment programme failed to show conserved amino acids in this region because of the missing amino acids from the CaTPI1 and CJ sequence.

The alignment of these full length Tr-KPI proteins with other Trypsin Proteinase Inhibitor proteins from different plant species showed that the white clover KPIs share four conserved cysteine residues with Kunitz Trypsin Inhibitor Proteins from other plant species (Figure 3.5). Most members of the KPI family possess four cysteine residues to form the disulfide bonds (Cys63-Cys110 and Cys160-Cys169). This is supporting evidence that the white clover proteins are also putative KPI proteins. Tr-KPI1 and Tr-KPI2 have one extra cysteine residue sitting outside the first disulfide bond which is conserved in the *Pisum sativum* (GenBank-082771.1, GenBank: NP_001238220.1), *M. truncatula* (GenBank: AAM88404.1), *C. areitinum* (GenBank: AAT45474.1, GenBank: AAT45474.1) and *Cajanus cajun* (GenBank: NP_001238611.1) inhibitor proteins. Although some conserved residues are observed with other KPI proteins in the reactive site loop, this region is found to be highly variable when compared with the soybean KTI (Figure 3.5). The predicted reactive site is devoid of Arg or Lys which is the trypsin active site identified in soybean STI (Oliva *et al.,* 2011)


TrKPI1 MKHVSSLTLCFLLFVFITNLSLAFSN-EDVEQVLDINGNAIFPGGEYYILPALRG-PGGG
TrKPI4 MKPTMLTTLSLLLFALTICFPLAFS---SIEQLEDLNGNPIFYSTRFYIMPSIFG-AAGG
TrKPI5 MNPTLSITLSFLLFAFIS----NLSLNNAVQQVFDTNGNPVVSGKQYFIFPATDN-PKKG
TrKPI2 MKPMLSLIFSFFLFILITNLSLALSN-EADEQVLDSLGNPIYPGGKYYIFPVSHDETYGG
PsKTI1 MKPLSPLTLSFFLFVFITNLSLAFSN-EDVEQVLDFNGNPIFPGVQYFILPAIRG-PAGG
PsKTI2 MKPLSPLTLSFLLFVFITTLSLAFSN-EDVEQVLDVNGKPIFPGGQYYILPAIRG-PPGG
MtPI20 MKHFLSLTLSFFIFVFITNLSLATSN--DVEQVLDINGNPIFPGGQYYILPALRG-PGGG
CaTPI2 --MKQSFTLSFLLFVFLLNLSLAFSN-EDVEQVLDINGNPIFPGGKYYILPAIRG-PPGG
CaTPI1 -----MKSIVFLLFAILTNPLFAFSNNNAIEQVLDTNGNPLIPGDEYYIFPASDN-PKTG
GmKPIA MKPTLLLSLSFLPL-----FAFLALS-EDVEQVVDISGNPIFPGGTYYIMPSTWG-AAGG
GmKTIB MKSTI----FFALFLFC-AFTTSYLPSAIADFVLDNEGNPLDSGGTYYILSDITA---FG
GmKTIA MKSTI----F-FLFLFC-AFTTSYLPSAIADFVLDNEGNPLENGGTYYILSDITA---FG
CcPI- -----LLFAILTNPLFAFSNNNAIEQVLDTNGNPLIPGDEYYIFPASDN-PKTG
DrTI- -----SDAEKVYDIEGYPVFLGSEYYIVSAIIG-AGGG
TrKPI1 GVRIGKTGD--LKCPVTVLQDRREVKNGLPVKFTIPDIS--TGIIFTGTP-VEIEFFK--
TrKPI4 GLKLGETGK--LTCPLTVLQDYSEVINGLQLKFTPPGEIF-VDLISTDQPLKGIEFVE--
TrKPI5 GLTLNNVGDDDLKCPVTVLQNN--AITGLPVKFTIPQT--TTDNIVTGTD-LDIEFTE--
TrKPI2 GLRLAKTGD--SKCEVTALQDDNIVTESIPVKFSIPGIS--PGIIFTGTP-IEIEFTK--
PsKTI1 GVRLGRTGD--LTCPVTVLQDRQEVKNGLPVKFVIPEIS--PGIIFTGTP-IEIEYTK--
PsKTI2 GVRLGRTGD--LTCPVTVLQDRREVKNGLPVKFVIPGIS--PGIIFTGTP-IEIEYTK--
MtPI20 GVRLGRTGD--LKCPVTVLQDRREVKNGLPVKFTIPGIS--PGIIFTGTP-LEIEYTK--
CaTPI2 GVRLDKTGD--SECPVTVLQDYKEVINGLPVKFVIPGIS--PGIIFTGTP-IEIEFTK--
CaTPI1 GLTLNKISD--AECPVTVLQNN--ATRGLPVKFTLSGSNNTGNNILTNTD-LEIEFTK--
GmKPI GLKLGRTGN--SNCPVTVLQDYSEIFRGTPVKFSIPGIS--PGIIFTGTP-LEIEFAE--
GmKTIB GIRAAPTGN--ERCPLTVVQSRNELDKGIGTIISSPFR---IRFIAEGNP-LRLKFDSFA
GmKTIA GIRAAPTGN--ERCPLTVVQSRNELDKGIGTIISSPYR---IRFIAEGHP-LSLKFDSFA
CcPI GLTLNKISD--AECPVTVLQNN--ATRGLPVKFTLSGSNNTGNNILTNTD-LEIEFTK--
DrTI GVRPGRTRG--SMCPMSIIQEQSDLQMGLPVRFSSPEES--QGKIYTDTE-LEIEFVE--
TrKPI1 -KPNCA-KSSKWLVFVDNVIKKACVGIGSPENYPGVQTL-SGTFNIHKHE--SGFGYKLG
TrKPI4 -KPECA-ESSKWVVVEDDDFPRPYVGIGGIEGNKGKRII-NGSFKIVKH----GFGYKIV
TrKPI5 -KPDCA-ESSKWLLVTDDNTQQSYVGIGGPANYPGVELI-SGKFLVVKHG--TGGSYKIG
TrKPI2 -KPSCV-ESSKWLIFVDDVIQKACVGIGGPENYPHFKTLNDGRFYIEKHE--SGFGYKLG
PsKTI1 -KPNCA-KSSKWLVFVDNVIQKACVGIGGPENYPGVQTL-SGLFKIEKHE--SGFGYKLG
PsKTI2 -KPNCA-KSSKWLVFVDNVIQKACVGIGGPENYPGIQTL-SGLFKIEKHE--SGFGYKLG
MtPI20 -KPSCA-ASTKWLIFVDNVIGKACIGIGGPENYPGVQTL-KGKFNIQKHA--SGFGYNLG
CaTPI2 -KPNCA-ESSKWLIFVDDTIDKACIGIGGPENYSGKQTL-SGTFNIQKYG--SGFGYKLG
CaTPI1 -KPNCV-ESSKWIVFVDDFTPQGCVGIGGPENHLGLEIL-NGKFLIVRHA--SGYVYRFG
GmKPI -KPYCA-ESSKWVAFVDNEIQKACVGIGGPEGHPGQQTF-SGTFSIQKY----KFGYKLV
GmKTIB VIMLCVGIPTEWSVVED-LPEGPAVKIGENKD-----AV-DGWFRIERVSDDEFNNYKLV
GmKTIA VIMLCVGIPTEWSVVED-LPEGPAVKIGENKD-----AM-DGWFRLERVSDDEFNNYKLV
CcPI -KPNCV-ESSKWIVFVDDFTPQGCVGIGGPENHLGLEIL-NGKFLIVRHA--SGYVYRFG
DrTI -KPDCA-ESSKWVIVKDS--GEARVAIGGSEDHPQGELV-RGFFKIEKLG---SLAYKLV

Continued

```
TrKPI1                  FCIK---GSPTCLDIGRYDN--DEAGKRLNLTEHESYHVIFVDAASHEADQYIKSVV---
TrKPI4                  FCPRFTAPPGLCFDIGRHD---DENGMRLILTENDPFEIVFEIPRRSVA-----------
TrKPI5                  FCLD---STGDCGYLGLQVFNSGEGGSRLILTITDAYSVVFVDSASVKSE-NIVSAEGNP
TrKPI2                  YCVK---DSPTCLDIGRSGNVTEEGGFRLNLTHQVAYAVEFIDVTPFEA--RIKSVV---
PsKTI1                  FCVK---GSPTCLDVGRFDN--DEDGRRLNLTEHESFQVVFIQAEANDA-EFIKSVV---
PsKTI2                  FCIK---GSPTCLDVGRFDN--DEAGRRLNLTEHESFQVVFVEAEANDA-EFIKSVV---
MtPI20                  FCVT---GSPTCLDIGRFDN--DEAGRRLNLTEHEVYQVVFVDAATYEA-EYIKSVV---
CaTPI2                  FCVK---GSPICLDIGRYDN--DEGGRRLNLTEHEAFRVVFVDASSYED-GIVKSVV---
CaTPI1                  FCLD---VSGDCGLLGLNTFDSREGGSRLILTIFNSYNVVFVDVASVKSG-RIMPLKGGF
GmKPI                   FCIT---GSGTCLDIGRFDAKNGEGGRRLNLTEHEAFDIVFIEASKV--DGIIKSVV---
GmKTIB                  FCTQ-QAEDDKCGDIGISID-HDDGTRRLVVSKNKPLVVQFQKVDKESLAKKNH---G-L
GmKTIA                  FCPQ-QAEDDKCGDIGISID-HDDGTRRLVVSKNKPLVVQFQKLDKESLAKKNH---G-L
CcPI                    FCLD---VSGDCGLLGLNTFDSREGGSRLILTIFNSYNVVFVDVASVKSG-RIMPLKGGF
DrTI                    FCPKSKSSSGSCSDIGINYE--GRRSLVLKSSDDSPFRVVFVKPRSGSETES--------
                        :*         *  :*            *  :      : *

TrKPI1                  ------
TrKPI4                  ------
TrKPI5                  NIALPI
TrKPI2                  ------
PsKTI1                  ------
PsKTI2                  ------
MtPI20                  ------
CaTPI2                  ------
CaTPI1                  DASV--
GmKPI                   ------
GmKTIB                  SRSE--
GmKTIA                  SRSE--
CcPI                    DASV--
DrTI                    ------
```

Figure 3.5: Alignment of Tr-KPI deduced amino acid sequences with other Trypsin Inhibitors using ClustalW. MT (*M. truncatula* GenBank: AAM88404.1), CaTPI1 (*C. arietinum* kunitz trypsin inhibitor 1 GenBank: AAT45474.1), CaTPI2 (*C. arietinum* kunitz trypsin inhibitor 2 GenBank: AAT45474.1), GmKPI (*G. max* 1 GenBank: ACA23207.1), GmKTIA (*G. max* GenBank: AAF87095.1) and CJ (GenBank: ADB44827.1), CcPI (*C. cajan* GenBank: NP_001238611.1); GmKTIb (*G. max* GenBank: NP_001238220.1) DrTI(*Delonix rigia* GenBAnk: AAY84867.1); PsKTI1(*P. sativum* GenBank-Q41015.2) and PsKTI2 (*P. sativum* GenBAnk-082771.1).The yellow box indicates conserved cystine residues, grey box indicates reactive site; "*" denotes that the amino acid in that column are identical in all sequences in the alignment. ":" denotes that conserved substitutions have been observed and "." indicating semi-conserved substitutions.

Table 3.1: Predicted localization of Tr-KPI proteins from different plant species[1].

Protein	cTP	mTP	SP	Others	Loc	RC
Tr-KPI1	0.004	0.018	0.997	0.041	S	1
Tr-KPI2	0.003	0.011	0.991	0.047	S	1
Tr-KPI4	0.009	0.008	0.995	0.034	S	1
Tr-KPI5	0.005	0.014	0.997	0.036	S	1
PsKTI1	0.006	0.011	0.996	0.037	S	1
PsKTI2	0.005	0.013	0.996	0.039	S	1
MtPI20	0.007	0.011	0.995	0.040	S	1
CaTPI1	0.004	0.015	0.996	0.037	S	1
CaTPI2	0.011	0.017	0.965	0.095	S	1
GmKPI	0.012	0.014	0.990	0.058	S	1
GmKTIA	0.013	0.007	0.988	0.081	S	1
GmKTIB	0.016	0.009	0.987	0.072	S	1
CcPI	0.148	0.032	0.452	0.320	S	5
DrTI	0.116	0.048	0.150	0.799	-	2

[1] The predictions were undertaken using TargetP 1.1 Server (Emanuelson *et al.,* 2000) where the location assignment is based on the predicted presence of any N-terminal presequences. cTP: sequence contains a chloroplast transit peptide; mTP: Sequence contain a mitochondrial targeting peptide; SP: Secretory Pathway i.e sequence contain a signal peptide; '-': Any other location; Loc: Prediction of localization; S: secretory pathway; RC: Reliability Class which is the measure of the size of the difference between the highest and second highest output scores, where 1 indicates strongest prediction. There are five RCs, defined as follows:

1. difference >0.800
2. 0.800>difference>0.600
3. 0.600>difference>0.400
4. 0.400>difference>0.200
5. 0.200>difference

(http://www.cbs.dtu.dk/services/TargetP/)

3.1.4 Structure of Kunitz proteinase inhibitor gene family in Medicago truncatula

BLAST searching for similar sequences using the JCVI *M. truncatula* Genome Annotation (updated version 3.5; http://www.jcvi.org/cgi-bin/medicago/annotation.cgi?page=pseudo_annot) for Tr-KPIs yielded only Kunitz proteinase inhibitors in this model species, further supporting the proposition that these proteins do belong to the KPI family. The highest sequence similarity for Tr-KPI1 was 76% (Table 3.2), Tr-KPI2 was 73% (Table 3.3), Tr-KPI4 was 63% (Table 3.4) and Tr-KPI5 was 70% (Table 3.5) while the lowest being 37%, 46% 38% and 38% respectively. The *Medicago* database (Mt3.5) gave 27 matches when the GBrowse was searched with the keyword '*Kunitz*'. It was found that out of 27 matches, 13 of these were not annotated and/or partial sequences. Therefore, the remaining 14 annotated full length genes were used for further phylogenetic analysis.

The phylogenetic analysis revealed that Tr-KPI4 followed by Tr-KPI5 are more distantly related and formed separate clades with *Medicago* proteins whereas Tr-KPI1 and Tr-KPI2 are sitting in the same clade (Figure 3.6).

Table 3.2: Result of JCVI BLAST search using Tr-KPI1 aa sequences

Accession No	Gene	Identity (%)	Positive (%)	E value
IMGA|Medtr6g059730.1	Kunitz-type trypsin inhibitor	76	83	1.7e-87
IMGA|Medtr6g059680.1	Kunitz type Trypsin Inhibitor	67	78	2.9e-76
IMGA|Medtr6g009650.1	Kunitz type Trypsin Inhibitor	48	64	6.3e-42
IMGA|Medtr6g059810.1	Kunitz type Trypsin Inhibitor	46	58	6.9e-36
IMGA|Medtr6g059650.1	Kunitz type Trypsin Inhibitor	45	55	1.5e-46
IMGA|Medtr6g059530.1	Kunitz type Trypsin Inhibitor	45	58	2.3e-35
IMGA|Medtr6g065460.1	Kunitz type Trypsin Inhibitor	43	57	5.2e-34
IMGA|Medtr6g059410.1	Kunitz type Trypsin Inhibitor	44	57	1.2e-33
IMGA|Medtr6g059760.1	Kunitz type Trypsin Inhibitor	37	51	21.4e-19
IMGA|Medtr6g078070.1	Kunitz type Trypsin Inhibitor	30	49	4.9e-19
IMGA|Medtr6g078100.1	Kunitz type Trypsin Inhibitor	32	45	1.3e-18
IMGA|Medtr6g044810.1	Kunitz type Trypsin Inhibitor	35	46	3.4e-18
IMGA|Medtr6g059780.1	Kunitz type Trypsin Inhibitor	42	54	5.6e-18
IMGA|Medtr6g044780.1	Kunitz type Trypsin Inhibitor	37	44	1.5e-17

Table 3.3: Result of JCVI BLAST search using Tr-KPI2 aa sequences

Accession No	Gene	Identity (%)	Positive (%)	E value
IMGA|Medtr6g059680.1	Kunitz type Trypsin Inhibitor	73	81	1.0e-73
IMGA|Medtr6g059730.1	Kunitz type Trypsin Inhibitor	65	74	2.2e-64
IMGA|Medtr6g009650.1	Kunitz type Trypsin Inhibitor	50	65	1.9e-42
IMGA|Medtr6g059810.1	Kunitz type Trypsin Inhibitor	51	64	1.6e-38
IMGA|Medtr6g059530.1	Kunitz type Trypsin Inhibitor	46	59	1.9e-33
IMGA|Medtr6g059410.1	Kunitz type Trypsin Inhibitor	45	58	1.7e-32
IMGA|Medtr6g059650.1	Kunitz type Trypsin Inhibitor	45	58	2.0e-29
IMGA|Medtr6g065460.1	Kunitz type Trypsin Inhibitor	44	54	9.3e-25
IMGA|Medtr6g059760.1	Kunitz type Trypsin Inhibitor	38	52	2.2e-20
IMGA|Medtr6g044810.1	Kunitz type Trypsin Inhibitor	35	49	1.5e-17
IMGA|Medtr6g044780.1	Kunitz type Trypsin Inhibitor	38	48	6.8e-17
IMGA|Medtr6g059780.1	Kunitz type Trypsin Inhibitor	46	55	8.2e-17

Table 3.4: Result of JCVI BLAST search using Tr-KPI4 aa sequences

Accession No	Gene	Identity (%)	Positive (%)	E value
IMGA|Medtr6g059650.1	Kunitz type Trypsin Inhibitor	62	72	2.7e-66
IMGA|Medtr6g065460.1	Kunitz type Trypsin Inhibitor	53	64	4.9e-41
IMGA|Medtr6g059530.1	Kunitz type Trypsin Inhibitor	48	62	8.3e-40
IMGA|Medtr6g059410.1	Kunitz type Trypsin Inhibitor	47	61	1.2e-38
IMGA|Medtr6g059730.1	Kunitz type Trypsin Inhibitor	45	60	1.3e-34
IMGA|Medtr6g059680.1	Kunitz type Trypsin Inhibitor	44	60	9.1e-34
IMGA|Medtr6g059760.1	Kunitz type Trypsin Inhibitor	44	56	3.7e-30
IMGA|Medtr6g044780.1	Kunitz type Trypsin Inhibitor	38	52	6.8e-29
IMGA|Medtr6g044810.1	Kunitz type Trypsin Inhibitor	34	50	6.3e-26
IMGA|Medtr6g009650.1	Kunitz type Trypsin Inhibitor	38	53	3.6e-23
IMGA|Medtr6g059810.1	Kunitz type Trypsin Inhibitor	38	52	1.2e-22

Table 3.5: Result of JCVI BLAST search using Tr-KPI5 aa sequences

Accession No	Gene	Identity (%)	Positive (%)	E value
IMGA|Medtr6g009650.1	Kunitz type Trypsin Inhibitor	70	83	2.2e-78
IMGA|Medtr6g059730.1	Kunitz type Trypsin Inhibitor	50	65	3.0e-51
IMGA|Medtr6g059680.1	Kunitz type Trypsin Inhibitor	49	64	2.4e-49
IMGA|Medtr6g059810.1	Kunitz type Trypsin Inhibitor	42	55	6.9e-36
IMGA|Medtr6g059650.1	Kunitz type Trypsin Inhibitor	42	56	2.1e-34
IMGA|Medtr6g065460.1	Kunitz type Trypsin Inhibitor	50	60	6.4e-30
IMGA|Medtr6g059530.1	Kunitz type Trypsin Inhibitor	40	53	5.3e-29
IMGA|Medtr6g059410.1	Kunitz type Trypsin Inhibitor	40	53	1.3e-27
IMGA|Medtr6g059760.1	Kunitz type Trypsin Inhibitor	39	53	8.7e-24
IMGA|Medtr6g044810.1	Kunitz type Trypsin Inhibitor	39	50	2.5e-22
IMGA|Medtr6g044780.1	Kunitz type Trypsin Inhibitor	38	51	9.0e-22

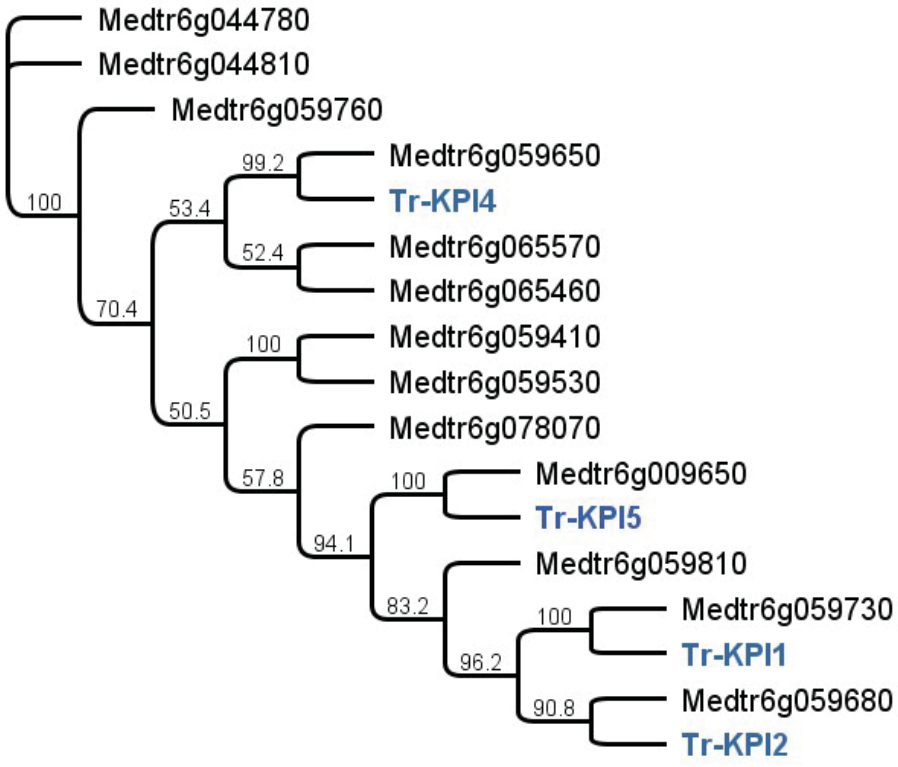


Figure 3.6: Phylogenetic relationship with *M. truncatula* KPI and Tr-KPI proteins. A neighbour joining tree was built with Geneious Pro 5.4.3 using standard parameters.

3.2 Expression of Tr-KPIs in different tissues of white clover

To determine if the four distinct *Tr-KPI* genes display tissue-specific expression, RNA was extracted from leaf, stolon, petiole, root, seed and inflorescence tissue (Figure 3.7) using the Hot Borate method. The qRT-PCR products obtained for all tissues for each gene were confirmed by sequencing.

The overall result of mRNA abundance of the four *Tr-KPI* genes showed that *Tr-KPI1* and *Tr-KPI5* are highly expressed in different tissues when compared with *Tr-KPI2* and *Tr-KPI4* and the results support differential expression of the *Tr-KPI* genes in white clover. qRT-PCR data showed that *Tr-KPI1* is mostly highly expressed in the leaf followed by the inflorescence, root and seed flour (Figure3.7). *Tr-KPI2* is found to be expressed more in the inflorescence tissue followed by root, seed flour and leaf. Among the five different tissues, *Tr-KPI4* is highly expressed in dry seed followed by the stolon and root. The expression of this gene is below the detection level in the leaf, petiole and in the inflorescence tissue (Figure 3.7B). The expression of *Tr-KPI5* was found to be highest in the inflorescence followed by leaf, petiole, stolon, root and dry seed (Figure 3.7B). These results suggest tissue-specific expression of four distinct *Tr-KPI* genes and therefore, further studies were undertaken to observe whether expression of these genes is developmentally-regulated in this plant species.

3.2.1 Expression of Tr-KPI genes in leaves at different developmental stages

To study the expression of *Kunitz* genes in leaves at different developmental stages, the following tissues were collected: the stolon apical structure, first emerged leaf, first fully expanded leaf, mature leaf and onset of senescing leaf (Figure 3.8A).

The expression of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* was detected in all the leaf tissues tested using qRT-PCR. Higher expression of *Tr-KPI2* and *Tr-KPI5* was observed in the apical structure followed by the first emerged leaf when compared with the other leaf tissues (Figure 3.8B). The expression level of *Tr-KPI1* was almost the same in the different leaf tissues, but showed a slightly higher expression in the apical and the first emerged leaf tissues. Out of these three *Tr-KPI* genes, *Tr-KPI5* was found to be more abundant in the apical structure and in the first emerged leaf followed by *Tr-KPI1* and *Tr-KPI2.* The transcript abundance of *Tr-KPI4* was below the detection level in the different leaf developmental stages examined.

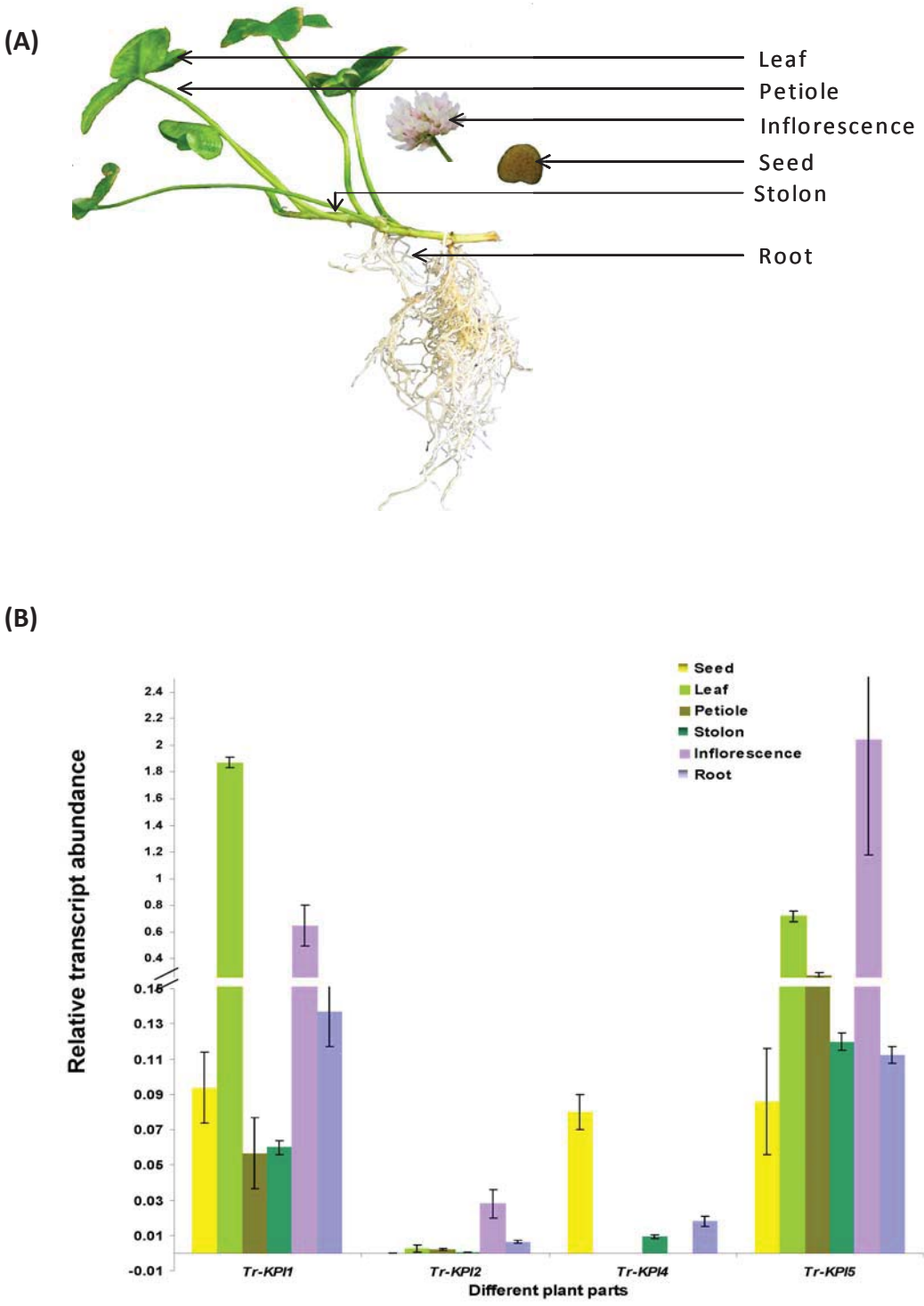


Figure 3.7: Expression of *Tr-KPI* genes in different plant parts of white clover. (A) Different tissues of white clover used for RNA extraction. (B) Expression of *Tr-KPI* genes in different tissues of the white clover plant, as indicated. qRT-PCR was performed using three technical replicates with pooled tissues from 8 plants using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value of three technical repeats and the error bars represent mean ± SEM.

3.2.2 Expression of Tr-KPI genes in different regions of the root

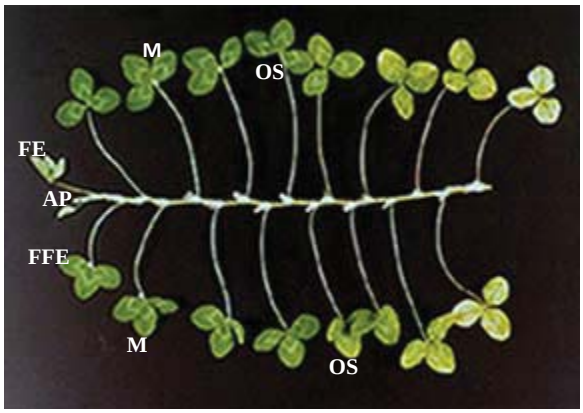
To more closely analyse the expression of *Tr-KPI* genes in the root, white clover stolons were grown in 600mL-capacity PVC pipes with liquid 0.5X Hoagland's media. For collecting the samples, roots were divided into three developmental regions designated as (i) the elongation zone (EZ) which is 1 to 1.5 cm from the tip of the root including the root cap, root tip and zone of elongation, (ii) the zone of visible lateral roots (VL), a the region falling in between the EZ and mature root with emerging visible lateral roots, and (iii) the mature root zone (MR) which is the zone of maturation proximal to lateral roots.

The expression of all four *Tr-KPIs* was observed in each root developmental zone. Almost similar level of expression was observed for *Tr-KPI1* in all three different parts of root showing slightly higher abundance in EZ followed by VL and MR. *Tr-KPI2* followed the same pattern as *Tr-KPI1* but the transcript abundance was much lower. The abundance of *Tr-KPI4* was observed to be higher in the VL zone compared to EZ and MR zone (Figure 3.9B). Higher expression of *Tr-KPI5* was observed in the EZ, followed by the VL and MR.

3.2.5 Expression of Tr-KPIs during nodule development

The expression level of the *Tr-KPIs* was also investigated during root nodulation. Nodules were collected from plants growing in 0.5X Hoagland's media and three different developmental stages were collected: young developing nodules less than 7 days old, crown nodules developing from the primary root (17 days old), and mature nodules of more than 30 days old (Figure 3.10, A). In the analysis, *Tr-KPI1* and *Tr-KPI5* were found to be more prominent in these different stages of nodule development (Figure 3.10, B) and a higher level of expression was found in the crown nodule for both genes. *Tr-KPI2* showed lower expression when compared with *Tr-KPI1* and *Tr-KPI5,* whereas only a basal level of expression was observed for *Tr-KPI4* gene in all stages examined.

(A)



(B)

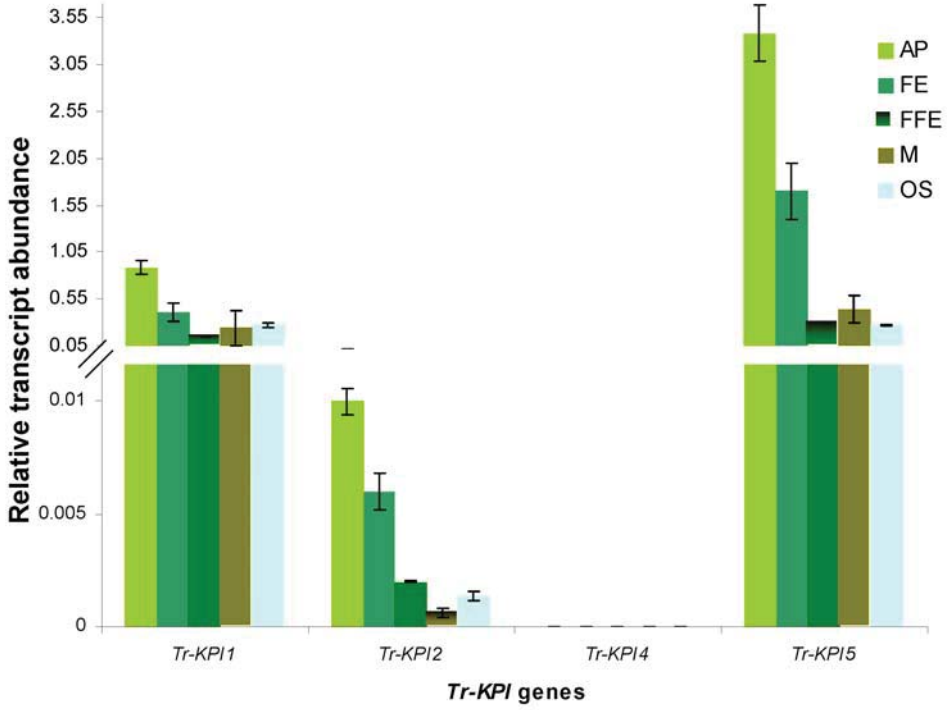
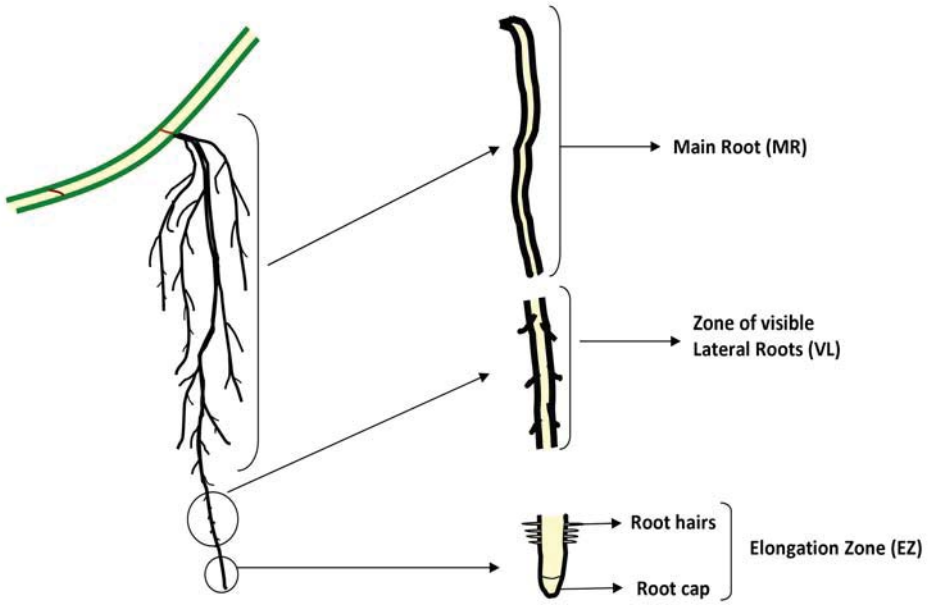


Figure 3.8: A. Leaf developmental stages along a single stolon of white clover. AP denotes the apical structure, FE: first emerged leaf, FFE: first fully expanded leaf, M: Mature leaf and OS: onset of senescing leaf. Image was adapted from Hunter *et al.,* (1999). B. Expression of *Tr-KPIs* in different leaf developmental stages as indicated. qRT-PCR was performed using three technical replicates with pooled tissues collected from 15 plants using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value and error bars representing mean ± SEM.

(A)



(B)

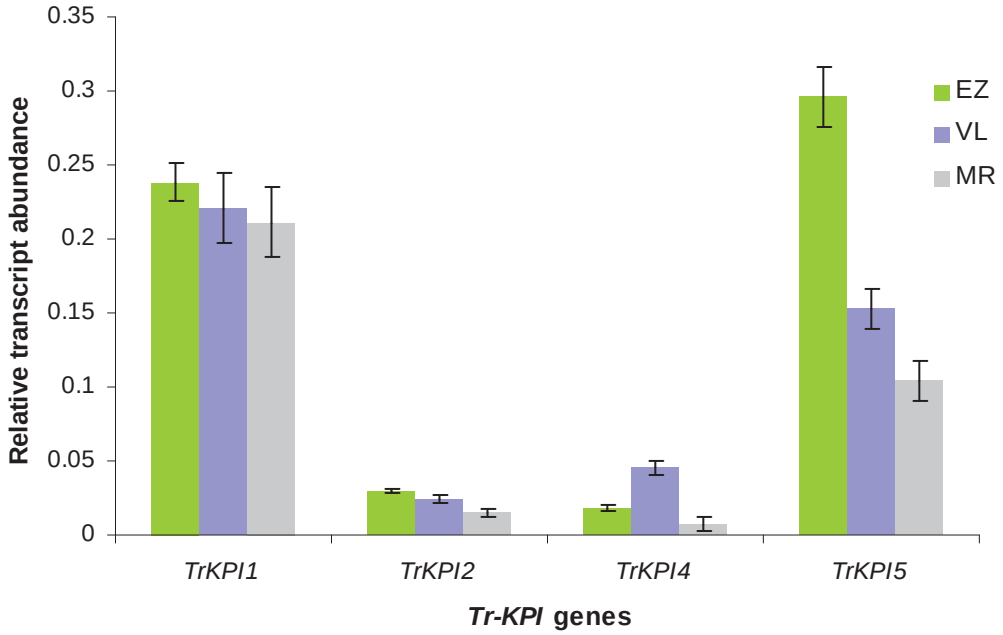
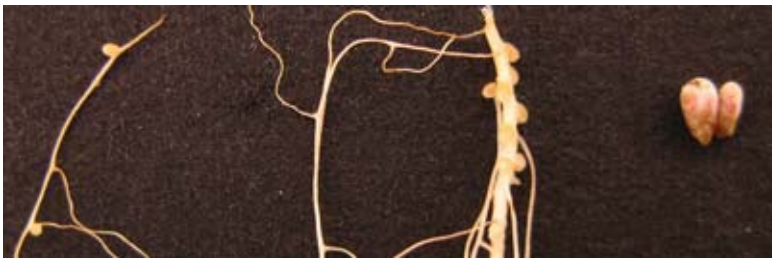


Figure 3.9: A. Stages of root development in white clover used for expression studies. EZ denotes the elongation zone, VL: visible lateral root zone, MR: mature root zone with the lateral roots excised. B. Expression of *Tr-KPIs* in the different root parts. qRT-PCR was performed using three technical replicates. Data presented is the mean of 5 biological replications using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value and error bars representing mean ± SEM.

A)



Young Nodule (7days old) | Crown Nodule (17 days old) | Mature Nodule (30 days old)

B)

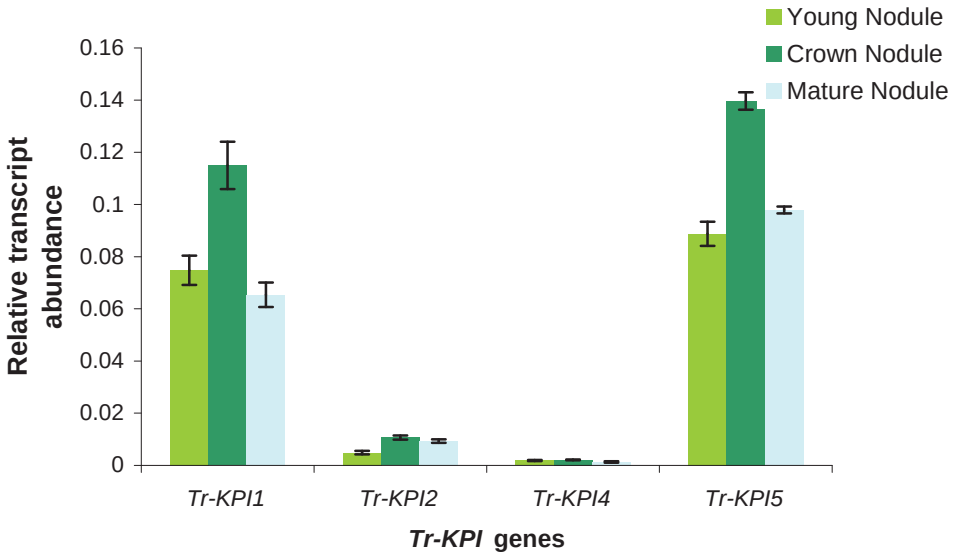


Figure 3.10: Nodule development stages (A) and expression of *Tr-KPIs* during the stages of nodulation, as indicated (B). qRT-PCR was performed using three technical replicates with pooled tissues collected from 9 plants using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value and error bars representing mean ± SEM.

3.2.5 Expression of Tr-KPI genes during germination

To study the expression level of *Tr-KPI* genes during germination of white clover, seeds were placed on three layers of water-soaked sterile filter paper on petri-dishes, covered with aluminium foil and imbibed at 25^0C. Radical emergence began by 10 h and was almost complete by 24 h, and so the whole imbibing seeds were harvested after 1 h, 3 h, 6 h and 12 h and the embryonic axes (without the cotyledon) were excised and collected at 18 h, 24 h, 36 h, 48 h and 72h (Figure 3.11A).

The presence of *Tr-KPI1, Tr-KPI4* and *Tr-KPI5* transcripts was observed in the dry seed flour and at different time points during germination (Figure 3.11B). Using qRT-PCR, a higher relative expression for *Tr-KPI1* and *Tr-KPI5* was observed at 24 h germination onwards whereas *Tr-KPI2* was highly expressed at 48 h onward. The expression of *Tr-KPI4* was detected in dry seed flour and up to 6hr of imbibition and then only a basal level of expression was detected from 24 h onward.

A)

B)

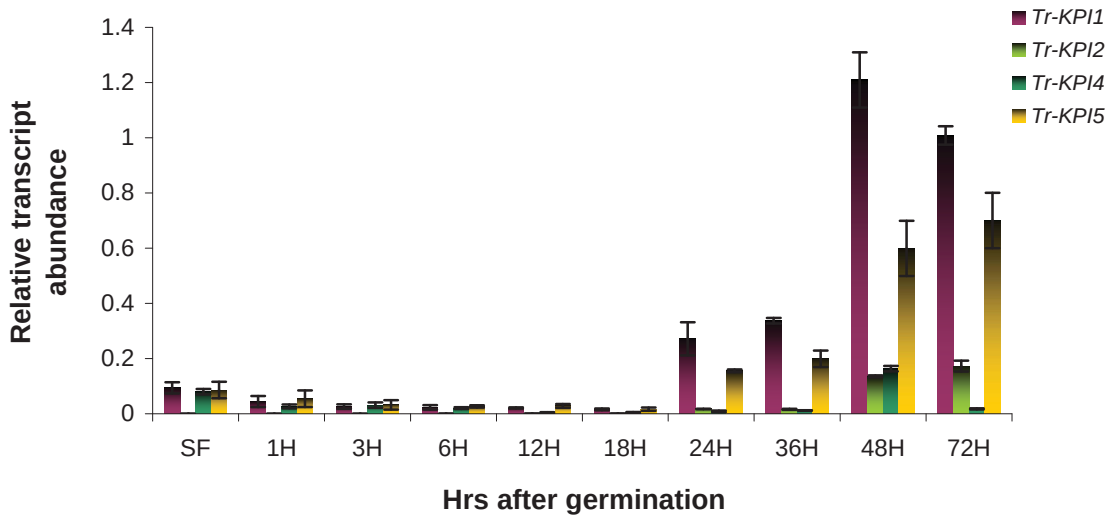


Figure 3.11: A. Imbibed and germinated seeds of white clover. The purple arrow indicates the emergence of the radical. B. Expression of *Tr-KPIs* in germinating seeds and seedlings of white clover at the post-imbibition time points indicated. SF= seed flour. qRT-PCR was performed using three technical replicates with *β-actin* and *GAPDH* as internal control and the experiment was repeated twice. Each data point represents the mean value and the error bars represent mean ± SEM of the technical replicates.

3.3 Tr-KPI genes and biotic stress

3.3.1 Response of Tr-KPI genes by mechanical wounding

Different groups of plant proteinase inhibitors have been reported to be induced by mechanical wounding in plant species belonging to a variety of families. These include chickpea from the Leguminosae (Jime´nez *et al*., 2008), poplar from the Salicaceae (Major and Constabel, 2008) and potato from the Solanaceae (Turra` *et al.*, 2009). Therefore, to study whether the *Tr-KPI* genes in white clover are also responsive to mechanical wounding and to determine whether the transcript of any specific inhibitor gene displays differential expression, four node stolon cuttings were excised and grown on vermiculite for four weeks before the wounding experiment was undertaken. Wounding was performed on the first fully expanded leaf (FFE) using forceps to induce crush damage at three sites on each trifoliate (Figure 3.12). To study local expression, the FFE leaves were harvested at 0, 1 h, 3 h, 6 h, 12 h and 24 h time points post-wounding. To investigate the systemic response of the *Tr-KPI* genes*,* the root tissues subtending from the 3rd and 4th nodes were also collected separately (Figure 3.12).

3.3.1.1 Local response of Tr-KPIs in leaf tissue

The expression of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* in the control tissue varied at different time points of the day when compared with 0 h, but all three followed a similar pattern suggesting that they may be diurnally regulated (Figure 3.13). In response to mechanical wounding, all three *Tr-KPI* genes showed a significant increase in transcript abundance in the wounded tissue from 3 h of wounding to 24 h when compared with 0hr. The transcript abundance decreased initially up to an hour compared to the control and from one hour onward all three genes were upregulated significantly up to 3 h. *Tr-KPI1* showed a 45-fold increase at 3 h after wounding whereas both *Tr-KPI2* and *Tr-KPI5* showed 12-fold increases suggesting a quick response to wounding when compared with the unwounded control. The transcript level decreased at 6 h and 12 h for *Tr-KPI1* and at 6 h for *Tr-KPI2*, whereas *Tr-KPI5* maintained a similar level of expression in the wounded tissue. At 24 h, *Tr-KPI1* showed a 4-fold, *Tr-KPI2* showed a 2-fold and *Tr-KPI5* showed a 5-fold increase when compared with the control suggesting a prolonged response to a single event in the wounded tissue. Therefore, the result of the local expression to wounding suggests that *Tr-KPI1* showed the highest earlier response followed by *Tr-KPI2* and *Tr-KPI5,* but all at 3 h (Figure 3.13).

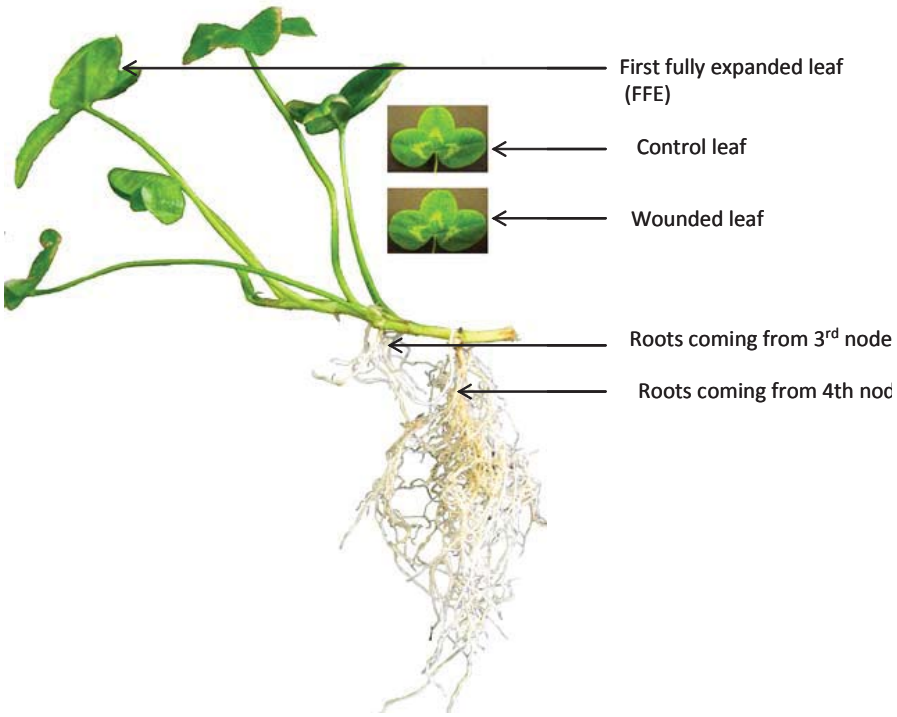


Figure 3.12: Stolons of the white clover cultivar 'Huia' with the leaf and root tissues used to study wounding response of *Tr-KPI* genes, as indicated

3.3.1.2 Systemic response of Tr-KPIs in root tissue in response to wounding in the leaves

The *Tr-KPI* gene family did show a systemic response in root tissue in response to mechanical wounding of leaf tissue. The expression level of *Tr-KPI1* did not show any significant difference in the control tissue over the first 3hr. However, wounding significantly induced the expression of *Tr-KPI1* from 6 h to 24 h in the roots from the 3rd node when compared with both control and the level of expression at 0 h (Figure 3.14). For *Tr-KPI2*, a similar kind of expression pattern was observed as *Tr-KPI1.* For *Tr-KPI4,* significant increases in the transcript level were observed at 1 h, 6 h and 12 h in the control tissue compared to 0 h whereas, in the wounded tissue, the transcript abundance followed the same pattern as *Tr-KPI1* and *Tr-KPI2.* For *Tr-KPI5,* a significant difference in the wounded tissue was observed only at 24 h compared to both 0 h and control 24 h. Among the four genes studied in the root growing from 3rd node, *Tr-KPI4* showed the higher transcript abundance at 6 h followed by *Tr-KPI2* and *Tr-KPI1.*

However, in the roots from the 4th node, no significant difference in the transcript level was observed for *Tr-KPI1* and *Tr-KPI5* in the control tissue compared to 0hr whereas, only *Tr-KPI2*

and *Tr-KPI4* showed a significant increase at 3 h and 6 h compared to 0 h. In the wounded tissue, *Tr-KPI1* showed a significant increase in transcript abundance at 3, 12 and 24 h when compared with both 0 h and the control. *Tr-KPI2* showed a significant increase in transcript abundance at 3 h compared to both control and 0 h. *Tr-KPI4* showed down-regulation in expression in the wounded sample compare to the control at 3, 6 and 12 h (Figure 3.15). Similar to *Tr-KPI2, Tr-KPI5* showed a significant increase only at 3 h after wounding when compared with both 0 h and control.

To conclude, the *Tr-KPI* genes in white clover are induced in both local and systemic tissues in response to mechanical wounding. Therefore, the next objective of the study was to see whether these genes are responsive against insect herbivory.

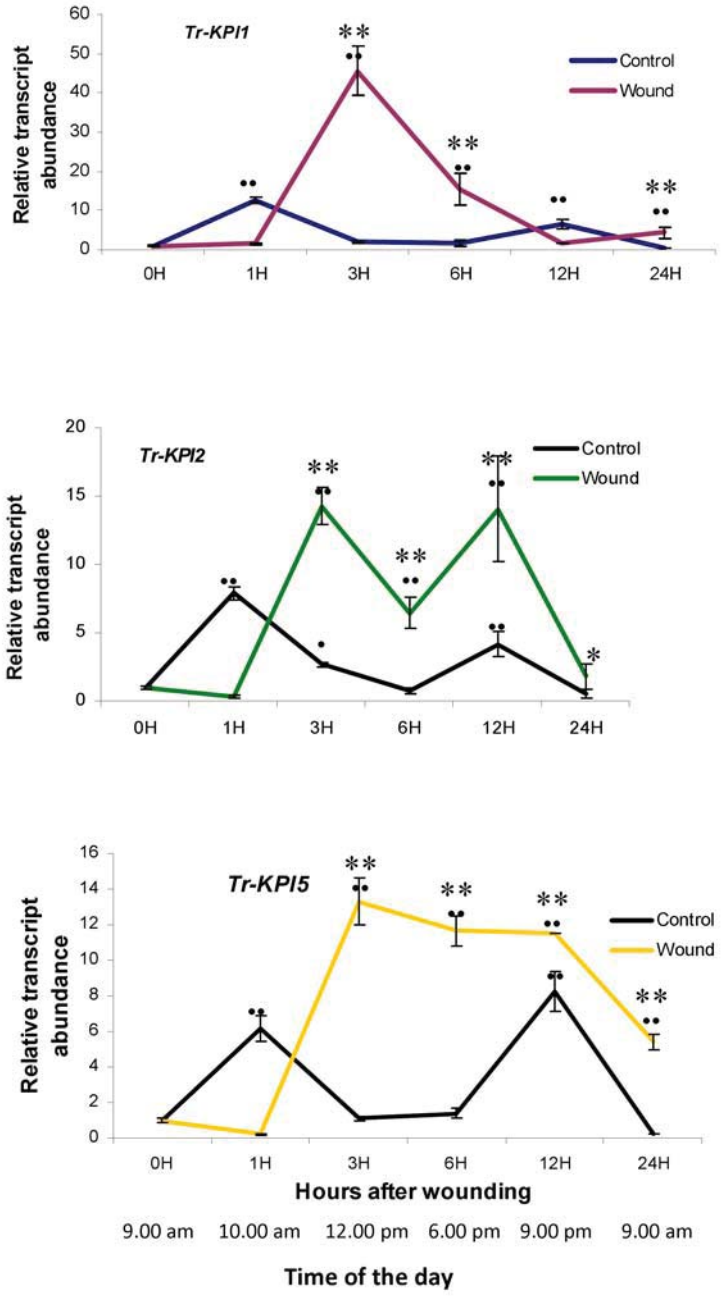


Figure 3.13: Changes in *Tr-KPI* gene expression, as indicated, in response to mechanical wounding of the first fully expanded leaf (FFE). The hours denote post-wounding. Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of two biological replicates, each consisting of pooled tissue from three plants. '**' indicate statistically significant (P<0.01) differential expression in comparison with untreated plants and '•' and '••' indicates statistically significant (P<0.05; P<0.01) differential expression in comparison with 0 h using student's t-test.

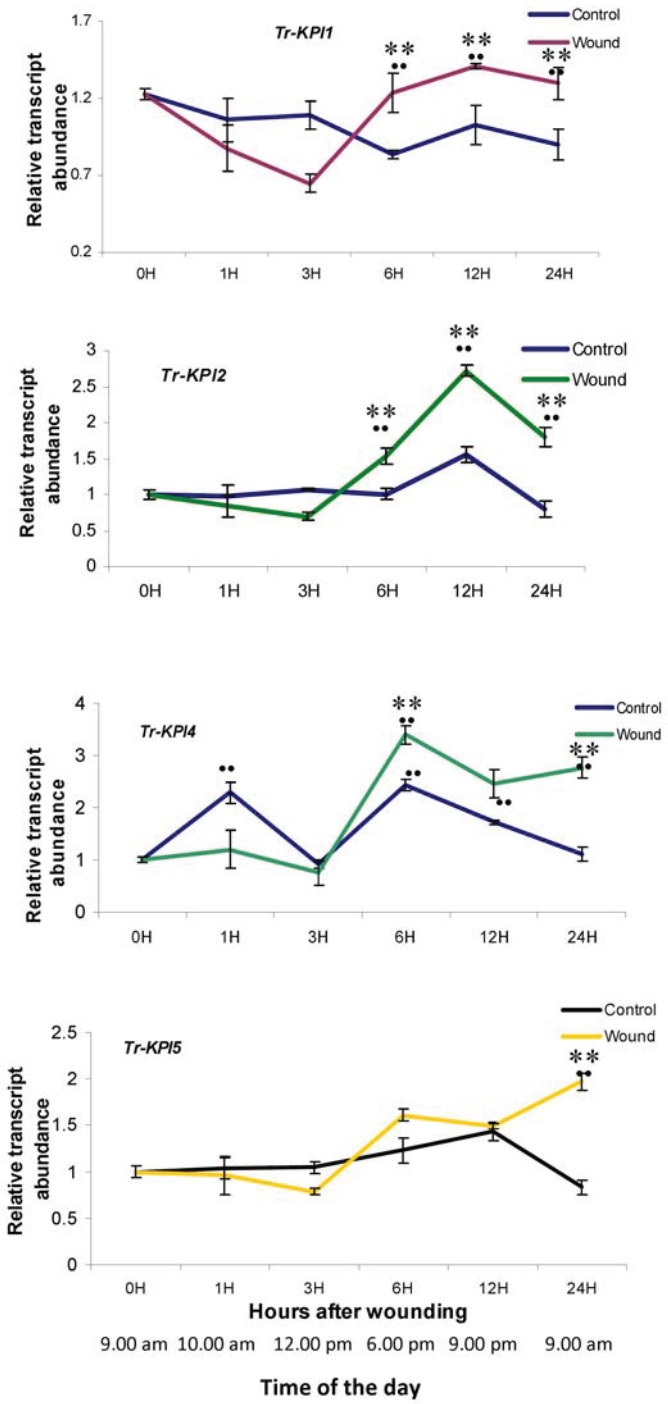


Figure 3.14: Changes in *Tr-KPI* gene expression, as indicated, in roots subtending from the third node in response to mechanical wounding on the first fully expanded leaf (FFE). The hours denote post-wounding. Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of two biological replicates, each consisting of pooled tissue from three plants. '**' indicate statistically significant ($p<0.01$) differential expression in comparison with untreated plants and '•' and '••' indicates statistically significant ($P<0.05$; $p<0.01$) differential expression in comparison with 0 h using student's t-test.

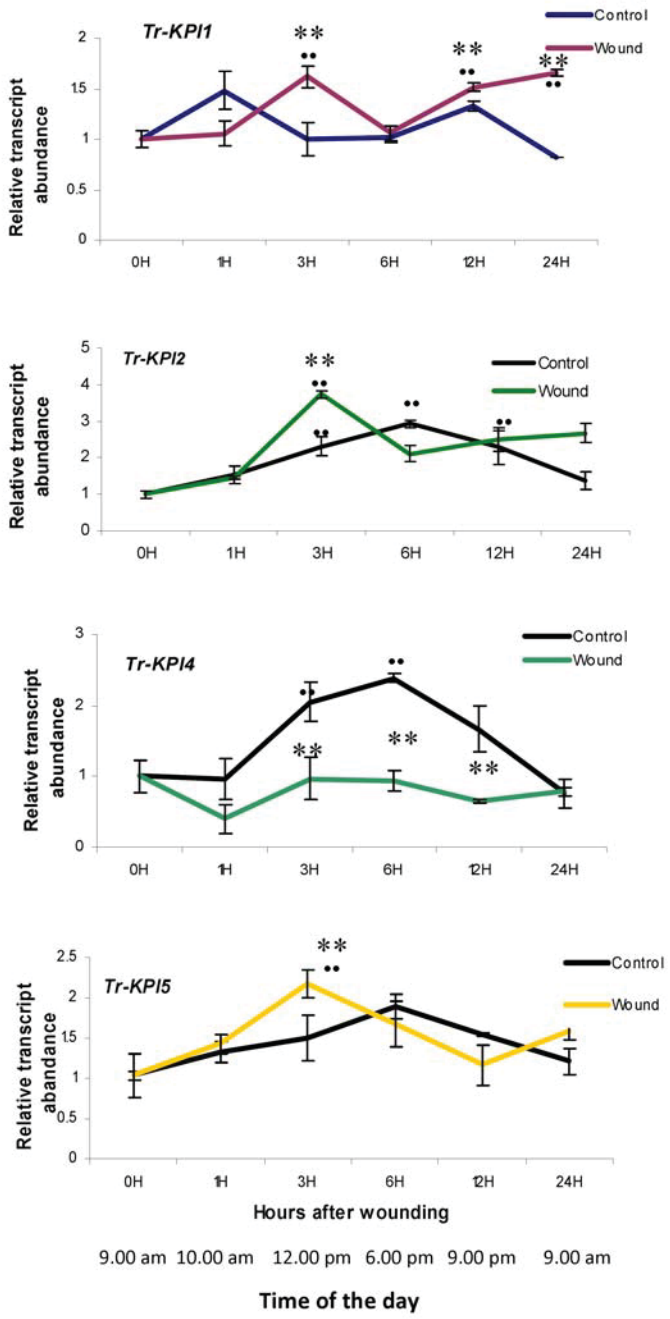


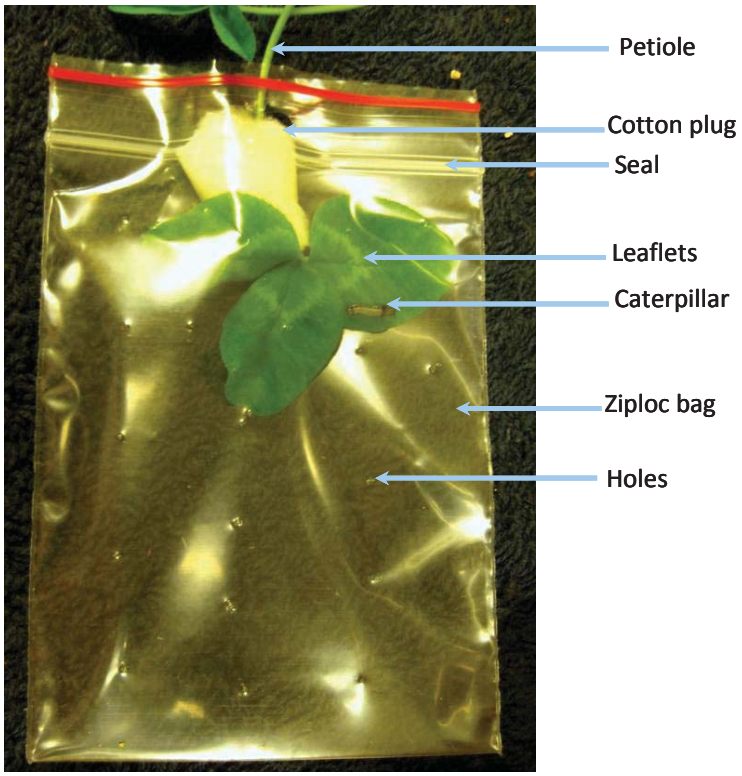
Figure 3.15: Changes in *Tr-KPI* gene expression, as indicated, in roots subtending from the fourth node in response to mechanical wounding on the first fully expanded leaf (FFE). The hours denote post-wounding. Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of two biological replicates, each consisting of pooled tissue from three plants. '**' indicate statistically significant (P<0.01) differential expression in comparison with untreated plants and '•' and '••' indicates statistically significant (P<0.05; P<0.01) differential expression in comparison with 0 h using student's t-test.

3.3.2 Changes in Tr-KPI gene expression in response to insect herbivory

Protection of plants from plant insect herbivores and pathogens is a well known function of proteinase inhibitors which are induced as a means of defense. The results from the wound study revealed the response of *Tr-KPI* genes in white clover by mechanical wounding (Figure 3.13, 3.14 and 3.15). Therefore, to investigate whether the transcript level of *Tr-KPIs* shows any change in local and systemic tissue in response to insect herbivory, white clover plants were subjected to feeding by the tobacco cutworm (*Spodoptera litura*), a generalist herbivore.

For this study, stolons of the cultivar Huia were grown in soil for five weeks, and the first fully expanded leaf (FFE) of white clover was selected for insect feeding. The whole lamina, consisting of the three trifoliates, of the first fully expanded leaf was placed inside a ziploc plastic bag (65 mmx75 mm). Little holes were made on the bag by needle to prevent condensation and the junction of the petiole and lamina was encircled by a cotton plug before the bag was sealed to prevent any contusion or wounding effect, and also to prevent any detachment of lamina and petiole as a consequence of caterpillar feeding (Figure 3.16 A). Early third instar larvae were selected for the experiment, and one larva was placed per Ziploc bag with the help of forceps and the bags were sealed. An identical setup was followed for the control plants, except without larvae. To study the local *Tr-KPI* responses, *Spodoptera* inoculated FFE leaves were collected at 1 h, 3 h, 6 h, 12 h and 24 h (Figure 18 B), and to investigate the systemic response of *Tr-KPIs* genes, apical tissues and root tissues were also collected at the same time points. The experiment consisted of three biological replications each containing pooled tissues from two stolons.

A)



B)

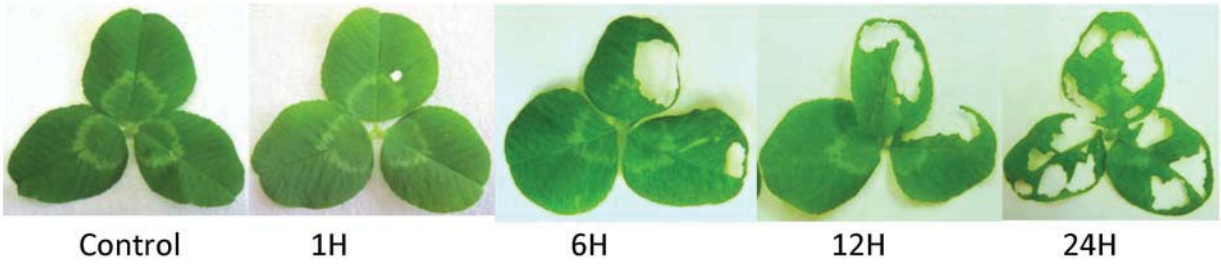


Figure 3.16: A. Experimental setup for herbivory trial by *Spodoptera litura*. B. The first fully expanded leaf after feeding at 1 h, 6 h, 12 h and 24 h.

3.3.2.1 Response of Tr-KPIs in local tissues by insect herbivory

The transcript abundance of *Tr-KPI1* in the control leaf tissue showed significant increases at 3, 6 and 24 h when compared with 0 h, while the expression of *Tr-KPI1* increased 3- fold within 1hr of insect feeding, then decreased at 3 h and from 6 h onward a very high abundance of the transcript was observed up to 24 h in the local tissue subjected to insect feeding when compared with both the control and 0 h (Figure 3.17). The expression of *Tr-KPI2* in the control tissue only showed a significant increase at 24 h compared to 0hr whereas, in the insect fed leaf, the transcript did not change in expression up to 3 h and then increased from 6 h to 24 h when compared with both the control and 0 h. Although the expression of *Tr-KPI4* was not detected in the leaf tissue by qRT-PCR in the expression (Figure 3.7) or wounding study (data not presented), there is a possibility that this gene is expressed at a level in the leaf and could be induced by *Spodoptera* feeding. Therefore, the expression level of *Tr-KPI4* was also measured in the insect fed leaf tissue, but the transcript was undetectable by qRT-PCR (data not shown). For *Tr-KPI5,* the transcript level did not show any significant variation in the control tissue against 0hr whereas, in the insect fed tissue, significant increases were observed from 3hr to 24hr when compared with both control and 0 h (Figure 3.17). Among the three genes, *Tr-KPI1* was found to be the most highly expressed and the earliest induced gene which showed around 1600-fold increase (24 h) followed by *Tr-KPI5* (160-fold, 24 h) and *Tr-KPI2* (170-fold, 24 h).

3.3.2.2 Response of Tr-KPIs in systemic tissues by insect herbivory

In the control plant, the abundance of *Tr-KPI1* in the apical tissue varied significantly at 1 h, 3 h, 12 h and 24 h when compared with 0 h which may suggest diurnal fluctuation, whereas in the insect fed plant, the transcript level started to increase from 6 h, but differed significantly at 12 h against 0 h, and at 24 h against both 0 h and the control (Figure 3.18). *Tr-KPI2* showed a similar pattern to *Tr-KPI1* where in the control a significant difference was observed at 1 h, 3 h and 12 h when compared with 0 h. The transcript abundance started to increase in the insect-fed plants from 6 h and varied significantly at 12 h against 0 h and at 24 h against both the control and 0 h. In the case of *Tr-KPI5*, though the mRNA level went up at 3 h in the insect-fed plants when compared with 0 h, the transcript level did not change at 6 h and 12 h when compared with 0 h and control (Figure 3.18). However, at 24 h, the transcript level increased significantly when compared with the control and 0 h. Among the different genes, *Tr-KPI1* was

found to be the most highly expressed gene with a 10-fold increase followed by *Tr-KPI2* and *Tr-KPI5.*

In the root tissue, the transcript abundance of *Tr-KPI1* did not show any significant variation in the control sample compared to 0 h but in the insect-fed plants, the transcript level increased slightly from 1 h onward to 6 h compared with the control and 0 h, and then decreased significantly at 12 h and 24 h when compared with 0 h (Figure 3.19). A similar expression level of *Tr-KPI2* was observed in the control with a significant decrease in the transcript level at 12 h and 24 h when compared with 0 h. *Tr-KPI4*, which is undetectable in the local leaf tissue of the insect-fed plants, showed a significant increase in the root tissue from 1 h to 6 h compared with the control and 0 h and a decrease at 12 and 24 h when compared with the control and 0 h (Figure 3.19). *Tr-KPI5* showed the same kind of response as *Tr-KPI4* in the roots of the insect-fed plants. The transcript abundance for this gene showed a significant increase at 1 h and 3 h, and then a significant decrease at 12 h compared with 0 h, and at 24 h compared with both 0 h and control (Figure 3.19).

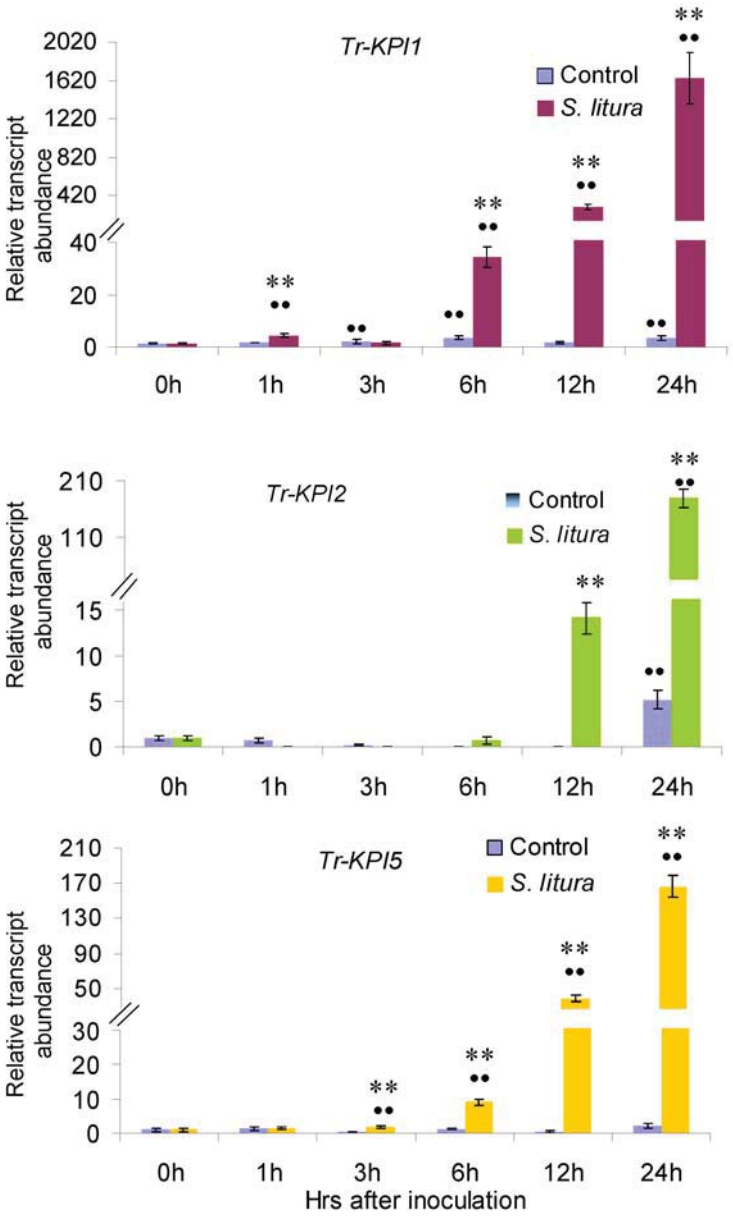


Figure 3.17: Changes in *Tr-KPI* gene expression, as indicated, in response to insect herbivory of the first fully expanded leaf (FFE). Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of three biological replicates, each consisting of pooled tissue from two stolons. '**' indicate statistically significant (P<0.01) differential expression in comparison with untreated plants and '••' indicates statistically significant (P>0.01) difference against 0 h using student's t-test.

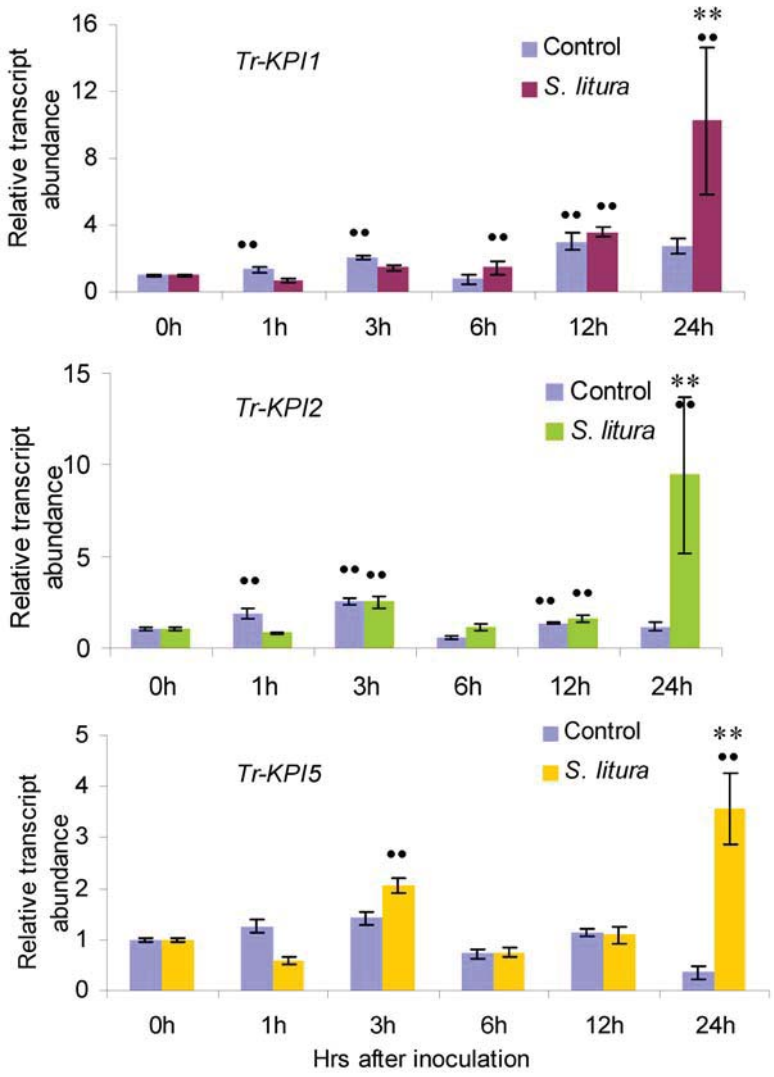


Figure 3.18: Changes in *Tr-KPI* gene expression, as indicated, in apical tissue in response to insect herbivory of the first fully expanded leaf (FFE). Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of three biological replicates, each consisting of pooled tissue from two stolons. '**' indicate statistically significant (P<0.01) differential expression in comparison with untreated plants and '••' indicates statistically significant (P<0.01) difference against 0 h using student's t-test.

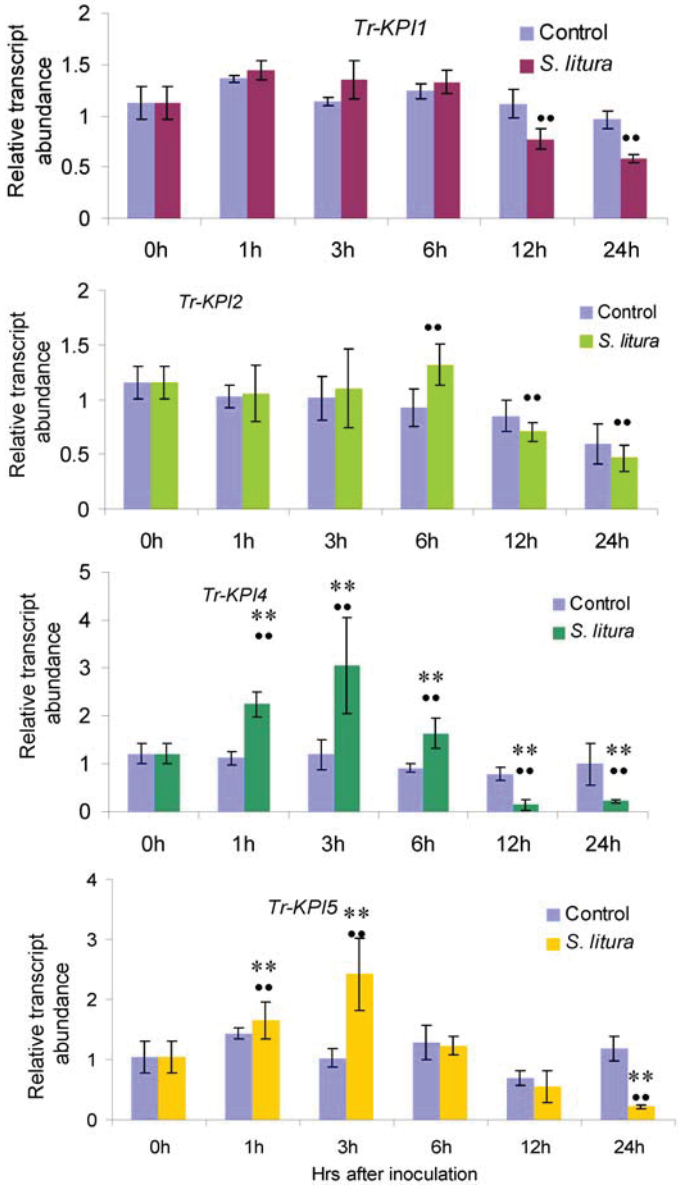


Figure 3.19: Changes in *Tr-KPI* gene expression, as indicated, in roots in response to insect herbivory of the first fully expanded leaf (FFE). Relative transcript abundance was determined by qRT-PCR and was normalized using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of three biological replicates. '**' indicate statistically significant ($P<0.01$) differential expression in comparison with untreated plants and '••' indicates statistically significant ($P<0.01$) difference against 0 h using student's t-test.

3.3.3 Response of Tr-KPI genes in white clover to nematode feeding

3.3.3.1 Response of Tr-KPI genes in white clover to root-knot and cyst nematode feeding

The wounding and insect herbivory studies evidently indicated that *Kunitz proteinase inhibitor* genes in white clover are induced in both local and systemic tissues. Therefore, to see whether the infestation of nematode in the roots of clover plants can also influence the expression of *Tr-KPI* genes both locally and systematically, cultivar Huia, which is a moderately susceptible host to nematodes (Watson *et al.*, 1996), was inoculated with eggs of the white clover cyst (*Heterodera trifolii*) and root-knot (*Meloidogyne trifoliophila*) nematode, collected from AgResearch Grasslands. The leaf and root tissues were harvested at 4 and 8 days after inoculation. In the Huia cultivar, *M. trifoliophila* produced galls by four days and both cyst and root knot nematode 2nd stage juveniles were observed in the stained roots collected after 4 days (Figure 3.20).

The transcript level of *Tr-KPI1* was found to accumulate significantly in the leaf tissue at day 8 after inoculation by *H. trifolii* and *M. trifoliophila*. Interestingly, the transcript level was significantly increased in the root tissue infected by the cyst nematode at day 4 and day 8, but not by the root-knot nematode (RKN) (Figure 3.21). The *Tr-KPI2* transcript accumulated significantly both in the root and leaf tissue at 8 days after inoculation by *H. trifolii* (Figure 3.21) and in the leaf tissue at day 4 after infestation by the RKN. In contrast, a decrease in *Tr-KPI2* expression was observed at day 4 after inoculation in root tissue of plants infected with RKN. As previous data suggested that *Tr-KPI4* is undetectable in the leaf tissue, only root tissues were used to study the expression pattern in response to nematode infestation. *Tr-KPI4* showed an earlier significant induction at day 4 and day 8 after inoculation by cyst nematode and a down-regulation in the case of root-knot nematode infestation at day 4 (Figure 3.21). For *Tr-KPI5*, a significant abundance in transcript level was observed at day 4 and day 8 in responses to cyst nematode infestation in the local tissue whereas a significant systemic response was observed in the leaf tissue at day 8. Infestation by RKN down-regulated gene expression at day 8 in the root tissue but did not display any variation in abundance in the leaf tissue (Figure 3.21).

A) Control at day 4

B) Root infested with *M. trifoliophila* at 4 day

C) Root infested with *H. trifolii* at 4 day

D) Root infested with *M. trifoliophila* at 4 day

E) Root infested with *H. trifolii* at 4 day

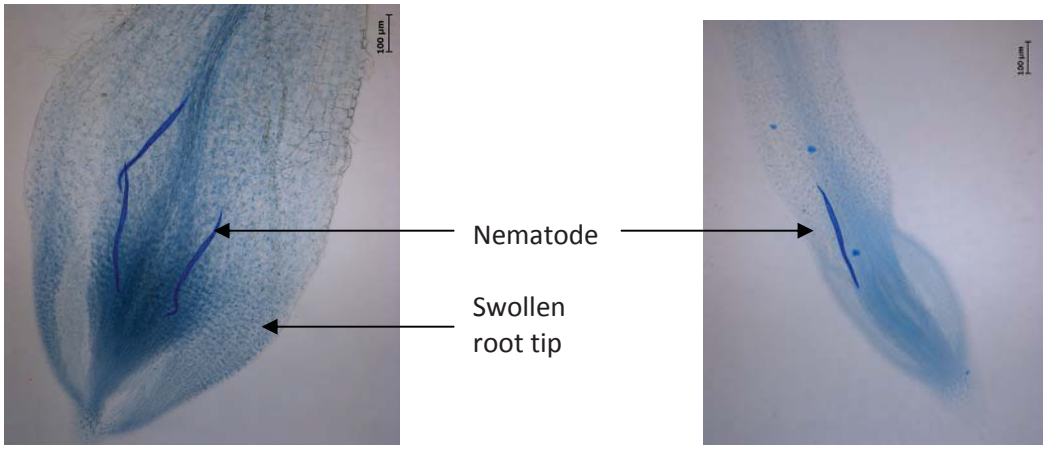


Figure 3.20: Stolon of white clover cultivar Huia at day 4 after inoculation by cyst and root knot nematodes. (A) Control plant at day 4, (B) and (C) infected stolons with root-knot and cyst nematode respectively. (D) and (E) Microscopic image of root-knot and cyst nematode inside the root after staining with 0.05% (v/v) aniline blue.

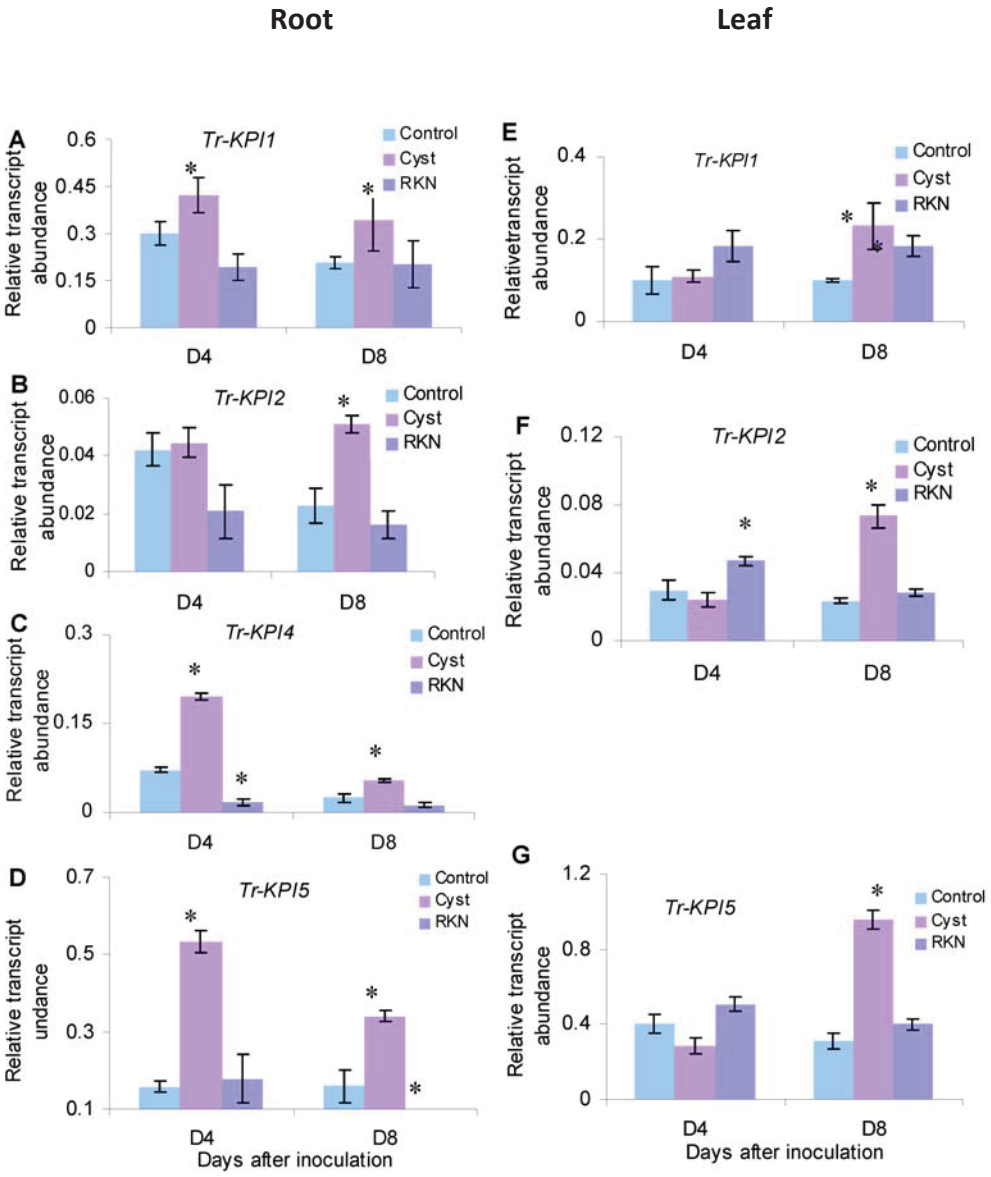


Figure 3.21: Changes in *Tr-KPI* gene expression in response to nematode infestation, as indicated, in leaf and root tissue. The relative abundance of mRNA in root tissue root tissue A) *Tr-KPI1*, B) *Tr-KPI2*, C) *Tr-KPI4* and D) *Tr-KPI5* and in leaf tissue E) *Tr-KPI1*, F) *Tr-KPI2* and G) *Tr-KPI5.* qRT-PCR was performed using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of three biological replicates. '*'indicate statistically significant (P<0.05) differential expression in comparison with untreated plants using student's t-test.

3.3.3.2 Response of Tr-KPI genes in white clover resistant and susceptible lines infested with Heterodera trifolii

The qRT-PCR data (section 3.5.1) suggested that the expression of the *Tr-KPI* gene family of white clover is influenced in response to infestation by the cyst nematode *H. trifolii*. Therefore, to further investigate the influence of cyst nematode infestation on *Tr-KPI* gene expression, the white clover breeding line 17 (resistant) and line 23 (susceptible) were examined. The plants are referred to as lines as they are the material from the breeding programme conducted by AgResearch Grasslands, Palmerston North. The plants were inoculated with the egg mass as mentioned previously (chapter 3.5.1, Figure 3.22), and the relative abundance of *Tr-KPI* genes in the root tissues was determined by qRT-PCR in both resistant and susceptible lines at day 4 and day 8 after inoculation, and were compared with the control (non-inoculated).

The transcript abundance of *Tr-KPI1* increased significantly at day 4 and day 8 after inoculation in the resistant line 17R and in the susceptible line 23S when compared with the control (Figure 3.23). The transcript level of *Tr-KPI2* also increased significantly at 4 and 8 days after inoculation in the resistant line 17R, but in the susceptible line 23S, a significant increase in transcript accumulation was observed only at day 4 (Figure 3.23). The expression of *Tr-KPI4* increased greatly at day 4 and 8 after inoculation in the resistant line 17R, but did not show any response to nematode infestation for the susceptible line 23S. The mRNA abundance of *Tr-KPI5* increased significantly at day 4 and day 8 for the resistant line 17R, and while the mRNA level increased significantly at 4 days of infestation, the transcript level did not change compare with the control at day 8 in the susceptible line 23S (Figure 3.23).

A)

B)

C)

Figure 3.22: Experimental setup for nematode inoculation of the resistant (17R) and susceptible line (23S). (A) Stolons of resistant and susceptible lines just before inoculation with eggs from cyst nematode. (B) Control and inoculated stolons of the resistant line 17R harvested at day 4, (C) control and inoculated stolons of the susceptible line 23S harvested at day 4.

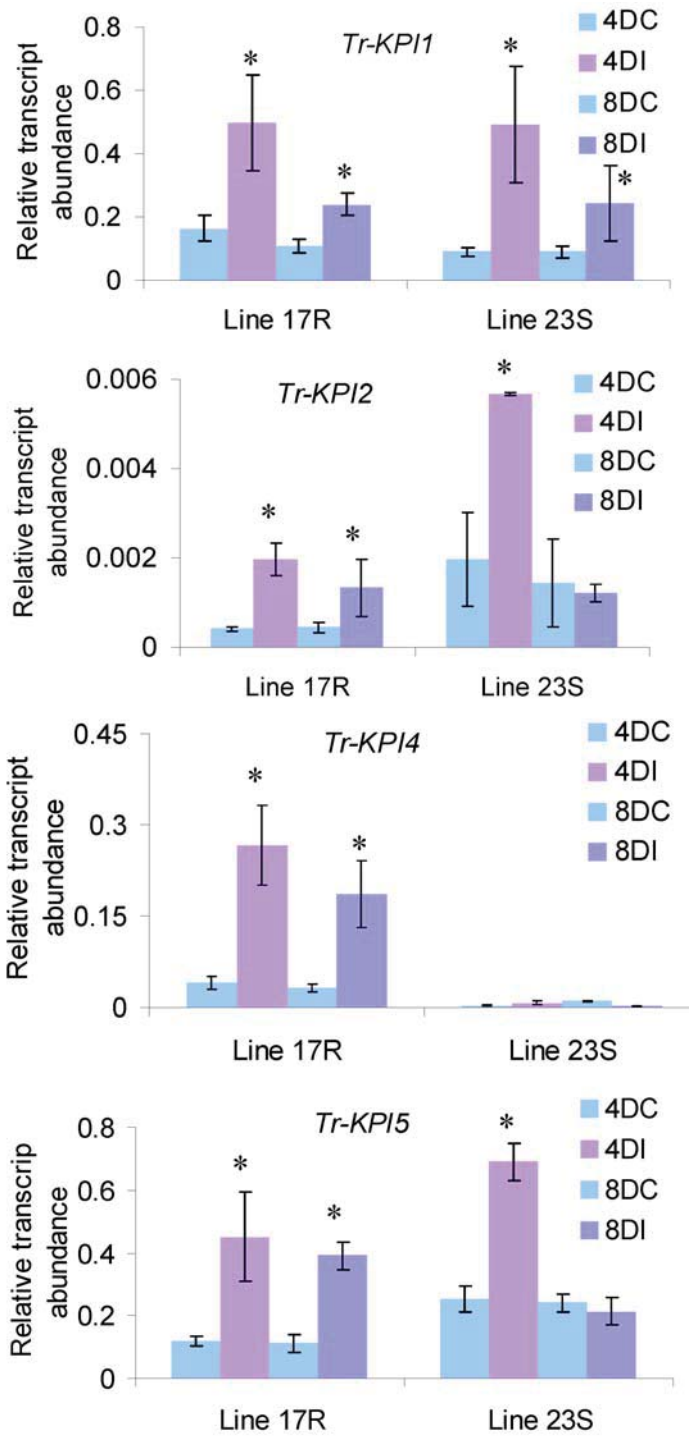


Figure 3.23: Changes in *Tr-KPI* gene expression, as indicated, in root tissue in the resistant (17R) and susceptible (23S) lines in response to cyst nematode (*Heterodera trifolii*) infestation. qRT-PCR was performed using two internal reference genes: *β-actin* and *GAPDH*. Bars represent mean ±SEM of three biological replicates. '*' indicates statistically significant ($P<0.05$) differential expression in comparison with untreated plants using student's t-test.

3.4 Tr-KPI Genes and abiotic stress

The results of the biotic stress studies suggested that changes in the expression of the *Tr-KPI* gene family may be part of the plant defense mechanism. Therefore, the next priority was to investigate the influence of abiotic stress on changes in gene expression. Two different kinds of abiotic stress were applied: osmotic stress (water deficiency) and nutrient limitation (phosphorus deficiency).

3.4.1 The influence of water deficit on Tr-KPI gene expression

To investigate the effect of water deficit, two different sets of treatments were imposed: one without a previous water deficit treatment [called non pre-stressed (NPS)] and a second treatment involving one week of water deficit treatment, a week of rehydration and then the water deficit treatment [called pre-stressed (PS)] (Figure 3.24). The plants (cultivar Huia) were grown in a mixture of vermiculite and perlite at a ratio of 1:1 in 1.2 L-capacity pots (Figure 3.24 A) under 65% relative humidity with 14hr light and 10 hr dark. The temperature of the growth room was maintained at 22^0C.

The same experimental setup was used previously by Dr. Alluh Nikmatullah, but using soil and a drought resistant ecotype Tienshan and a drought tolerate cultivar Kopu to study the expression level of the 1-aminocyclopropane-1-carboxylic acid (ACC) oxidase (ACO) and ACC synthase gene families (Nikmatullah, 2009). Therefore, for further analysis, the same cDNA preparation was used in the present study to examine the expression of the *Tr-KPI* genes in the Tienshan ecotype and cv. Kopu.

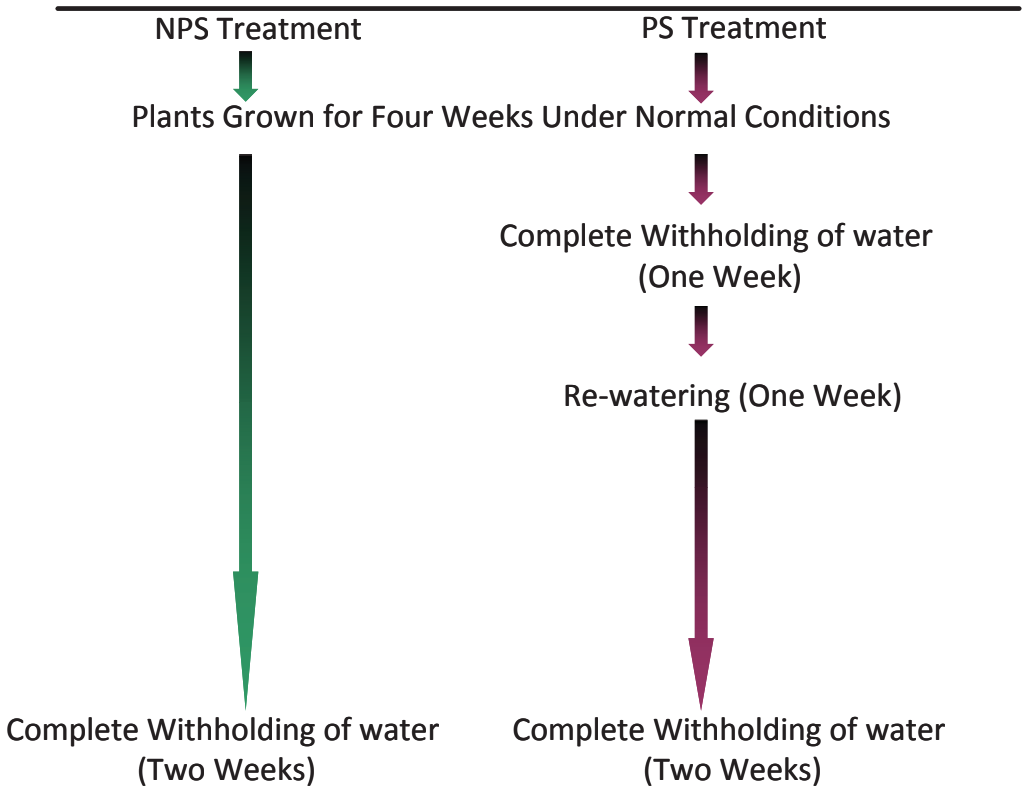


Figure 3.24: Schematic depiction of the two different water deficit treatments imposed on white clover cultivar Huia grown in a vermiculite and perlite mixture (1:1).

3.4.1.1 Change in moisture content of the media

In the water deficit treatment, the moisture content of the media was monitored on a daily basis and used as an indicator that a water deficit was imposed on the plants. An equal amount of 0.5 x Hoagland's media was supplied to the plants and at the start of the experiment, all excess water was allowed to drain away from the media and the media water content determined using the gravimetric method. The experiment was continued for thirteen days. During this period, the moisture content of the media gradually decreased (Figure 3.25B) and as wounding has an influence on the expression of *Tr-KPI* genes (Figure 3.13, 3.14, 3.15), the stolons were allowed to grow without any removal of leaves over this period.

In both the NPS and PS treatments, a similar trend of a decrease in moisture content was observed (Figure 3.25C). The initial moisture content of all the pots was maintained at approximately 74% by applying an equal amount of Hoagland's media. However, the PS treated plants were found to lose moisture quite rapidly compared to the NPS treated plants.

After 13 days of the water deficit treatment, the final moisture content was approximately 32% for both NPS and PS treated plants (Figure 3.25C).

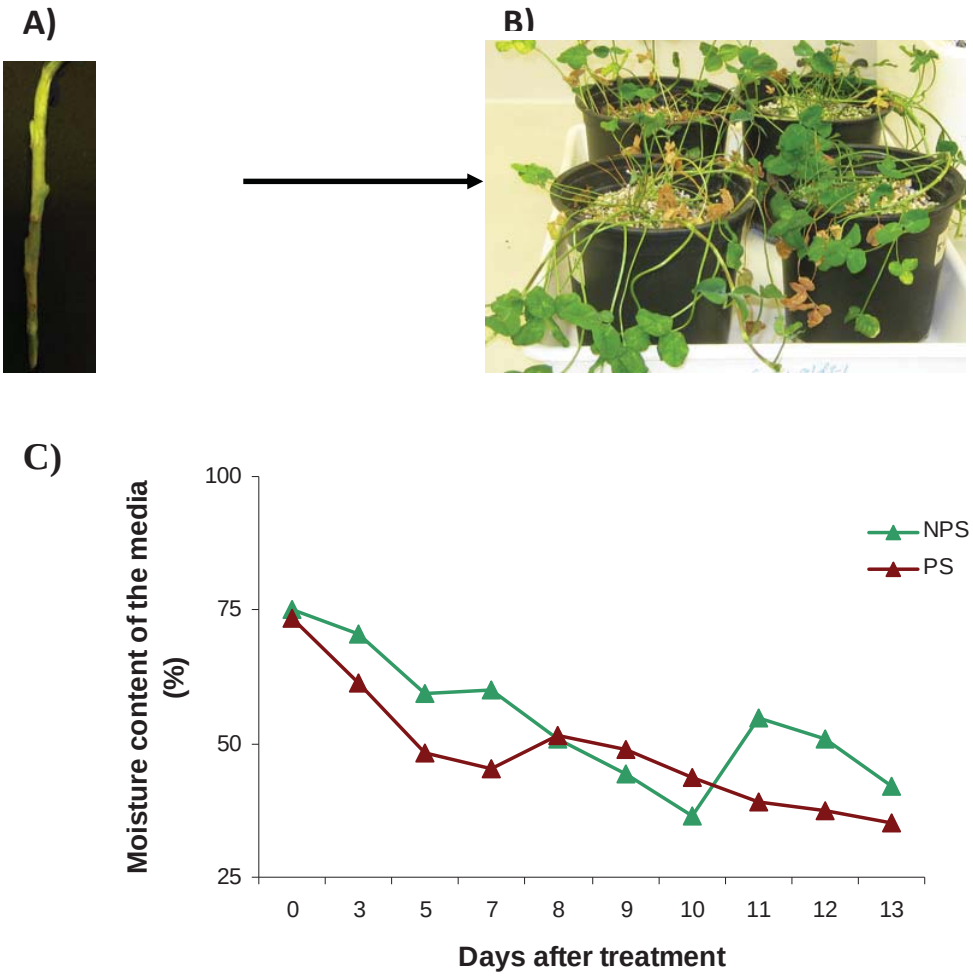


Figure 3.25: Experimental set up of the water deficiency study and moisture content of the media. (A) Four noded stolon of white clover plant cultivar Huia. (B) Plants growing in vermiculite and perlite (1:1) mixture. (C) Changes in media moisture content (n=3) determined by the gravimetric method, for the treatment day, as indicated over a 13 day time course.

3.4.1.2 Expression of Tr-NCED1 in the first fully expanded leaf of white clover in response to water deficit

To act as a further measure of the progress of the water deficit, the expression level of the 9-cis-epoxycarotenoid dioxygenase (NCED1) gene in white clover was examined. The sequence of the *Tr-NCED1* was obtained from AgResearch Grasslands white clover EST database and qRT-PCR primers were designed. The result of qRT-PCR analysis of NPS and PS treated leaf material showed that *Tr-NCED1* was upregulated significantly from a 65% moisture content to 32% in the NPS treatment. In comparison to the NPS treatment, the level of the *Tr-NCED1* expression did not show any increase or decrease (Figure 3.26) in response to the PS treatment although, in a broad sense an increasing trend was observed from 65% to 38% moisture content in response to the PS treatment.

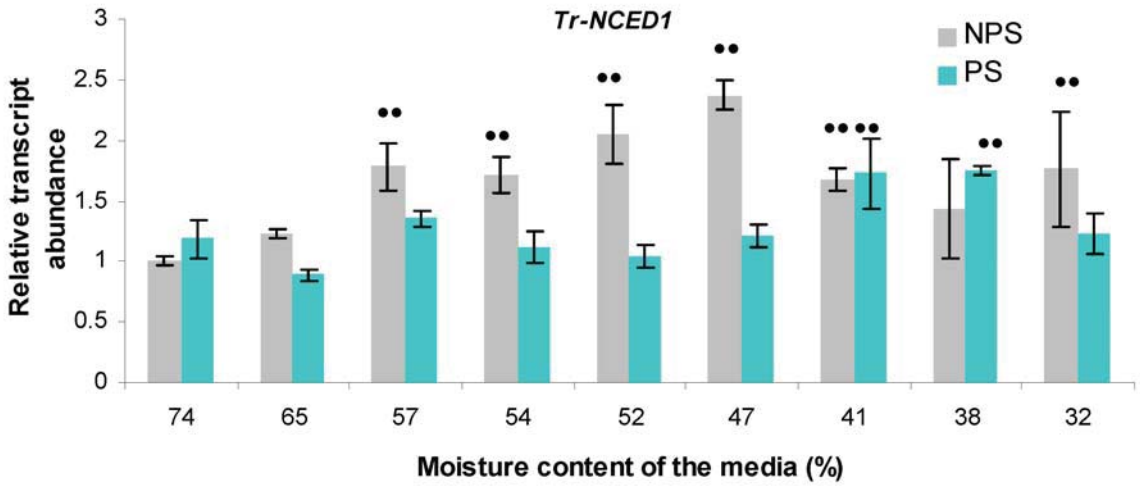


Figure 3.26: Relative expression of *Tr-NCED1* in the first fully expanded (FFE) leaves of NPS and PS treated plants at the moisture content of the media, as indicated in cultivar Huia. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from at least four stolons were used for each data point and error bars representing mean ± SEM. '••' indicate statistically significant ($P<0.01$) differential expression using student's t-test.

3.4.1.3 Expression of Tr-KPIs in the first fully expanded leaf of white clover in response to water deficit

As mentioned previously, the moisture content of the media did not decrease at the same rate for both NPS and PS treated plants, and so changes in the expression of *Tr-KPI* genes were plotted based on decrease in moisture content of the media rather than on a per day basis.

In the NPS-treated first fully expanded leaf (FFE), expression of *Tr-KPI1* significantly increased with a decrease in moisture content from 57% to 32% when compared with the initial moisture content of 74%. For the PS treatment, the mRNA abundance of *Tr-KPI1* also started to increase significantly from 57% until the end of the time-course where the moisture content was 32% (Figure 3.27). However, in the PS treated leaves, expression of *Tr-KPI1* was significantly higher when compared with the NPS treated leaves at the same moisture content. For example, at a 41% moisture content, *Tr-KPI1* expression had increased by 1-fold in the NPS and 5-fold in the PS treatment when compared with the control.

The transcript abundance of *Tr-KPI2* in the NPS treatment did not change when compared with the control (74% moisture content; Figure 3.27). In fact, the transcript level showed a trend of a decrease from 57% to 32% moisture content. However, in the PS treatment, a higher level of transcript abundance was observed at a 57% to 32% moisture content change which represents a 4-fold increase compared to the initial moisture content.

In common with *Tr-KPI1, Tr-KPI5* also showed an increased level of expression over a 57% to 32% change in moisture content in the NPS treatment when compared with the control. However, unlike *Tr-KPI1* and *Tr-KPI2*, *Tr-KPI5* was induced earlier at a 65% moisture content in the case of the PS treatment and maintained a higher transcript abundance than the NPS treatment, up to a 32% moisture content (Figure 3.27).

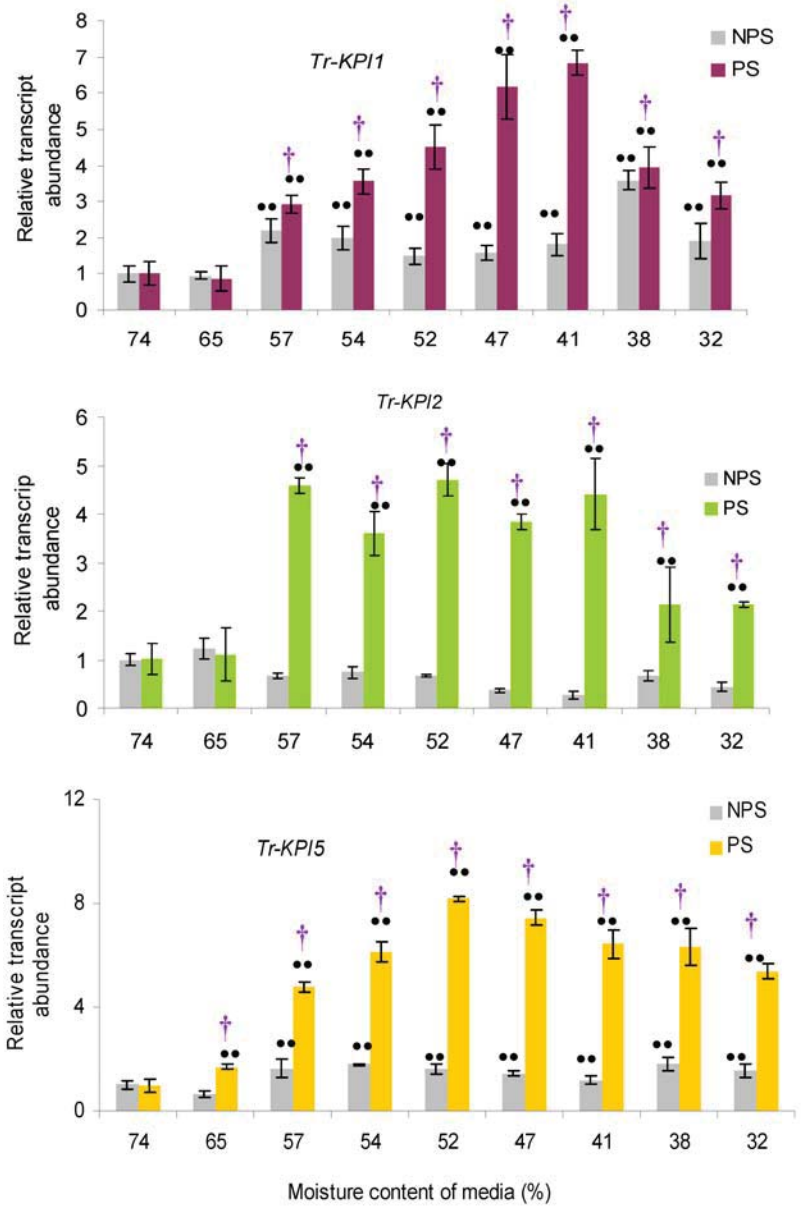


Figure 3.27: Expression of *Tr-KPIs*, as indicated, in the first fully expanded (FFE) leaves in response to NPS and PS treatments at the moisture content of the media, as indicated in cultivar Huia. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from at least four stolons were used for each data point and error bars representing mean ± SEM. '••' indicate statistically significant (P<0.01) differential expression in comparison with initial moisture content and '†' indicates statistically significant (P<0.01) expression in the PS treatment in comparison with the NPS treatments using student's t-test.

3.4.1.4 Expression of Tr-KPIs in roots of white clover in response to a water deficit

Changes in expression of the *Tr-KPI* genes in the root tissue were not as marked as observed in the leaf tissue for both NPS and PS treatment (Figure 3.28). Though *Tr-KPI1* transcript abundance increased in the PS treated root tissue from 54% to 38% moisture content, the values did not vary significantly when compared with the NPS treatment. A decrease in the transcript level was observed for both *Tr-KPI1* and *Tr-KPI2* for the NPS treatment from 65 to 32% moisture content when compared with the initial moisture content (74%). The *Tr-KPI2* transcript level increased at 54%, 47% and 38% moisture contents. For, *Tr-KPI4*, which is mainly expressed in the root tissue, transcript abundance (Figure 3.28) did not follow any increasing or decreasing trend for both PS and NPS treatments. For *Tr-KPI5,* the transcript level increased significantly in the PS treatment when compared with the NPS treatment from a moisture content of 65% to 38%. At 32% moisture content, *Tr-KPI5* transcript levels decreased in both NPS and PS treatments compared with the initial moisture content (74%).

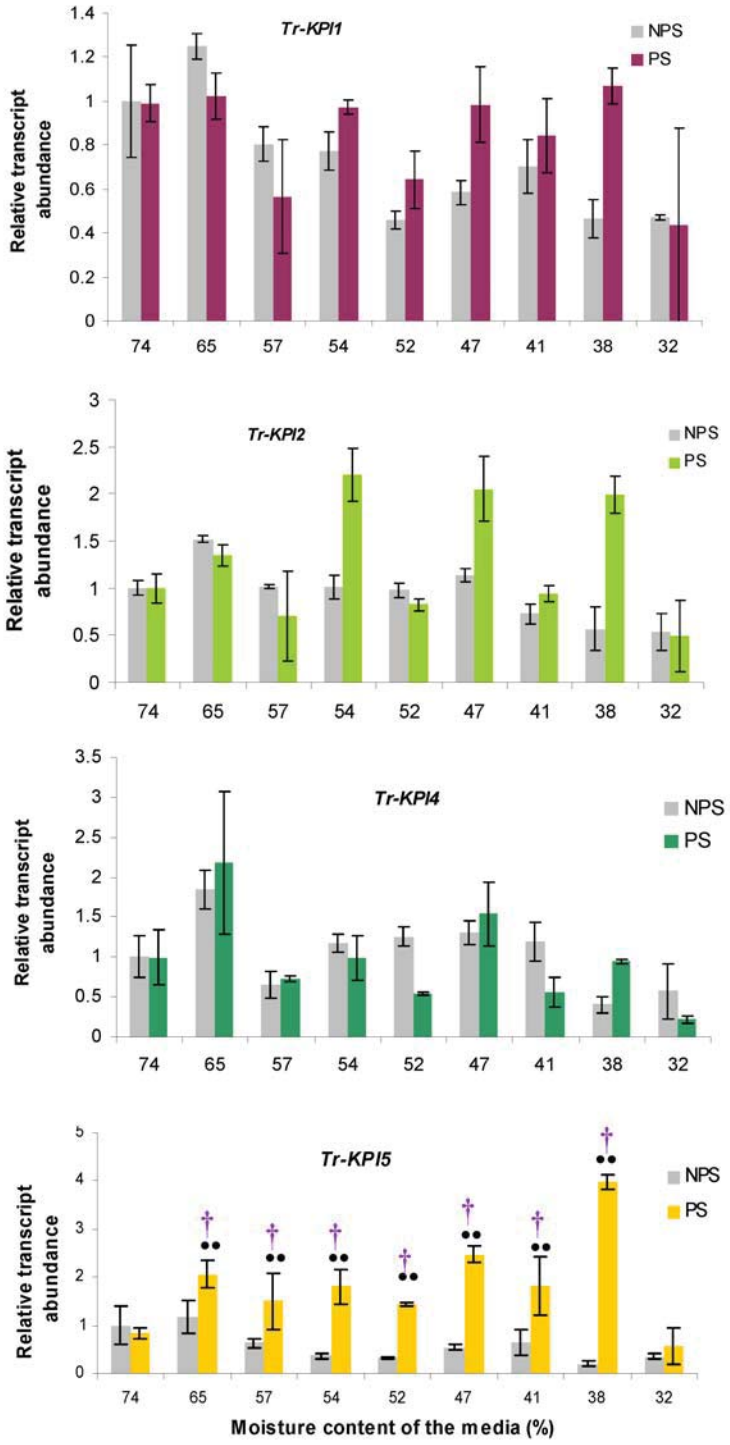


Figure 3.28: Relative expression of *Tr-KPIs*, as indicated, in root tissue of NPS and PS treated plants at the moisture content of the media, as indicated in cultivar Huia. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from at least four stolons were used for each data point and error bars representing mean ± SEM. Statistical analysis was performed only for *Tr-KPI5* .'••' indicate statistically significant (P<0.01) differential expression in comparison with initial moisture content and '†' indicates statistically significant (P<0.01) expression in the PS treatment in comparison with the NPS treatments using student's t-test.

3.4.1.5 Expression of Tr-KPIs in the first fully expanded leaf of white clover ecotype Tienshan and cultivar Kopu in response to a water deficit

The small-leaved white clover ecotype Tienshan, which is considered to display some drought tolerance and cv. Kopu, which is considered to be more drought susceptible, were used to observe changes in Tr-*KPI* gene expression in response to the NPS and PS treatments. The experiment was conducted previously by Dr. Alluh Nikmatullah in 2009 (Nikmatullah, 2009). The cDNA from those experiments were used to analyze the *Tr-KPI* gene expression level.

In the NPS treatment of Tienshan, the transcript abundance of *Tr-KPI1* did not vary markedly with a decrease in moisture content from 30% to 5% (Figure 3.29). However for the PS treatment, the transcript level increased by ca. 3-fold when the moisture content dropped to 26% from 30% and then remained at a ca. 3-6-fold increase as the SWC decreased to 16%. After this, expression increased ca. 12 to 15-fold as the SWC dropped to 6%. For *Tr-KPI2,* expression remained broadly similar in both the NPS and PS treated leaves, and also in terms of the changes in SWC (Figure 3.29). A single increase in expression in the PS treatment (from 30% SWC) was observed at 18% and 13% SWC, where difference between treatments was also observed, while a significant increase in the NPS treated leaves (from 30% SWC) was observed at 21% and 10% SWC. For*Tr-KPI5*, a significant increase in expression was observed at 18%, 16%, 10% and 8% SWC when compared with the initial moisture content (30%) under the NPS treatment. For the PS treatment, a significant increase in the expression level was observed when the moisture content dropped down to 23%, 21% and 18% and then again at 13% and 6% from the initial moisture content 30% (Figure 3.29). Differences between the treatments were observed at 23%, 21%, 18%, 13% and 6% SWC.

For cultivar Kopu, the *Tr-KPI1* transcript level showed an increasing trend from 29% to 14% moisture content for the NPS treatment, with significant differences from fully hydrated (29% SWC) at 23% and 14% SWC and a decrease at 11% and 9% moisture content (Figure 3.30). For the PS treatment, the transcript level increased significantly as the moisture content decreased from 27% to 9%. For *Tr-KPI2,* though expression was maintained at the same level for the NPS treatment, with a decrease in moisture content when compared to 29%, for the PS treatment, the transcript level was at least 0.5 to 1 fold higher when compared with the NPS treatment at the same moisture level. Interestingly, *Tr-KPI2* transcripts were undetectable at 11% and 9% moisture content for the PS treatment. For *Tr-KPI5,* the transcript level did not show any difference in the NPS treatment. However, for the PS treatment the transcript level increased 4-fold with a decrease in moisture content from 29% to 27% and maintained the same level up

to a 17% SWC following a decrease at 14% and 11% moisture content and an increase again at 9% moisture content (Figure 3.30).

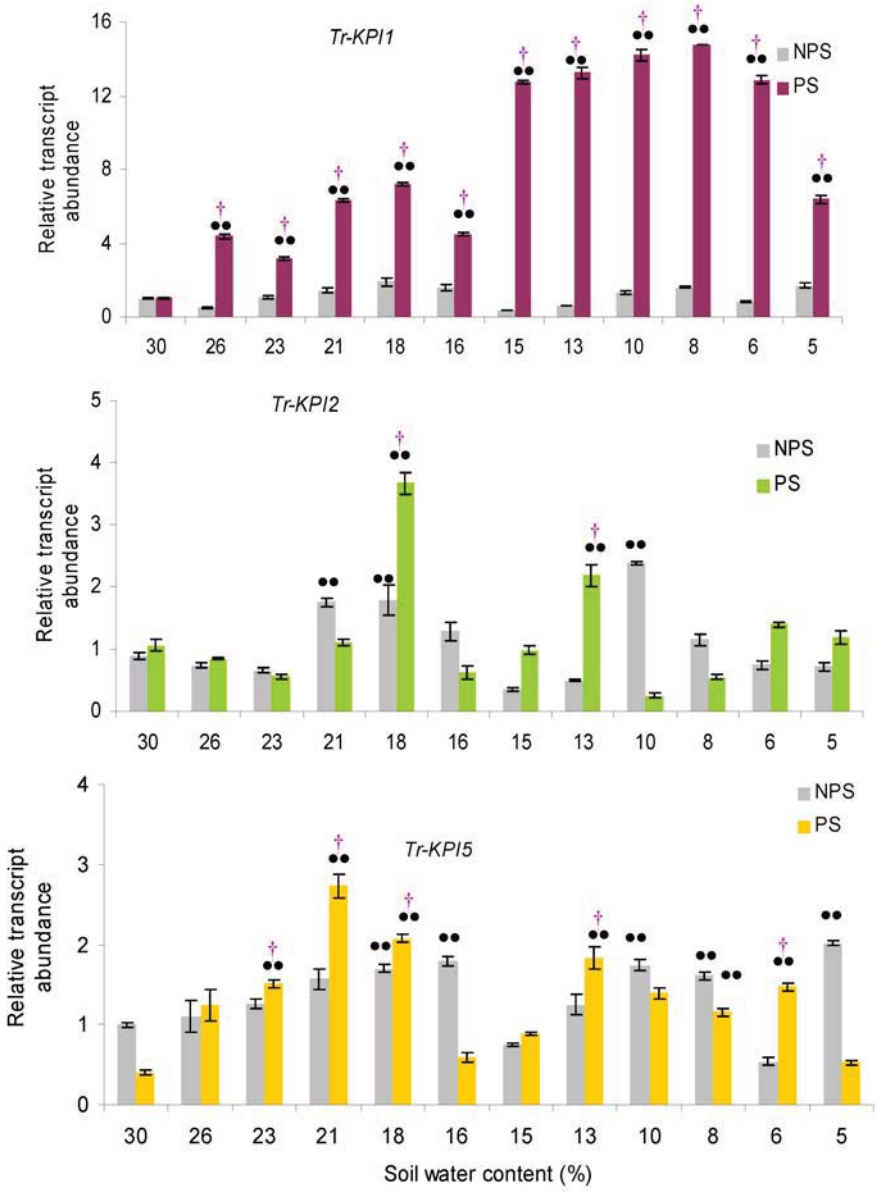


Figure 3.29: Relative expression of *Tr-KPI* genes, as indicated, in the first fully expanded (FFE) leaves of NPS and PS treated ecotype Tienshan at the moisture content of soil (SWC), as indicated. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene and error bars representing mean ± SEM. '••' indicate statistically significant (P<0.01) differential expression in comparison with initial moisture content and '†' indicates statistically significant (P<0.01) expression in the PS treatment in comparison with the NPS treatments using student's t-test.

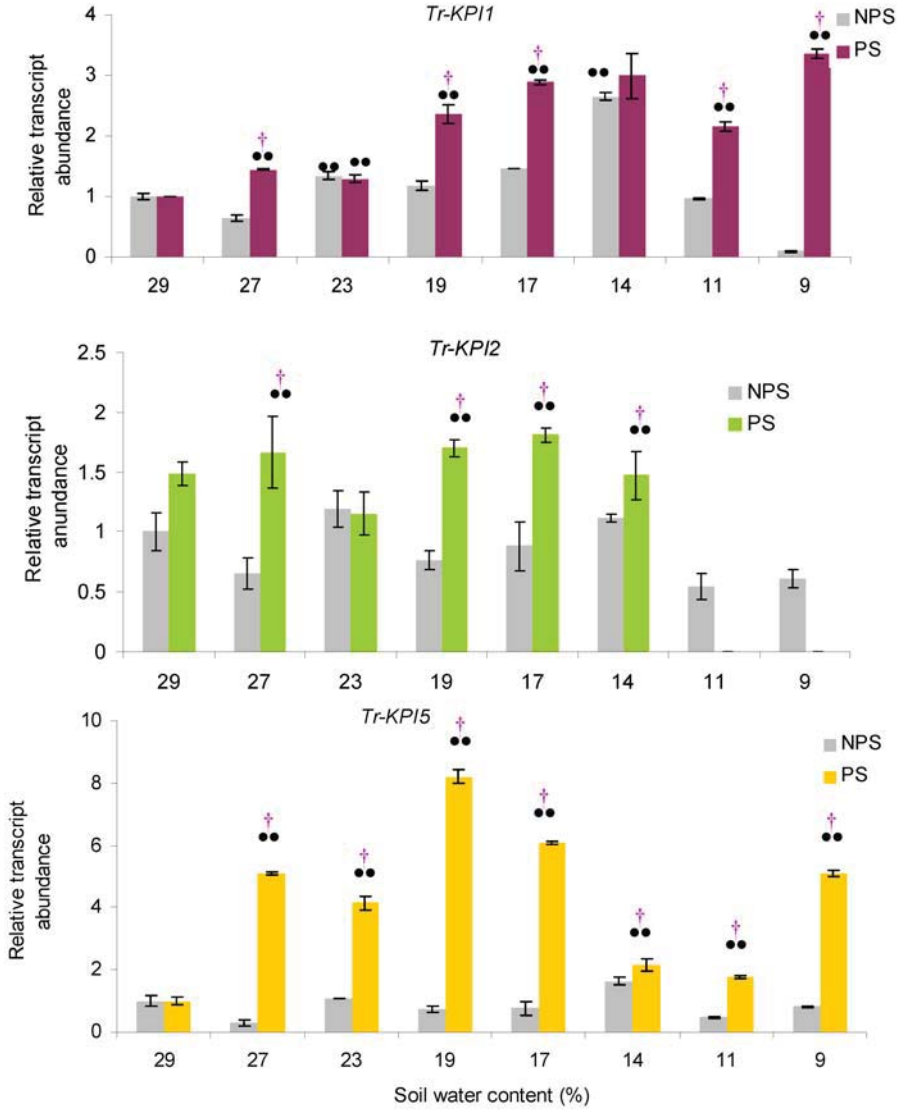
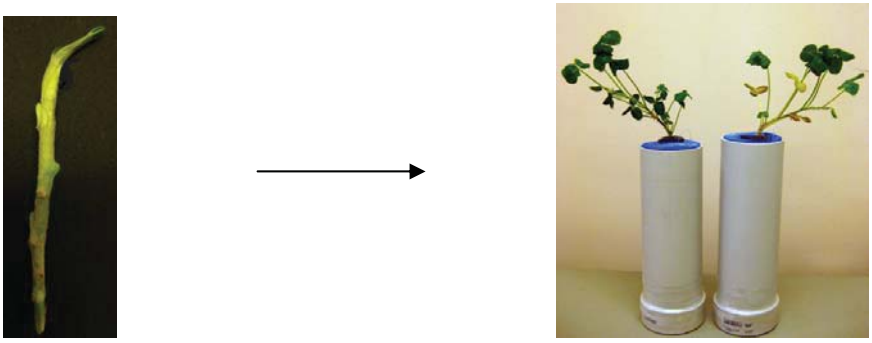


Figure 3.30: Relative expression of *Tr-KPIs*, as indicated, in the first fully expanded (FFE) leaves of NPS and PS treated cultivar Kopu at the moisture content of soil (SWC), as indicated. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene and error bars representing mean ± SEM. '••' indicate statistically significant (P<0.01) differential expression in comparison with initial moisture content and '†' indicates statistically significant (P<0.01) expression in the PS treatment in comparison with the NPS treatments using student's t-test.

3.4.2 Changes in the expression of the Tr-KPI genes in response to change in phosphorus supply

Phosphorus deficiency occurs mainly in low-input managed grasslands and may affect the growth of forage legumes particularly white clover. Therefore, changes in *Tr-KPI* gene expression in response to phosphorus deficiency was investigated in this study.

Stolons of the white clover cultivar QRT-PCR were cut at four internodes from the apex and rooted in vermiculite supplemented with 0.5 X Hoagland's medium (Appendix 8, Figure 3.30). After 5 days, the stolons were transferred to liquid 0.5 X Hoagland's media and grown at 22^0C with a 14 h light and 10 h night cycle for 14 days. For the phosphorus supply study, the stolons were either treated with Hoagland's medium with 5 µM P (Pi deficiency treatment) or 0.5 mM P (Pi sufficient treatment). The experiment consisted of two biological replications each comprising pooled tissues from three plants. The samples were collected at 0 h, 1 h, 6 h, 12 h, 1 day, 2 day, 3 day, 5 day and 7 day after treatment. To harvest the root tissues, the roots were divided into 3 different zones: (i) the elongation zone (EZ) which is 1 to 1.5 cm from the tip of the root including the root cap, root tip and zone of elongation, (ii) the zone of visible lateral roots (VL), a region falling in between the EZ and mature root with emerging visible lateral roots, and (iii) the mature root zone (MR) which is the zone of maturation with lateral roots excised, (see Figure 3.9). The first fully expanded leaf (FFE) was also harvested.



(A) Four noded stolon

(B) Stolon growing in 0.6 L PVC pipe during the study period

Figure 3.31: Experimental set up for phosphorus deficiency experiment.

In the P supply experiment, the qRT-PCR data for EZ showed significantly higher transcript abundance for all four genes studied in the P-supplied media over the time course of the trial, when compared with the 0 h level (Figure 3.32). Pi starvation significantly increased the transcript abundance for *Tr-KPI1* from 6 h onward to day 5 compared with both control (P-

supplied) and 0 h. For *Tr-KPI2,* Pi starvation decreased the transcript level upto 12 h and a significant increase was observed from day 1 to day 7 in the EZ tissue when compared with both the control and 0 h. The *Tr-KPI4* transcripts were highly induced at 6hr onward till day 6 when compared with both the control and 0 h. *Tr-KPI5* showed a similar pattern of transcript abundance as*Tr-KPI1,* where in the Pi starved tissue, the transcript level increased at 6 h and then from day 1 to day 5 (Figure 3.32).

In the VL zone, the transcript level did not vary significantly for *Tr-KPI1* in the control tissue except at 6 h when compared with 0hr (Figure 3.33). For *Tr-KPI2,* the transcript level varied significantly from 1 h and 12 h, for *Tr-KPI4* from 1 h to day 3 and for *Tr-KPI5* from 1 h, 6 h and from day 3 in the control tissue when compared with 0 h. In the Pi starved tissue, though the transcript level of *Tr-KPI1* and *Tr-KPI5* was significantly up-regulated at 6 h to day 1 compared to both control and 0 h, the transcript level did not change much from day 2. On the other hand, the response of *Tr-KPI2* and *Tr-KPI4* was much more prominent when compared with *Tr-KPI1* and *Tr-KPI5* (Figure 3.33) in the VL zone when the plants were exposed to Pi starvation. The transcript level of *Tr-KPI2* started to increase after 12 h when compared with 0hr of the treatment and remain significantly higher up to day 7 when compared with both 0 h and control. In the case of *Tr-KPI4*, the transcript level started to increase after just one hour of exposure to Pi deficiency and maintained a higher level up to day 3 when compared with both control and 0 h.

The expression level of the *Tr-KPIs* in the mature root zone control tissue varied significantly from 0 h. For *Tr-KPI1*, upregulation in the control was observed from 1 h to day 7, for *Tr-KPI4* from 1 h to 12 h, day 2 and day 3 and for *Tr-KPI5* from 1 h to day 7 (Figure 3.34). For Pi starved MR tissue, the *Tr-KPI1* transcript level significantly increased from 1 h to day 7 when compared with 0 h, and day 2 and day 7 when compared with control tissue. For *Tr-KPI2*, the transcript level significantly increased at 1 h to 12 h and day 2 when compared with both the control and 0 h. In the case of *Tr-KPI4*, the transcript level significantly increased at 1 h and 6 h when compared to both control and 0 h, and at 12 h and day 3 when compared to 0 h (Figure 3.34). For *Tr-KPI5*, the transcript level did not show any significant variation when compared with the control in the Pi starved tissue, but a significant level of increase was observed from 1 h to day 7 when compared with the 0 h time point.

Initially, Pi deficiency did not result in any increase in mRNA abundance for the *Tr-KPIs* in the first fully expanded leaf (Figure 3.35). In both P+ and P- leaf tissues, the transcript level

decreased when compared with 0hr. However, for *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* the transcript level increased significantly from day 1 to day 3 when compared with the control tissue.

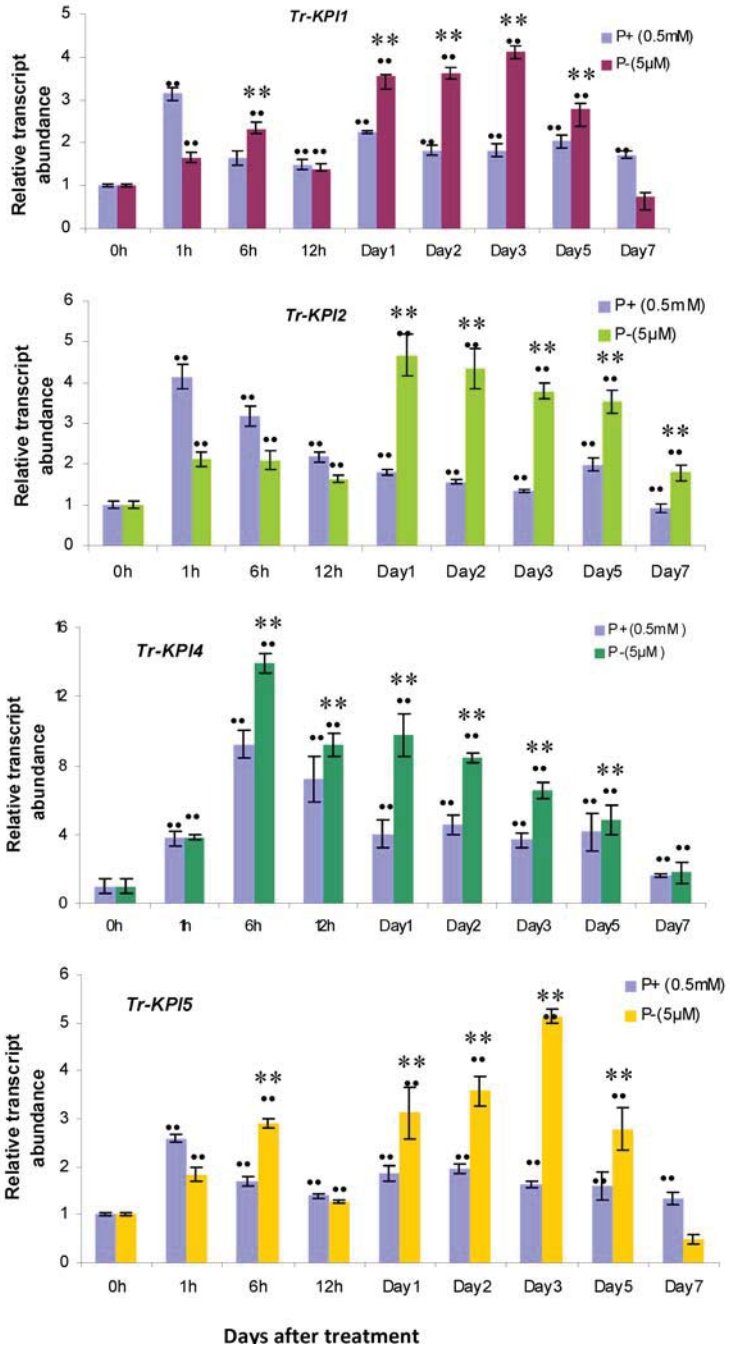


Figure 3.32: Relative expression of *Tr-KPIs*, as indicated, in the elongation zone (EZ) in control (P+) and Pi decificient (P-) treated plants using qRT-PCR. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from three plants were used for each data point and error bars representing mean ± SEM. '**' indicates statistically significant (P<0.01) differential expression in comparison with untreated plants and '••' indicated statistically significant (P<0.01) compared to 0 h using student's t-test.

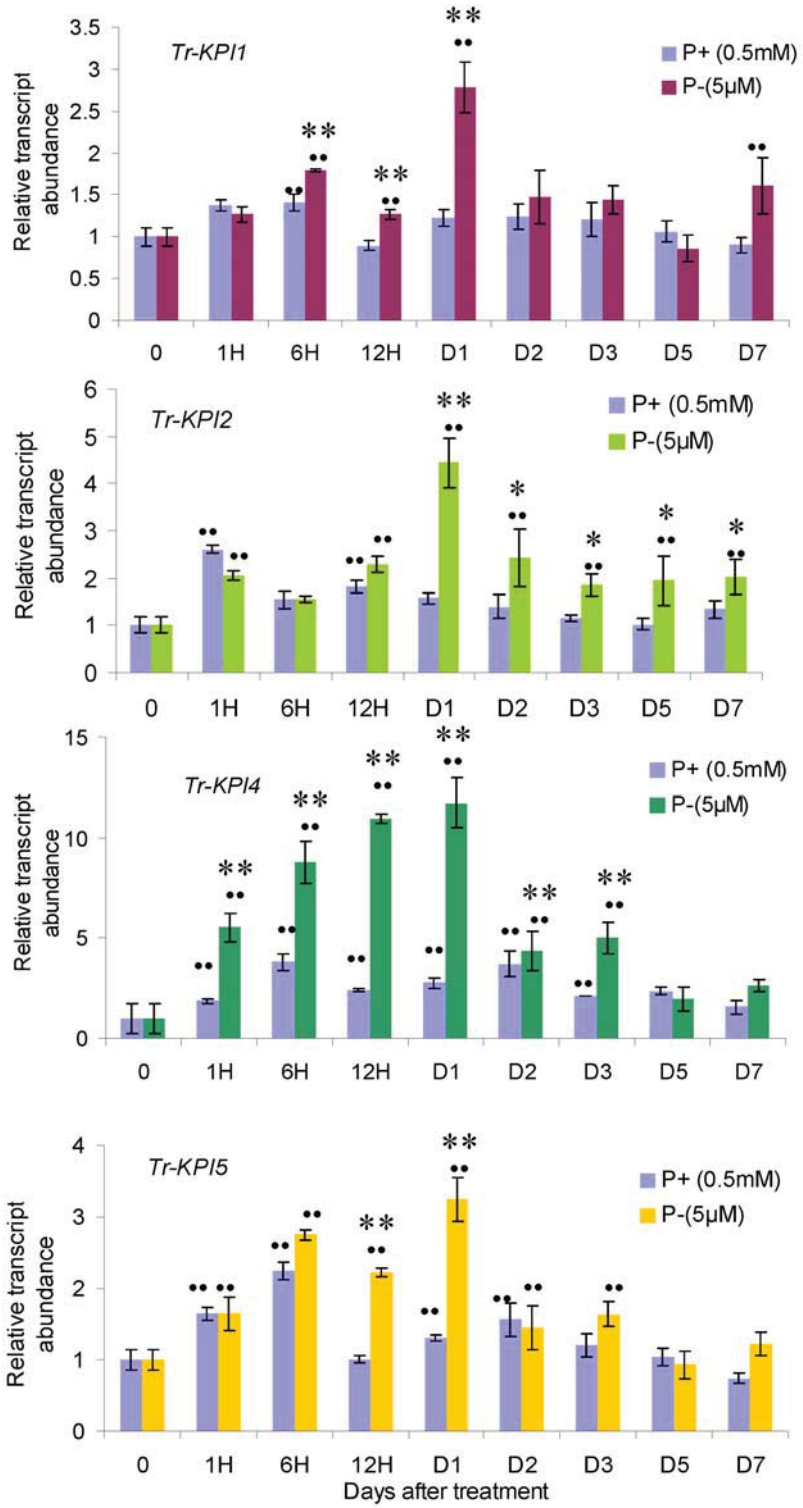


Figure 3.33: Relative expression of *Tr-KPIs*, as indicated, in the visible lateral root zone (VL) in control (P+) and Pi deficient (P-) treated plants using qRT-PCR. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from three plants were used for each data point and error bars representing mean ± SEM. '**' indicates statistically significant (P<0.01) differential expression in comparison with untreated plants and '••' indicated statistically significant (P<0.01) compared to 0 h using student's t-test.

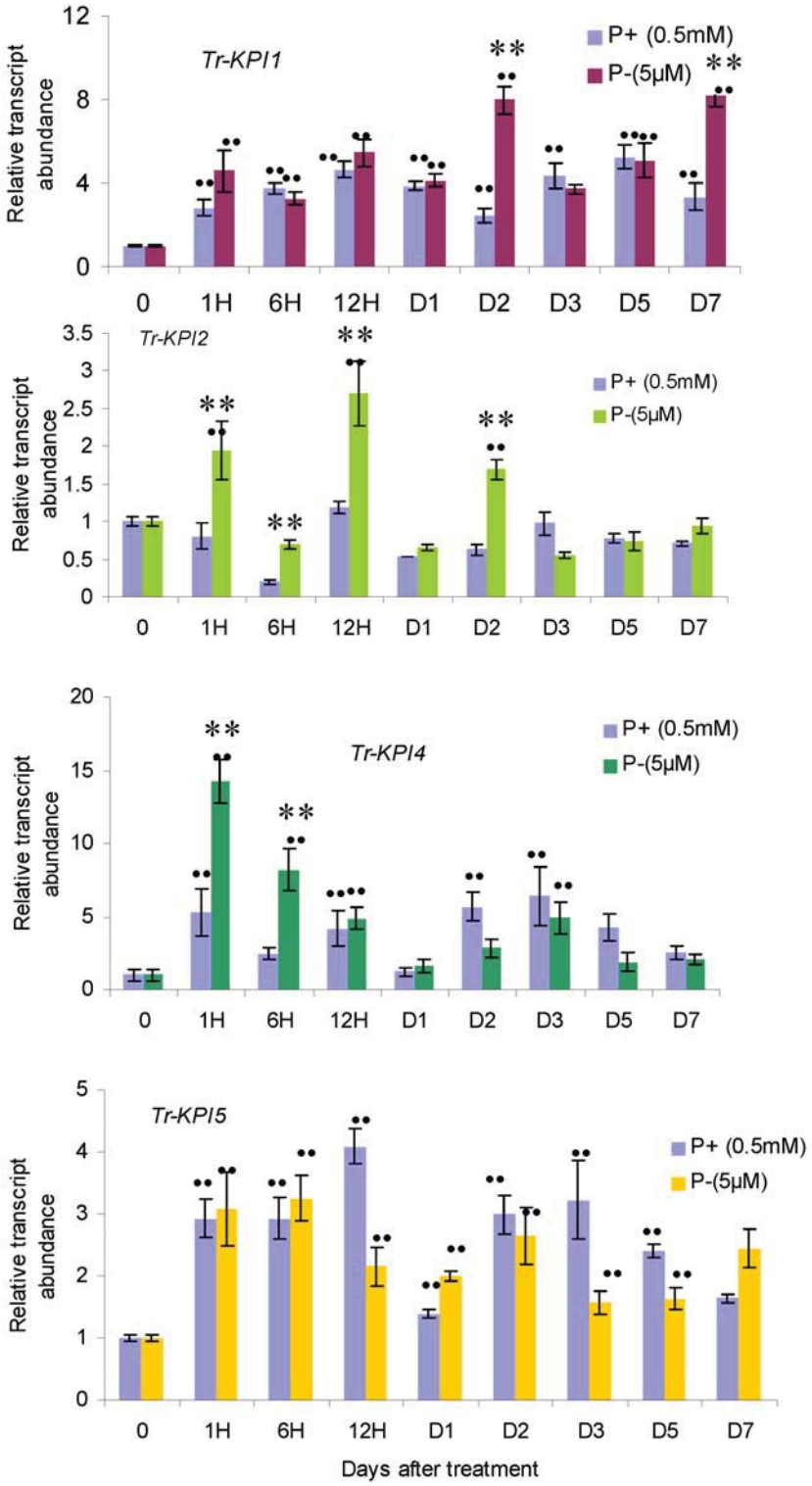


Figure 3.34: Relative expression of *Tr-KPIs*, as indicated, in the mature root zone (MR) in control (P+) and Pi deficient (P-) treated plants using qRT-PCR. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from three plants were used for each data point and error bars representing mean ± SEM. '**' indicates statistically significant (P<0.01) differential expression in comparison with untreated plants and '••' indicated statistically significant (P<0.01) compared to 0 h using student's t-test.

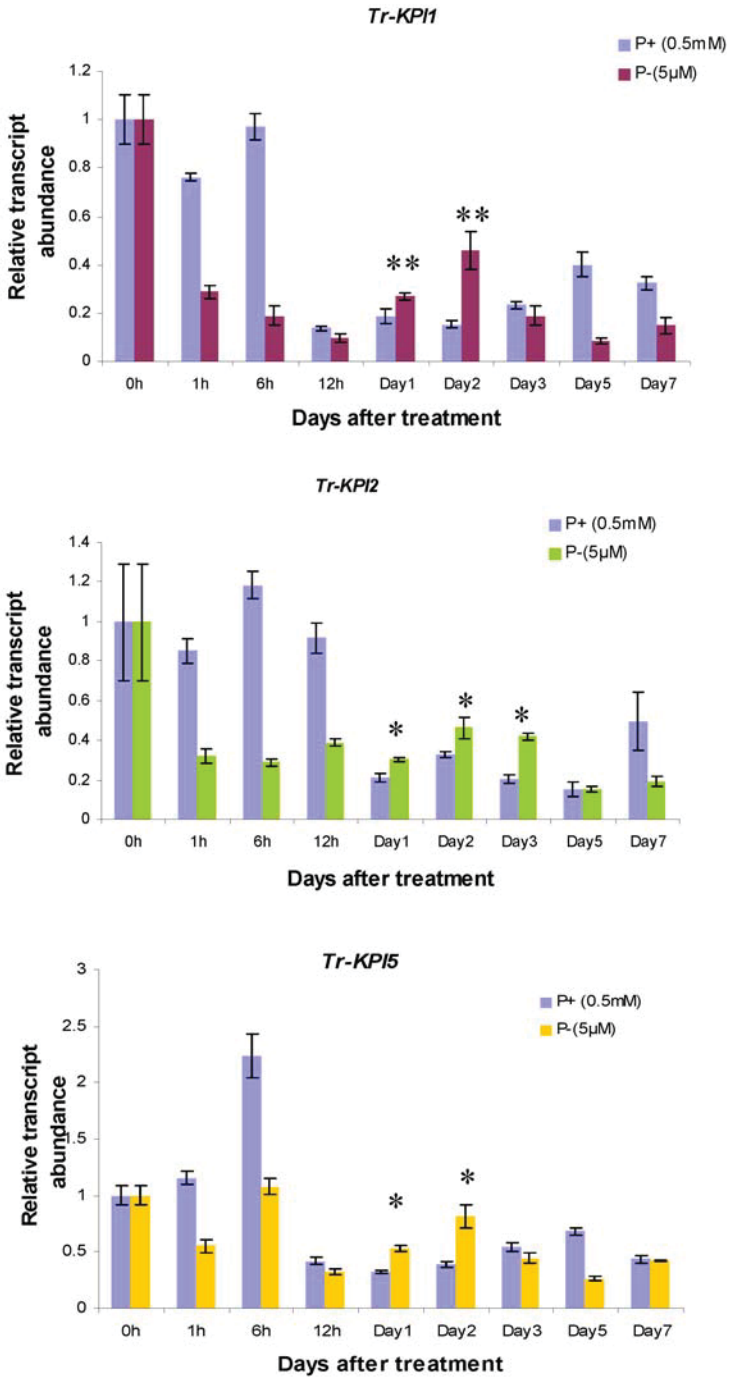


Figure 3.35: Relative expression of *Tr-KPIs*, as indicated, in the first fully expanded leaf (FFE) in control (P+) and Pi deficient (P-) treated plants using qRT-PCR. Results represent qRT-PCR using *β-actin* and *GAPDH* as reference gene where two biological replications each consisting of pooled tissues from three plants were used for each data point and error bars representing mean ± SEM. '*' indicates statistically significant (P<0.05) differential expression in comparison with untreated plants using student's t-test.

3.5 Exploration of Tr-KPI Promoter Sequences

The results of the developmental, biotic and abiotic stress studies suggest that the expression of the Tr-*KPI* genes in white clover must be controlled by complex cis-regulatory elements. Therefore, focus was given to isolate the promoter regions of these genes and to analyze putative transcription factors binding sites (TFBS).

To identify the 5'-UTR and the promoter regions of the selected *Tr-KPI* genes, the Genome Walking^{TM} protocol was employed using nested PCR primers and three different genomic DNA libraries with three blunt cutting restriction enzymes (chapter 2.3.5.1). Using this procedure, a 1.89Kb region for *Tr-KPI1* , 2.3Kb region for *Tr-KPI2*, 1.7Kb region for *Tr-KPI4* and 1Kb region for *Tr-KPI5* was isolated (Figure 3.37, 3.38, 3.39 and 3.40) and sequenced upstream of the ATG. To further confirm that these regions did corresponded to each *Tr-KPI* gene, the full sequences including the promoter sequence, and regions of the 3'UTR was isolated (Figure 3.36) by PCR and sequenced using the PFL (promoter with full length coding sequence) forward and reverse primer.

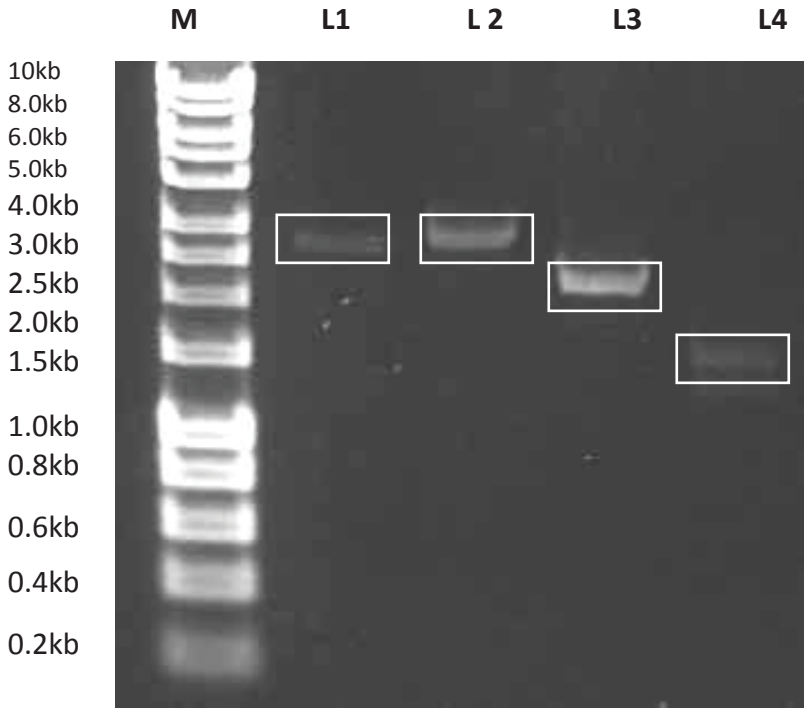


Figure 3.36: Isolation of full length *Tr-KPI* genes, as indicated, spanning the promoter region, reading frame, 5'UTR and part of the 3'UTR region using PCR.

PCR fragments were generated using *Tr-KPI* forward and reverse primers (PFL primer sets, Appendix 6). PCR products were separated using agarose (1%, w/v) gel electrophoresis and visualized after ethidium bromide staining.

M: Markers, with the relative size of the bands is indicated in the left hand side.

L1: The promoter region and part of 3' UTR region of *Tr-KPI1* (2.76Kb)

L2: The promoter region and part of 3' UTR region of *Tr-KPI2* (3.052kb)

L3: The promoter region and part of 3' UTR region of *Tr-KPI4* (2.33kb)

L4: The promoter region and part of 3' UTR region of *Tr-KPI5* (1.787kb)

>Tr-KPI1

TCTATGGCCTCAAGCAAGCTCCTCGCCAATGGTTTGAGAGGCTTCAGGGTGCTCTGCTGCAACTTGGATT
TTTGTCAAGCAAGTGTGATCCTTCCTTGTTCACCTATAATTCCAAAGGCAATACAGTCTATCTGCTTGTG
TATGTAGACGACATTATCATCACCAGCAATAATTCTTCTCTGCTGCAGTCTTTCATTACACAGCTTAATC
AGGCATTCTCTCTCAAGCACCTTGGCTCACTTGATTATTTTTTGGGTATAGAGGTTAACCAACTTCCCAA
TGGCTCATTACTTCTTACTCAGTCCAAATACATTCGTGATCTTCTCAACAGGACAAACATGCTTGAATCA
ACACCTGTATCTACTCCTATGCAGTCCAGTTGCAAGTTGAGCAAGGATGGCTCCCCTGCTCTATCTGACC
CCTTTATGTACAGGTCTGTGGTTGGTGCTTTACAATATGCCACTCTTACTCGCCCTGAGATTTCTTATGT
TGTTAACAAAGTGTGTCAGTTCATGTCTTGTCCTTTGGAGGCTCATTGGGTTGCTGTCAAGCGTATTTTG
AGGTATCTCAAGGGTACCCTTCATCATGGTCTCTGCCTATCCCCTGCCATTCCCTCTACAGCTCCCTCTC
TCAAGATGTTTTGTGATGCTGACTGGGCCTCCGACCCTGATGACAGGCGTAGTACCTCTGGTGCTGCTCT
TTACTTTGGTCCCAATCTCATTTCTTGGTGGTCTCGCAAGCAGCTTGTTGTTGCCAGATCAAGCACCGGG
GCTGAGTACAAAAGCCTTGCCCATGCCACTGCTGAGCTACTCTGGGTTCAAACTCTGCTGGCTGAACTCC
TTGTATCATTTACCTCGCCTGTCATCTATTGTGATAATCTCTCTGCTGTTTCTCTGGCCCATAATCCTGT
CATGCACTCCAGGACCAAGCACATGGAAATTGACCTGTTTTTGTTCGCGAAAGGTCCTCTCAAAGCAACT
TTCTGTGCTGCATATTCCTGGTACTGATCAGTGGGCTGATGTTCTGACCAAGCCTCTCTCCTCTTCCAAA
TTTCTTGAGCTCCGGTCCAAACTCAATGTTACCTCTGGCAGGCCCCCTTGAGTTTGAGGGGGGAGTATTA
GAGATAAGGCTAGCTTAGTTCTGTAGATTATGTAAATAGAACTATATGGTATAGAAATGATACAGCTAAG
ATAGAATAACTAACCTAAGAGTTAGTTATGCCAGCTGGCATGCCAGTTGTATAGAACTGACAGGTGTCAG
TATCTGATTACTTGTAACCGGCTATAGCAGGTAGCCAAGTGTGACTATAAATGTAACTAGTCAACACTCT
CTGTAACTAACCAAAAAAACAGAATGCAATGAAATGAGATTCTTGAATCACTCTTTACTTTCTATTAAAA
TCCAATTTTTGAAAGACTATATTTATACGCTTAAGCGACACAACATTACTATACATGTCATTTTGAACTT
TTTTTGAAAGATTTACAGTCTTTGATACTAAAAACCAAAAATATCAAATGTAAATGTATAAAAATATGTT
AATTGTCAGAGCACGCTTTATCATATTTAACCAGATGATGAATTATATTAGTTTTTGTTAATTTATTTAT
CTTTATGAAAAGAATGGGAGCTGGTAGTAAAAAATAAAACCTGATATTGAAGATCTTATCCAGTAGCTTG
CCACGTGAGGAAACATTTGTGTCACATTATATATTAAGAAATATAAGCATAATAATTGAATCAATAAACT
AGTGAAAGTTGAAAGAGCTCTTTAATATTTCCTTACAATTATTATATATAGAAGCAGATGCAACCCACCA
AAACACAAACAAATTATCAAGATTAAATCTTATCTAACCATGAAGCATGTTTCATCACTCACCCTTTCCA
TTCTCTTCTTTGTTTCCATCACCAATCTTTCACTAGCATTCTCAAATGAAGATGTTGAGCAAGTGCTCGA
TATAAATGGTAACGCCATCTTCCCAGGTGGCGAATACTACATTTTACCAGCACTTCGCGGCCCAGGAGGC
GGAGGAGTAAGGATTGGCAAAACCGGAGACTTAAAGTGTCCGGTTACTGTCTTACAAGATCGTAGAGAAG
TCAAAAATGGTTTACCAGTGAAGTTTACCATACCAGATATAAGTACTGGTATAATCTTCACTGGTACACC
AGTTGAGATCGAGTTCTTTAAGAAACCTAATTGTGCTAAATCATCGAAATGGTTGGTATTCGTTGATAAT
GTTGTTAAAAAAGCTTGTGTTGGTATTGGTGGTCCTGAAAATTATCCTGGTGTGCAAACATTGAGTGGAA
CATTTAATATTCATAAACATGAATCTGGATTTGGTTATAAACTTGGATTTTGCATTAAGGGTTCACCTAC
TTGTTTGGATATTGGGAGGTATGACAATGATGAAGCTGGAAAACGTTTGAATTTGACTGAACATGAGTCT
TATCATGTTATCTTTGTTGATGCTGCTTCTCATGAAGCTGATCAGTATATTAAGTCTGTTGTTTAATTTA
TAATTTATAATTAGCCATAAAGCTTATCATATGTAGCTTTTGTTTTTGCTGTGTTGAACTTTGTTGTTGC
ACAATTTAAGCTATAATAAGTGGAGTAAGATATTTTTACATTTTTTGTGTTGTTGATTGATTTATTCCCT
TTTCTAATTGTTCTTTTCTATCTATTATTCCTTTTCCAAATTCATAATTTAATCTTTTAGAGGAGAAAAC
TAAGGGTTGAGTTAAGTTTAAAAGAGAAAAGATAAGATTTTAATTTTATTTTCTTTGTTATATATTGGAA
AATTTAAAGATTTCAACCAATTAATTCCT

Figure 3.37: Full length *Tr-KPI1* gene, as indicated. The yellow highlighted region indicates the start and stop codons, the red highlighted region indicates the position of transcriptional start site (TSS) which is identified using FGENESH (http://linux1.softberry.com), the blue highlighted region indicates the presence of the GSP1 primer set and the grey highlighted region is indicating the presence of GSP2 primers sets. The green highlighted region is the position of the primers sets used to amplify the whole sequence from genomic DNA.

>Tr-KPI2

ACTCAAAAGGGATACAAAACTCATATCTTTTTGTATATGAGTTAAAATACAAGAAGGATAGTTAGAAACC
CAATTCCTATTGGGCTCTCTCAAAGCCCATCGAGCTAAAGCATGGGCTGCACGATTTCCATCTCTGCTGA
CCCAGGAAACAGAAACGTTGTTCAGTTTGCTTAGAACCCTAGAGCTGTTGCGAATACTCTTCCCCCAATT
CGTTCTGGGAAAATTCTTCCGAGCAATGGCGTTAACAATAATCGCAGAATCGAGTTCAATCGAGGTGTCT
TTAAGCTGGTTCGCTTGAGCAAAGAGGATTGCTTCATGAAGACCTGTAGCTTCCGCCATAGCCGCATCAT
TCGGTCCTTCACGCACTCTAGTCGCCGCTCCAACACAGCGGCCATCATCTCTTCGCAGGATCAAACCAAA
ACCCCAACGGCCATCATCTTTAAGATGAGCGTCCACGTTGAGCTTATGGCAGTTCCGGGAAGGGGGATTC
CAGCTCTTGTTGTTACGGGTTCCAGAAGTGGCGTGAGTTGATTTGCTAAGGCGGTTTTCGATCCAGTGCC
GGCGATACTCACTGAGGTTCTTCATAGCTTGAGCCACTACATCCTTTGCAGGTATGCTCTTGTTTTGAAA
AATTCTTTTGTTTCGAGCTATCCATATGCTGTATATGACTGTAGAAATAGTTTGCATGCATTCTTTGGTG
GAATTAATCAGCATGTAGCTTAGCCAATCTGGGAAACTCTGAGTTTGGACGTTCGATGTAGTGATAGTTA
GGGGACACCCAAACCAAACCTGTCTAGCCCACTCACATTGTTGAAAGATGTGATTTGTAGTTTCCGGTTC
TCTGTAGCACATAGGGCACATAGAGTCACACATAATGCCTCTATTCAGAAGATTTATCTTAGTGGGAATA
GCATTGTGCAGAATCCTCCAAAGTAGGTGTAATTGTTTTGGAGGGGCTTCAATTTTCCAGAGTTTGTTCC
AAATGTGAGCTCCCTGAAGTAGATTACTAGGTTGAGCTTGAGCAGAGGTTGTGATTGCCCACTCCATTTG
GGCGTTATACCCAGATCTTACAGTGTAGTGACCATCTTTAGTACCCTGCCAACAGATTAAATCATCTTCA
GAAGTGTTTGTGAGGGGGAGTTTCAAGATGTTATCTGCTTCTATAGGGATGAAATTCTGGGAGATGAGTT
GGGAATTCCATTGTTGACTGGTATGGTCAATAAGATCACTGACATGCTCCAAGTTAGAAGTGGTAGGTTT
GGTGCTCCAAGTGGTGTTCCCATCTTTAGGACTTATCCATCGATCCTCCCAAATGTTAATGTTCTTACCA
TTCCCAACCAGCCAGTAGCATCCTTTTTTTAGAATCCAACTGGCCTGTTGGATGCTTTGCCAAGAGTAGC
TGGGTCTGTTGCCTTTCCTAGCTTGTAGGAATTGGTGTTTAGGGAAGTATTTAGCTTTGAGGGTTTTTGC
CATAAGGGAGCCCGGTTCAGTTAAAATCCTCCATCCTTGCTTTGCCAAAAGGGCTTCATTGAAAGCCCTG
AGGTTCCTAAAACCCATGCCACCCAGTGTCTTCTGTTTGCAGGTTTTTTTCCAATTAACCCAGTGTATTT
TCCTTTTGTCAACATTGCTGCCCCACCAGAAATTGCACATCATGCTTTCCAATTTGTCACACAACCCTTT
AGGGATTAGAAAATTACTCATCAGGTAAGTAGGAATGGCCTGAGCAACTGCTTTTATAAGGGTGCCTCTT
CCTGCAAAGGACAGATTTTTTTCTTTCCAACCTTTCAGTTTTTTCCAGAGCTTGTCTTGAATAAAATTGA
AAACTTGAGCTTTAGATCTCCCAATTTGAGTTGGTTGCCCAAGGTATTTGGGGTAGTGGCTGACAATAGT
CATTGGGAGAATTTGTTGGATATCTTGCTTCATAGAGTGCTGAACCTTTTTGCTAAAAACCAGTTCAGAT
TTGTTATAGTTGACCAGCTGTCCAGAGGCCATTTGATATTGTGTGATGATTGACTGAATTTGGGTAGTTT
CTTCCTTGTTTGCCCTACAGAACATAAGACTGTCGTCAGCAAAGAAAAGATGAGTTATCTTTGTTAATAG
ATAAAACCAATTTACTAGTATTATTTTTGTTGGATTAAAACCAGATTTATTGTAGTTCTCTTTTTATTTT
GGTAAGTATTGTAGTTCTCTTATTAAAATATGGTCGTGCACCTCGTGAAGGTATAAAAAATGCCACACAA
ACTTATCAATCGAAAAGATAATAGGAGTAATAAAAAAGTGAAAAGTTGTGCCCTTCGTGGAACGCTTGCA
ATTTGTTGTAGAAAAAAACACAAATAAGGTATTCTTCCCTTCTGCGCACCAAAACACAAACAAAAAACAT
TATCTAACCATGAAGCCTATGTTATCACTCATCTTTTCTTTCTTCCTCTTTATTTTAATCACCAATCTTT
CACTTGCTTTATCAAATGAAGCTGTTGAGCAAGTATTGGACTCTCTTGGTAACCCCATTTTTCCCGGTGG
AAAATACTACATTTTTCCGGTATCTCATGATGAGACATATGGTGGAGGATTAAGACTAGCAAAAACCGGC
GATTCAAAGTGTGATGTTACTGCCTTACAAGATGACAATATAGTTATCGACAGTATACCAGTGAAATTCA
GTATACCAGGAATAAATCCTGGTATAATTTTCACTGGTACGCCGATTGAGATTGAGTTTACAAAGAAGCC
TAGTTGTGTTGAATCATCGAAATGGTTGATATTTGTTGACGATGTTATTCAAAGAGCTTGTGTTGGTATA
GGTGGTCCTGAAAATTATCCTCATTTTAAAACATTGAATGATGGTAGATTTTATATTGAGAAACATGAAT
CTGGATTTGGTTACAAGCTTGGTTATTGTGTGAAAGATTCTCCTACTTGTTTGGATATTGGGAGATCTGG
TAATGTGACGGAGGAAGGTGGGTTTCGGTTGAATTTGACTCATCAAGTGGCTTATGCGGTTGAGTTTATT
GATGTTACTCCTTTTGAAGCTAGAATTAAGTCTGTTGTTTGATACCTTGATTGTTAGTCTAACTAGTAAT
GTAATCTTGATTCTTGTTTTGAATGGCTATAAGTTTGCATAATATGCAAACATTGAATAAGTAATTGTTG
AATACTCTATGTTATATTTACTCTTTAATCGATATCTAATTGAGC

Figure 3.38: Full length *Tr-KPI2* gene, as indicated. The yellow highlighted region indicates the start and stop codons, the red highlighted region indicates the position of transcriptional start site (TSS) which is identified using FGENESH (http://linux1.softberry.com), the blue highlighted region indicates the presence of the GSP1 primer set and the grey highlighted region is indicating the presence of GSP2 primers sets. The green highlighted region is the position of the primers sets used to amplify the whole sequence from genomic DNA.

>Tr-KPI4

AAAGTTGAATATGCACATCCTGGTACAAATAAAAAATAAAACATATTAAGATTGATTCAAATATAAAAAC
AAAATCTAAATTTAACAAAAATAAACCATATTGCTGTGGAGTGGAGAAAACTATCTAAATCCCACAATCA
ACTCCTAAAATGTAGGGATTATTAGCTAAAAGATAGAAACAAATCCTAATAAATATGACAACTAAGCAAC
AATAGGCAAGGTCAAAAAATAGAAAACCATATAATACAAATAAATCTCAATTGGTTGATTTCTAATAAAA
AATTCAATACCCAATACAAATAAAAAACTACAGTGGTATGATTTCAAATAGTTTAGGGATTATTGGTTGA
TTTCTAATATCAATTATGTATTTTGAATGAGAGCAGTAACTACAAATATACATACCATGAATCATGAATA
TAATAAAAATAAAAAACTACAAATGTTGCTGCTTTCAAAGACAAAATTCCAGAAAAAAAATTATGAATCT
ACTTCCAAATTCATCAAACTAAAGCTAATACTGTTCAATGACATACCTTTTTCCTCAAATTCATGACTCA
TGAGAAGCAAAAATTGGGACTGCACATGAAAAAAATAAATAACTATGTGAGAAATGAAGAAGCAACTATA
TAATACTCAAATTAGGTAATAGTATTCACAAAAAACTGAAATCAAACCTAATTCATAGGAGAAAATTCAA
CAATTAAAGTTCACAATACTTACCCAAATCAAACCCTAATTTGTTAACCTGAACAATAAAAATAAAACAT
TGGAAGAAGAAAAAAATCAAAATTAGAATATGAAAACAAATCATGAAATAAGGAAGGATTAAGAAACAAT
AGATTTAGAAAATCAAAAGACAAACCTTGACCATGTACCTTTTTCCTCAAACCCATGAGTAACAATCTCT
TATTCACATGAAAAAGAAAGATGTGAGAAATGAAAATGAAACTGGAAATCAAAAACAAAAACAAAAATGA
ATTCGAAGTGCTAACCTTAGCTTCATCCGTATTATCTGCTTCATCATTCTCTCCCGTTGGGTTGATGTCA
AAAGAGAGGGAGAAATGGAATCCGAAAATTTGTGTGTAGAGAGGAAGAAGAGAGAGTGAGAGATTGAAAG
GGATTTTGTGTTGTAGAAGGTGTTTGAATTTCAAAACTTCCTGCTTTGTTTGTGTCAGGTTGAATATGAT
AAAAGCCTTCCAACATGCTGAAGGTTAGAGAAATTTGTTGTCTGCTAGTGAAAGTGAAAGTGACCGGACC
TAATTACTTTTAGACTAACTGGACCATTATTGGACTGGGCTGGAGAACAAAGCGGCCCATTTAAGCTTTT
CTTCATAACAATTTCCCTCTCTTTTCATATTCAAAAAAATTGCCTAAAAGAAGCTAAGTTTTTGGTATTT
ATTGCTAACTAAAATCCTATGTGGCATCTCTTATAGATTGACACTTGTTTTTTTTTGCTAAAACCTTGGT
GCTGCTTGTGTAGTTTTAGGATTACTTGCATGTGCTCTATAGTATTATTAATATATTAATTTAGATAGAC
ATGAACGATTTCTTAATTATACATAGAGTAATATTTTTTAATACCCATGCAGAGTAATATTAGTAGTATG
CCCAAGCTAACCTTAAATCAAGTCTATAAATAGAGATTTAAGACATGAACTGATTACAAAACAAATACAC
ACATTCATCGCTTAACAACCATGAAGCCTACAATGCTTACCACCCTTTCTTTACTCCTCTTTGCCTTAAC
CACCTACTTTCCATTAGCTTTTAGTTCTAACGAACAACTAGCAGACTTGAATGGAGACCCCATCTTTTAT
TCTACTCGTTTCTATATTATGCCATCTATCTTTGGAGCTGCAGGTGGTGGACTCAAGCTTGGTAAAACTG
GAAAATTGACATGTCCACTTACTGTACTTCAAGATTATTCTGAAGTTATCAATGGTCTGCAACTAAAATT
TACCCCTCCAGGTGAAATTTTCGTTGATTTGATAAGTACAGACCAACCACTGAAAGGTATTGAATTTGTA
GAGAAGCCAGAGTGTGCTGAATCCTCCAAGTGGGTGGTGGTCGAAGACGATGATTTCCCTCGACCATATG
TCGGTATTGGTGGTATTGAAGACAATAAAGGTAAGAGGATCATAAATGGTAGCTTTAAAATTGTGAAACA
TGGTTTTGGATACAAGATTGTGTTTTGTCCTCGATTCACTGCACCACCTGGTCTTTGTTTTGATATTGGA
AGGCATGATGATGAGAATGGAAGGCGTCTCATCCTCACTGAAAATGATCCTTTTGAAATTGTCTTTGTGA
TTCCTAGAAGATCTGTTGCTTGAATTAAATCTCCAAAATAAAACTAAATAAATTATGTACACTTGAGAGG
TTAGCTGATCACTTGTAAGACTTTAACCTTGTAATCTCAAGTTACTCCGTTTAAATTTCTTCTAAGACAG
TTGTAAAATGACCATTAGTGTCATTTTAATTAAC

Figure 3.39: Full length *Tr-KPI4* gene, as indicated. The yellow highlighted region indicates the start and stop codons, the red highlighted region indicates the position of transcriptional start site (TSS) which is identified using FGENESH (http://linux1.softberry.com), the blue highlighted region indicates the presence of the GSP1 primer set and the grey highlighted region is indicating the presence of GSP2 primers sets. The green highlighted region is the position of the primers sets used to amplify the whole sequence from genomic DNA.

>Tr-KPI5

AACCACTTGATGCCAGAATTGATCCCCGAATCAGATTAAATTGGTGGTGCAAACCAAAAAAAAAAAAAAT
TTATTTAGGTTTGTGATTATATTARAATTTCTCATAAATGTTACCCTCTCTTTAATTAGCAATTTTCTTT
TTCATGTTTTCGTTATATCTTAATATCAAGATTGTTTACACAATTTCATTTTCATATGTACATTAATAAT
TTATTTACTAAAATATGTTTATCTGTCAATAAGCATAGCTTTTCACGTGGTTATATTTTTTTTTGAATGT
AGAGGTTGAGTTACAATAAAGTTATTTGCAAAAGCAAGAATGGAGTAATCAGAGAGTATAAAAAAAAAAT
AGGATCCAAATTCTAAAGAAATATGGTCAAATATAGTTGGCCACGTGAGTGTAAATCATATCGGTCATGT
GCTTTTACCAAATACAATCATAATAACTAAATCAATTGTTAGATTAATGATAAATTGGTGAATACAATTT
TGATAGGTCCAAAAATTAATAAACACTATTTGTAGATGTTAGGAATTAAGCAAGTTCAAAACAAAGGGAT
AGTTCAATAGATGAGAGAATCTCATGACTTAAGTACACTACCACATTGAGCACTTACTCGATATGAGACT
CGTAACATTGACACTGCCTTTAAGGGTCGACATCCACATTGCCACTTACTGGATGTAGGACTTAAGCACA
TTAGAAGAGTCAAACTCAAAAGACTAATCCAATAAGTGAGAGAGTCTCACGCCTTAAATACTGCATTGAA
CACTTTAAAATATTTGATGTAAAACTCCTAATAGTAAACCTATTTGAATTATTCACGTCAATGGTGTGCA
TCTCTAAATAAAATATTAAATAGTGGAGAGTGAATTTCTTGTTTACTATTATATAAAGAGCAGAGGAAAT
GAAAGCAACAAAACAAACACAAATAGTATACACATTATTAAGCAAAGCAAATCATGAATCCTACTTTATC
CATTACCCTTTCCTTCCTCCTATTTGCATTTATCTCCAATCTTTCACTAAACAATGCAGTTCAGCAAGTA
TTTGACACAAATGGTAACCCTGTTGTCTCTGGTAAACAATACTTCATTTTTCCAGCAACTGATAACCCTA
AAAAGGGAGGACTAACCTTAAACAATGTCGGCGACGATGATTCAAAATGTCCAGTAACTGTCCTACAAAA
CAATGCTATAACAGGTTTACCAGTTAAATTCACCATCCCACAAACCACCACCGATAATATCGTAACAGGT
ACCGATCTTGACATCGAGTTCACCGAGAAACCTGATTGTGCTGAATCATCAAAATGGTTACTAGTTACTG
ATGATAATACTCAACAAAGCTATGTTGGTATTGGTGGTCCTTCAAATTATCCTGGTGTAGAAATAATCAG
TGGAAAATTTTTAGTTGTGAAACATGGAACTGGTGGAAGTTATAAGATTGGATTTTGTTTGGATAGTACT
GGTGATTGTGGTTATCTTGGATTACAAGTGTTTAATTCCGGAGAAGGTGGTTCGCGTTTGATTTTAACTA
TAACTGATGCTTATTCTGTTGTGTTTGTTGATTCTGCTTCTGTTAAATCTGAAAATATTGTGTCTGCTGA
AGGTAATCCAAATATTGCATTGCCTATTTAAGTGCTTGTTTTTATGAATAAGTGGAAAAGTGATTGTAAC
ACTTTCTTAATTATAATAATTGATGTTTGTAGGAATGGATCTTGTAATCCGTTTCTTTATCTGATTAATA
GTTTGAGGTCAAAGTTTAATGAAAATTCTTCCCTGACATTTTTATTATTAA

Figure 3.40: Full length *Tr-KPI5* gene, as indicated. The yellow highlighted region indicates the start and stop codons, the red highlighted region indicates the position of transcriptional start site (TSS) which is identified using FGENESH (http://linux1.softberry.com), the blue highlighted region indicates the presence of the GSP1 primer set and the grey highlighted region is indicating the presence of GSP2 primers sets. The green highlighted region is the position of the primers sets used to amplify the whole sequence from genomic DNA.

3.5.1 Cis – Binding elements in the promoter sequences

The putative binding domains corresponding to the different transcription factors in each promoter region were identified using different bioinformatics tools. These can only suggest the possible TFBS (transcription factor binding sites) for their corresponding proteins, and not the functionality of a site. It also needs to be noted that possible TFBS in the promoter region for a particular gene can be functional depending on tissue type, developmental stages and under different biotic and abiotic stresses. Therefore, different bioinformatics programmes can only indicate the binding potentials of the regulatory elements.

The putative transcription factor binding sites (TFBS) in the promoter sequences of the *Tr-KPI* genes were analyzed with MatInspector using matrix family concept databases (Cartharius *et al.,* 2005). TFBS with the core similarity of 1.0 (indicates highest conserved bases of the matrix exactly match the sequence) and matrix similarity (which considers all bases over the whole matrix length) of 0.9 were considered as potential candidates (for different transcription factor binding frequency; Table 6). The promoters of *Tr-KPI1, Tr-KPI4* and *Tr-KPI5* comprised an almost conserved predicted transcriptional start site (TSS) which starts at 64bp, 66bp and 72bp upstream of the ATG, whereas the TSS site for *Tr-KPI2* starts at 169bp upstream of ATG (Figure 3.40).

A number of putative hormonal, developmental and stress responsive transcription factor binding sites have been identified in the promoter region for all four genes (Table 6). Two hormonal regulatory binding sites have been observed in the promoter region namely for the transcription factor auxin response factor (AREF) (Guilfoyle, *et al*., 1998) and ethylene insensitive 3 like factors (EINL) (Solano *et al*., 1998). The TFBS for growth and development related transcription factors such as the ASRC (AS1/AS2 repressor complex) (Guo *et al*., 2008), the plant specific floral meristem identity gene LEAFY (LFY) (Parcy *et al*., 1998), the floral meristem and organ identity related binding factor MADS box protein (Riechmann and Meyerowitz, 1997), and the sugar responsive transcription factor gene (SUCB) are also observed in the promoter regions of the *Tr-KPI* genes.

TFBS specific for legume plant storage proteins such as the legumin BOX family (LEGB) (Baumlein *etal*., 1992) , a nodulin consensus sequence binding factor NCS2 (Jorgensen *et al*., 1991), and two different soybean embryo factor SEF3 and SEF4, responsible for embryo specific expression were also observed (Kovalchuk *et al.*, 2012).

The TFBS for biotic and abiotic stress responsive transcription factors such as MYBS; (Jin and Martin, 1999), MIIG, MYCL ; (Boter *et al*., 2004) NAC (NACF, CNAC, SWNS, WNAC); (Nakashima *et al.*, 2012), ABA responsive element (ABRE) (Singh *et al*., 2002), dehydration stress responsive element (DREB); (Liu *et al.,*1998), and stress activated VIP1 (VRES); (Pitzschke *et al*., 2009) are observed in the *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* promoter regions.

The TFBS for environmental cue responsive transcription factor, such as GAP-Box (light response element); (Park, *et al*., 1996) are observed in *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* and LREM (Light responsive element motif not regulated by different light qualities); (Jeong and Shih, 2003) are observed in the *Tr-KPI2, Tr-KPI4* and *Tr-KPI5* promoter regions.

Figure 3.41: Putative transcription factor binding sites upstream of the ATG of the promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* genes, as indicated, as identified by MatInspector.

Horizontal lines are indicating the promoter sequence of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5*. Red, green, yellow and blue colour is representing A, T, G and C nucleotide respectively. ATG is representing the start codon

⇨ : indicating forward direction of the transcription factor binding site and

⇦ : indicating reverse direction of the transcription factor binding site

Different coloured arrow is indicating position of different transcription factor binding sites where TSS is transcription start site; ABRE is ABA response element; AREF is auxin response element; ASRC is AS1/AS2 repressor complex; CNAC is calcium regulated NAC factor; DREB is dehydration responsive element binding protein; EINL is ethylene insensitive 3 Like factors; GAPB is GAP-Box (light response element); GBOX is plant G-Box/C-Box bZIP proteins; LEGB is legumin BOX family; LFY is LFY binding factor; LREM is light responsive element motif not regulated by different light qualities; MADS is MADS box proteins; MIIG is MYB IIG-type binding factors; MYB is MYB like proteins; MYBS is MYB proteins with single DNA binding repeat; MYCL is Myc-like basic helix-loop-helix binding factors; NACF is CUC2 (cup-shaped cotyledon 2) transcription factors; NCS2 is nodulin consensus sequence 2; SEF3 is Soybean embryo factor 3; SEF4 is soybean embryo factor 4; SUCB is sucrose Box; SWNS is secondary cell wall NACs; VRES is VIP1 responsive element and WNAC is wheat NAC domain transcription factor.

Figure 3.41: Putative transcription factor binding sites upstream of the ATG of the promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* genes, as indicated, as identified by MatInspector. For legend see page 118.

Continued

Figure 3.41: Putative transcription factor binding sites upstream of the ATG of the promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* genes, as indicated, as identified by MatInspector. For legend see page 118.

Continued

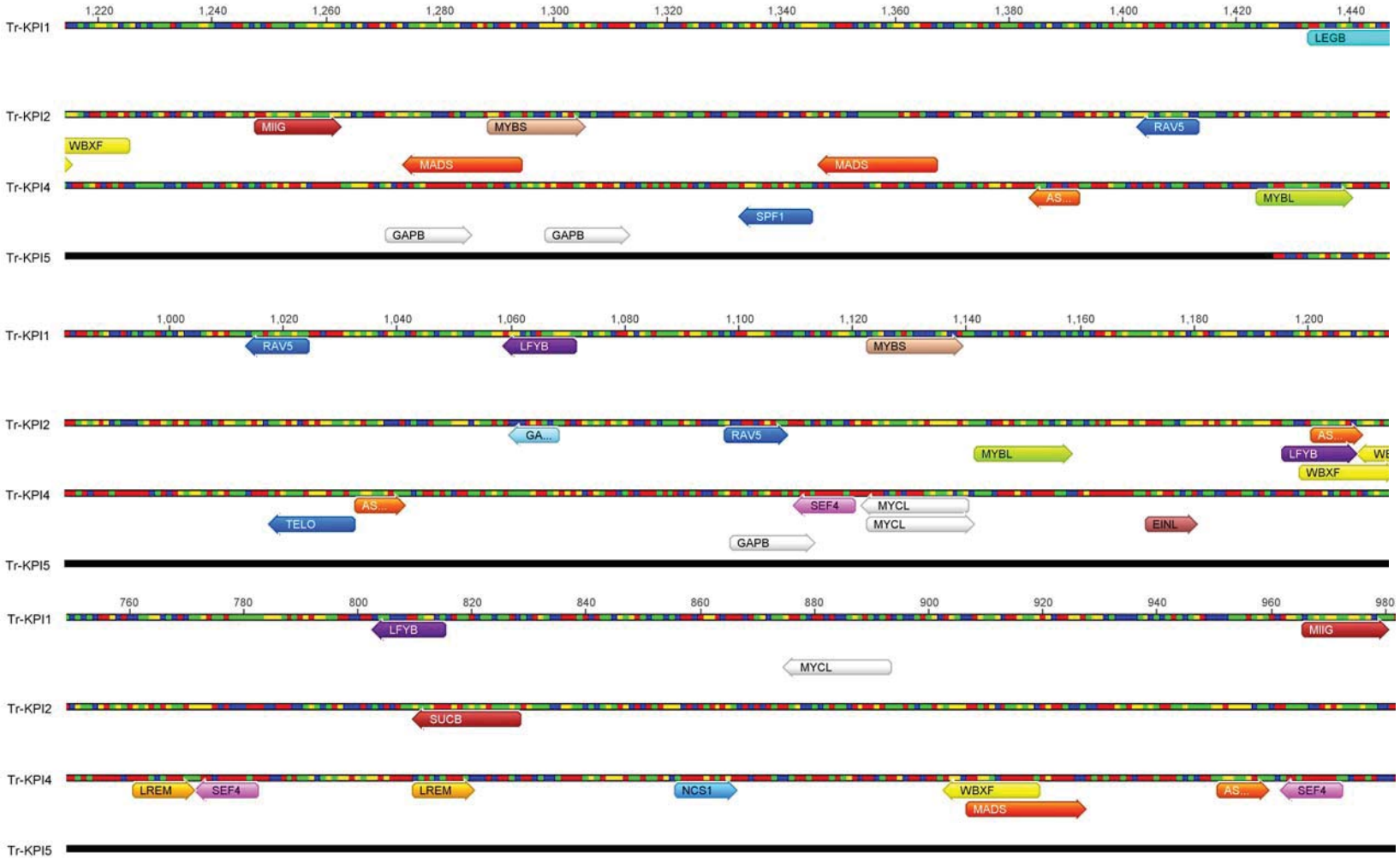


Figure 3.41: Putative transcription factor binding sites upstream of the ATG of the promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* genes, as indicated, as identified by MatInspector. For legend see page 118.

Continued

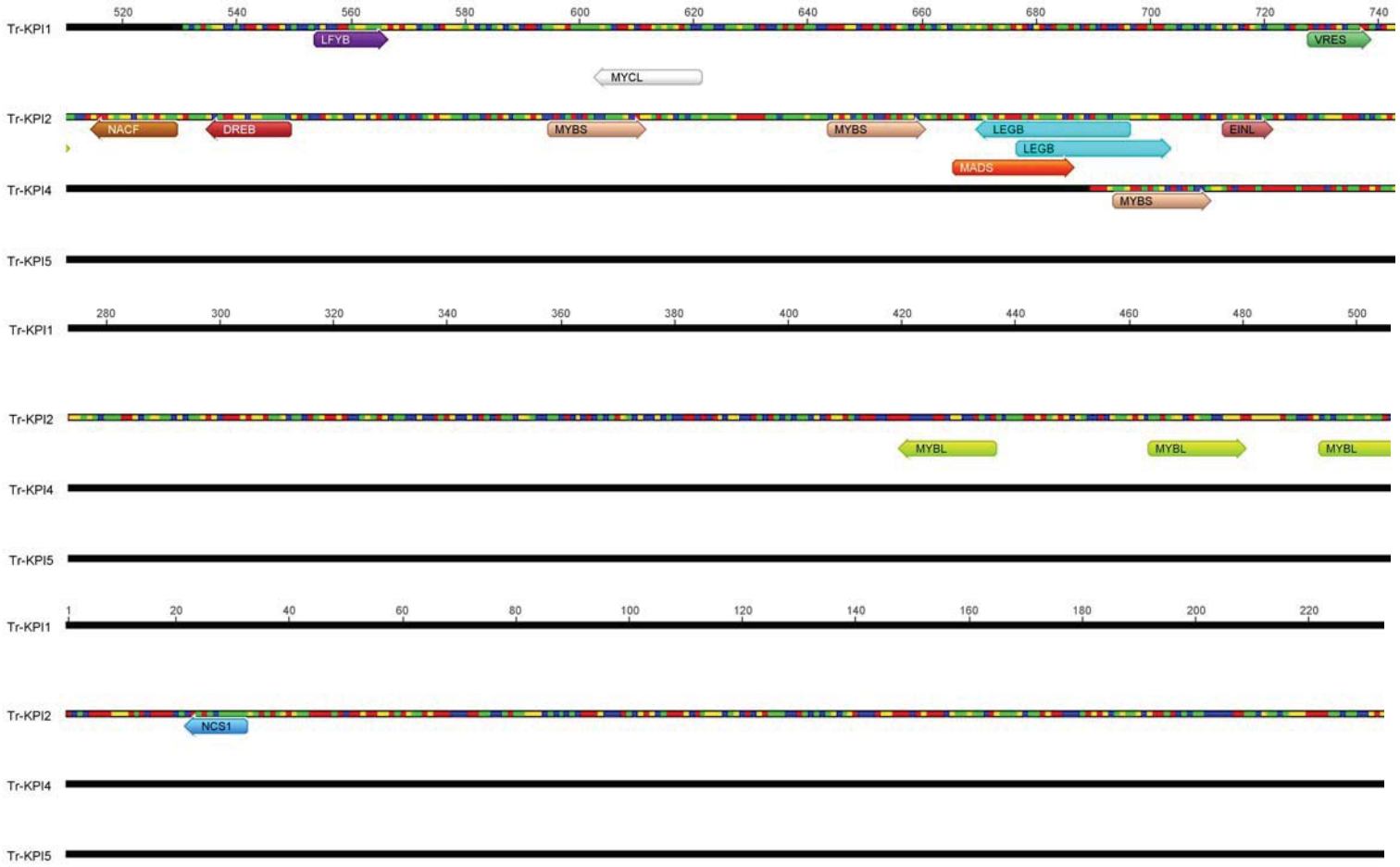


Figure 3.41: Putative transcription factor binding sites upstream of the ATG of the promoter region of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* genes, as indicated, as identified by MatInspector. For legend see page 118.

Table 3.6: Frequency of occurrences of Cis-binding elements in the promoter region of *Tr-KPI* genes, as indicated as predicted by MatInspector

Transcription Factor Family[1]	Detailed Family Information	Tr-KPI1	Tr-KPI2	Tr-KPI4	Tr-KPI5
ABRE	ABA Response Element	1	0	0	1
AREF	Auxin Response Element	1	1	0	0
ASRC	AS1/AS2 repressor complex	0	1	3	0
CNAC	Calcium regulated NAC factor	0	0	0	1
DREB	Dehydration responsive element binding protein	0	1	0	0
EINL	Ethylene insensitive 3 Like factors	1	1	4	2
GAPB	GAP-Box (light response element)	0	0	7	2
GBOX	PlantG-Box/C-Box bZIP proteins	0	0	0	4
LEGB	Legumin BOX family	0	2	0	0
LFY	LFY binding factor	3	1	1	0
LREM	Light responsive element motif not regulated by different light qualities	1	1	3	2
MADS	MADS box proteins	1	4	2	3
MIIG	MYB IIG-type binding factors	1	3	0	0
MYBL	MYB like proteins	3	6	2	2
MYBS	MYB proteins with single DNA binding repeat	3	3	1	1
MYCL	Myc-like basic helix-loop-helix binding factors	6	0	3	6
NACF	CUC2 (cup-shaped cotyledon 2) transcription factors	0	1	0	1
NCS2	Nodulin consensus sequence 2	0	3	1	0
SEF3	Soybean embryo factor 3	0	1	0	0
SEF4	Soybean embryo factor 4	2	1	4	0
SUCB	Sucrose Box	0	1	0	0
SWNS	Secondary cell wall NACs	2	0	0	0
VRES	VIP1 Responsive element	2	1	0	0
WNAC	Wheat NAC domain transcription factor	3	5	2	2

[1]Green: Growth and development responsive transcription factor binding sites
Orange: Legume specific transcription factor binding sites
Red: Stress Responsive transcription factor binding sites
Blue: Environmental cues responsive transcription factor binding sites

3.6 Knock-down of Tr-KPI genes expressing in white clover using RNAi

The knockdown lines for each of four *Tr-KPI* genes were developed in white clover using the RNAi mechanism (section 2.4). The lines were established in soil and well developed plants were selected to check the mRNA abundance level for each of the *KPI* lines. For the background control, three lines were selected that do not have the insert but had the same history as the RNAi lines and were named as the NC (negative control). The same wild type (cv. QRT-PCR) that was used in different expression study experiments was also used to categorize the expression level. Depending on the transcript abundance values, the knockdown lines were classified into three categories: high (same or high expression level compared with the NCs and wild type), low (low expression level compared with the NCs and the wild type) and medium (falling in between the high and low category). Leaf tissues were used for the experiment for *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* lines while root tissues were used for *Tr-KPI4* lines.

For *Tr-KPI1*, the high category involved those lines that have the transcript abundance value higher than 0.05, medium categories involved transcript abundance of 0.05 to 0.01 and low categories include the values of less than 0.01 (Figure 3.42). Seven lines were selected for high, six were selected for medium and only four lines were selected as the low expressing lines for *Tr-KPI1.*

For *Tr-KPI2,* the high category involved lines that had values more than 0.05, and four lines were identified in this category. Eight lines were identified for the medium category (values 0.05 to 0.01) and five lines were identified in the low category (less than 0.01). Interestingly, the expression of *Tr-KPI2* in the NCs and in the wild type was found to be lower when compared with the high expressing lines (Figure 3.42).

For *Tr-KPI4* RNAi lines, the expression level of the gene were studied in the root tissue as previous data suggested that this gene is more prominent in the root. Three lines were identified as high expressing (values >0.005), two lines as medium expressing (values 0.005-0.001) and four line were identified as low expressing (<0.0009).

For *Tr-KPI5*, three lines were identified as high (<0.1), two as medium (>0.1 to <0.05) and six were identified as low (>0.04) expressing lines (Figure 3.42).

3.6.1 Tr-KPI genes expression in the RNAi lines

The low expressing RNAi lines for *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* were selected to examine the expression of the other member of the gene family in each knockdown (KD) background. Therefore, the expression of *Tr-KPI1* was examined in the low category *Tr-KPI2, Tr-KPI4* and *Tr-KPI5* RNAi lines; *Tr-KPI2* was examined in the low category *Tr-KPI1, Tr-KPI4* and *Tr-KPI5* RNAi lines and *Tr-KPI5* was examined in low category *Tr-KPI1, Tr-KPI2* and *Tr-KPI4* RNAi lines (Figure 3.43, 3.44).

Tr-KPI1 was found to be highly expressed in three *Tr-KPI2* low category KD lines (L206, L207 and L220), compared to wild type and three negative control, whereas wild type level of expression was observed in line L204 and L215 (Figure 3.43). In the *Tr-KPI5* low category KD lines, *Tr-KPI1* showed the same level of expression when compared with the wild type and negative control. For *Tr-KPI2*, the expression level was found to be much lower when compared with the wild type and negative control in the *Tr-KPI1* low category KD lines, whereas, in the *Tr-KPI5* lines, *Tr-KPI2* maintained a low level of expression except for line L508 where a wild type level of expression was observed. The expression of *Tr-KPI5* in *Tr-KPI1* lines was found to be similar when compared with wild type and lower when compared with the negative controls. However, *Tr-KPI5* was found to be highly expressed in *Tr-KPI2* lines L206 and L207 when compared with the wild type and the negative controls (Figure 3.43).

The expression of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* was also studied in the root tissue of *Tr-KPI4* RNAi lines. A similar expression level was observed for *Tr-KPI1* and *Tr-KPI2* and a lower level of expression was observed for *Tr-KPI5* gene when compared with the wild type (Figure 3.44) and the negative control (data not presented).

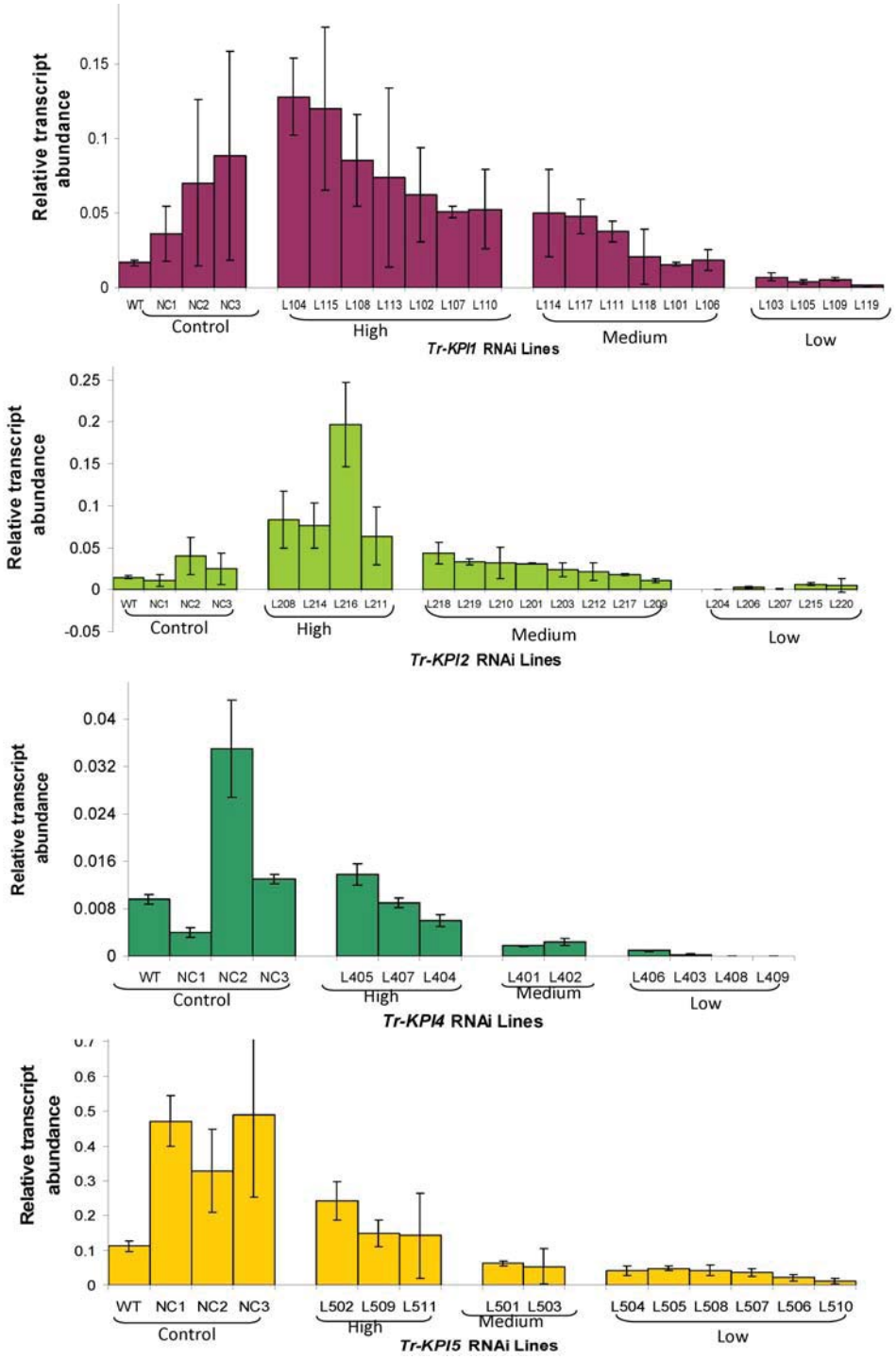


Figure 3.42: Transcript abundance of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* in their respective RNAi lines, as indicated. qPCR was performed using three technical replicates with tissues collected from different RNAi lines using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value of three technical repeats and the error bars representing mean ± SEM.

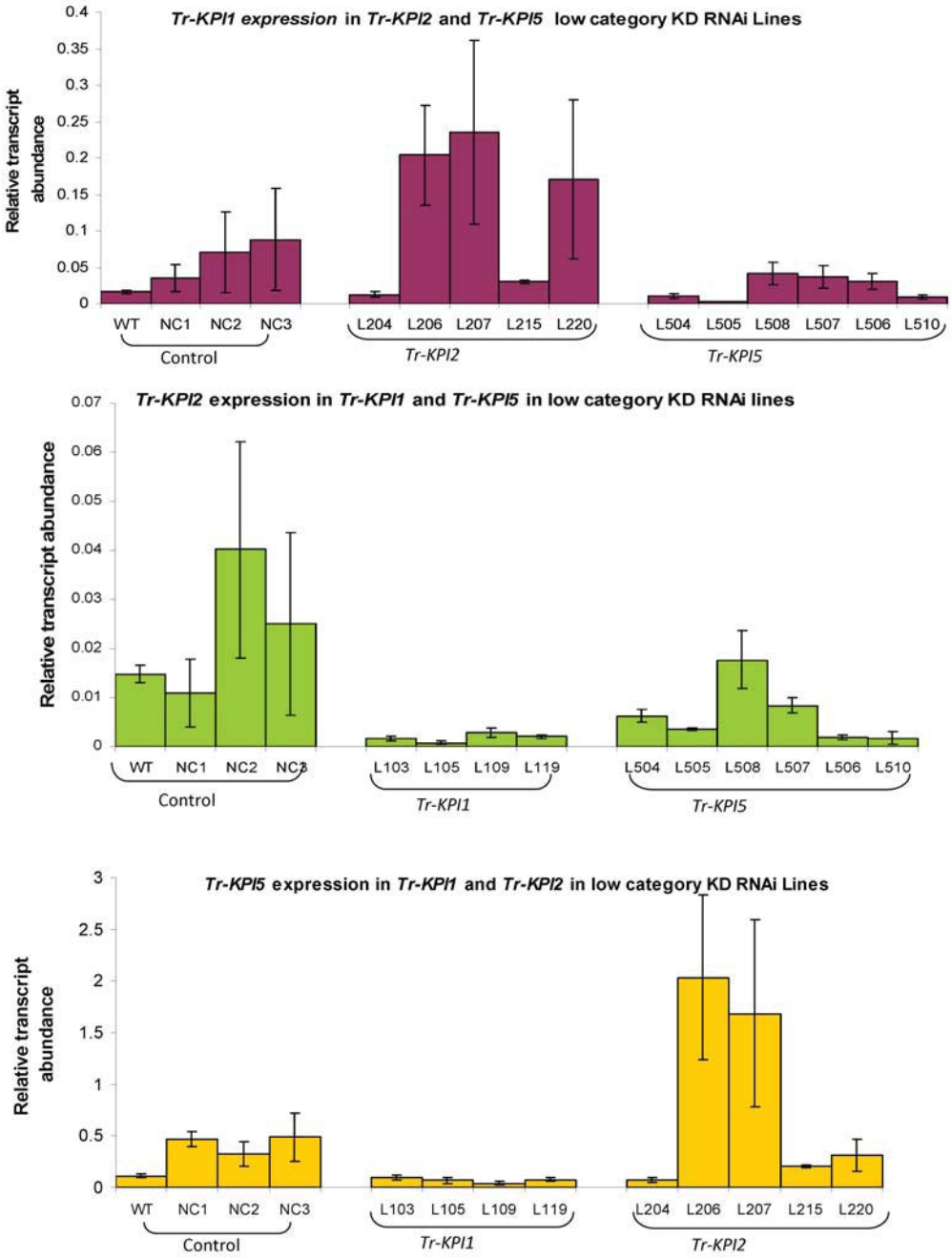


Figure 3.43: Expression of the *Tr-KPI* genes in the leaf tissues of low category RNAi KD lines, as indicated. qPCR was performed using three technical replicates with tissues collected from different RNAi lines using *β-actin* and *GAPDH* as internal control. Each data point represents the mean value of three technical repeats and the error bars representing mean ± SEM.

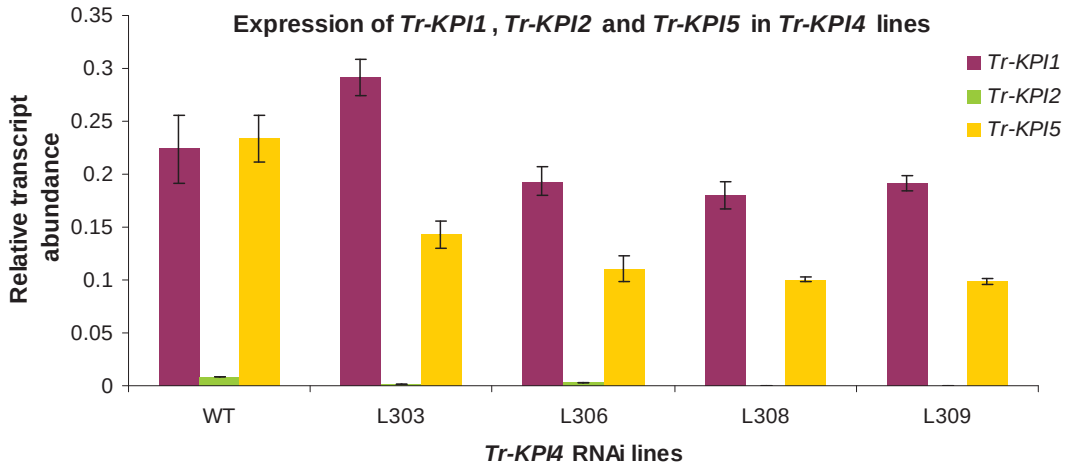


Figure 3.44: Expression of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* in the root tissues of *Tr-KPI4* RNAi lines, as indicated. qPCR was performed using three technical replicates *with tissues collected from different RNAi lines using β-actin* and *GAPDH* as internal control. Each data point represents the mean value of three technical repeats and the error bars representing mean ± SEM.

Chapter 4

Discussion

The search for the occurrence and functions of proteinase inhibitors including *KPIs* in crop plant species has been mainly focused on the development of transgenic plants resistant to insect herbivory. However, the endogenous function of these proteins other than for a protein storage purpose has always been overlooked. *KPI* gene families have thoroughly been studied in potato (Heibges *et al.,* 2003 a and b), poplar (Major and Constabel, 2008) and soybean (Jofuku and Goldberg, 1989), where the main focus was to determine the involvement of these proteins against insect herbivory. Though the role of *KPIs* as endogenous insecticides cannot be ignored, it is important to determine the additional functions of these *KPIs* as they may also be involved in controlling cellular activity under both normal and stress conditions. Here this thesis contributes to the current knowledge of the occurrence and regulation of *Kunitz proteinase inhibitors* in white clover. The examination of different plant developmental stages, biotic stress using two different herbivores and two different kinds of abiotic stress clearly indicates that members of the *KPI* gene family are regulated at the transcriptional level and suggests their possible involvement in plant growth and developmental and in response to biotic and abiotic stresses.

4.1 Occurrence of KPIs in white clover

Kunitz Proteinase Inhibitors are widespread in the plant kingdom and have been described from many plant species. More than one homologue of *KPIs* has been identified in different plant species: for example, two in chickpea (Hernandez-Nistal *et al.,* 2009), 11 in soybean (Jofuku and Goldberg, 1989), 21 in potato (Heibges *et al.,* 2003 a and b) and 31 in poplar (Philippe *et al.,* 2009). Assuming that *Kunitz Proteinase Inhibitors* are a multigene family in white clover, the first goal of the study was to identify the members of this gene family.

In white clover, one gene was identified using the AgReseach EST database which was later on amplified by gene-specific primers and named as *Tr-KPI1*. Degenerate primer sets were used to find further *Tr-KPI* genes from white clover and eight partial sequences were identified. Later, using genome walker and 3' RACE technology, three further full length genes were isolated and named as *Tr-KPI2, Tr-KPI4* and *Tr-KPI5*. Though repetitive attempts were made to clone the upstream and downstream regions for the four other partial sequences identified (*Tr-KPI3*, *Tr-KPI6*, *Tr-KPI7*, *and Tr-KPI8*; Figure 3.1B), this proved difficult as these genes are intron-less and they share a high degree of identity (for example, Tr-KPI3 shares 97% identity with Tr-KPI7; Tr-KPI6 shares 94% identity with Tr-KPI8 in the overlapping regions). However, it can not be stated with certainty that these eight partial and/or full length sequences comprise the whole

KPI family in white clover. As white clover is an allotetraploid (2*n* = 4*x* = 32) out-crossing plant species (Atwood and Hill, 1940), it might be speculated that these highly identical genes arise from the two different parents (for example *Tr-KPI1* and *Tr-KPI6* or *Tr-KPI8*). As mentioned previously, among the legumes, soybean has 11 members and *Medicago* has 12 members which are both diploid plant species. Therefore, it is likely that white clover might have more than 8 *Tr-KPI* genes in the genome. However, in this study, the four full length identified genes were the focus of further analysis.

BLAST searching with translated sequences of the identified Tr-KPIs using the NCBI and JCVI (against *M. truncatula* GBrowse) databases gave hits to only Kunitz-type Trypsin Inhibitor proteins. All four identified Tr-KPIs gave specific hits to the STI (Soybean Kunitz Trypsin Inhibitor) super-family conserved domain, strongly indicating that the white clover sequences belong to the KPI group. No sequence from any other gene family was identified by BLAST searching against GenBank and the JCVI database, again confirming that they are the orthologs of KPI. In order to confirm the identity of these four isolated genes, at least nine published homologues from different leguminous species were selected from the NCBI database and included in the present study for sequence comparison. The genes showed high sequence similarity with KPIs from *P. sativum, C. arietinum* and *M. truncatula.*

The identities of the four full length Tr-KPI proteins varied from 40 to 64% with each other. This high variation raises the possibility of diversified roles for the members of the Tr-KPI family in white clover, as it has been reported that for more distantly related proteins in a superfamily functional variation and substrate specificity is diverse (Todd *et al.,* 2001). However, existence of high dissimilarity among the members of KPIs is not uncommon in other plant species. For example, research with the *P. trichocarpa* Kunitz gene family suggests that 31 KPIs share about 96 to 38% identity to form six different clades and also suggests that tandem duplications might have played an important role in the expansion of this gene family (Philippe *et al.*, 2009). Poplars are long lived tree species, and therefore a large and diverse KPI family (31 members) might be crucial for its survival against multiple stresses (Philippe *et al.*, 2009). However, soybean and *Medicago* are annual plants and so the presence of more than 10 members could indicate the importance of different family members in different aspect during the shorter life cycle of these plants. When compared to annual legume species (soybean and *Medicago*), white clover is a stoloniferous perennial legume and therefore, the function and occurrence of some of the Tr-KPIs might be quite dissimilar. Thus it can be speculated that the Tr-KPIs are active against different proteinases under different conditions

that have evolved in white clover or other perennial plants as a result of more prolonged exposure to different biotic or abiotic stresses**.**

The identified Tr-KPIs contain four conserved cysteine residues which is one of the identifying characters of the KPI family. According to Oliva *et al*., (2011), plant Kunitz proteinase inhibitors especially from the Leguminosae can be classified based on their cysteine residues into three groups: Kunitz PIs with four or more cystine residues (forming two disulfide bonds), two or three cysteine residues (forming one disulfide bond) and one or no cystine residues. Tr-KPI1 and Tr-KPI2 have 5 cysteine residues and Tr-KPI4 and Tr-KPI5 have four cysteine residues. Therefore, according to this classification all of these four proteins fall into the same group.

The reactive site loop of Tr-KPIs is devoid of Lys or Arg which are present in the reactive loop of STI. However, Ile and Ser, which are present in other trypsin inhibitor reactive site residues (Major and Constabel, 2008) is present in the reactive site of Tr-KPI1 and Tr-KPI2. The presence of Glu-Ile in Tr-KPI4 and Thr-Thr in Tr-KPI5 reactive site loop does not match with any of the well identified KPI proteins, and this could indicate functional divergence of these four genes. Therefore, the functional characteristics of Tr-KPIs can not be envisaged on the basis of their sequence similarity. The sub-cellular localization prediction programme (TargetP 1.1 Server; Emanuelson *et al.,* 2000) , predicted that the Tr-KPIs are located in the secretary pathway, in common with KPIs from other plant species. Therefore, vacuole localization is a possiblity, which is the compartment that confines a variety of chemical compounds as well as proteins, including both enzymes and enzyme inhibitors that are involved in storage and active plant defense (Hollander-Czytko *et al.,* 1985).

Therefore, to determine whether these proteins are active or not, additional biochemical analysis with a wide range of commercially available proteinases is necessary. The high inequity at the active site found in this study also raises the possibility that some of the identified Tr-KPIs are not targeting the proteinases at all *in planta* as it has been reported by McCoy and Kortt (1997) that some proteins of the KPI family do not act as inhibitors. Glaczinski *et al.,* (2002) showed in tubers from cv. Provita that a Kunitz type proteinase inhibitor which had several amino acid substitutions when compared with the KPIs and no sequence similarity with a tuber invertase, did not inhibit any of the proteinases but show inhibitory activity against invertase. There is also a possibility that some of these genes are post-transcriptionally regulated or not translated. Selective translation of mRNA is an important mode of gene regulation of plants under both normal and stressed conditions (Bailey-Serres, 1999). For

example, selective translation of mRNA was observed in maize roots under oxygen deprivation where many normal cellular mRNAs are synthesized but showed a minimal level of translation (Bailey-Serres, 1999).

4.2 Tr-KPI genes differ in tissue specific expression and are developmentally regulated at the transcriptional level

The association of different members of serine proteinase inhibitor families with different plant developmental stages suggests their involvement in endogenous functions in different plants. For example, *PIN2* of the serine proteinase inhibitor family has been found to be constitutively expressed in potato and tomato flowers and potato tubers (Lorberth *et al.*, 1992). Later, the involvement of *PIN2* was found to be associated with seed development process using RNAi lines in which a lack of *PIN2* causes the formation of an abnormal endothelium (Chye, *et al.,* 2006), and so confirming the inferences from the transcriptional data. The transcriptional regulation of different members of poplar *KTI* gene family has been observed in apical tissue, young leaves and flowers (Major and Constabel, 2008). The poplar study confirmed that some of the *KPIs* are constitutively expressed in different plant parts but does not provide any information about any changes in expression during the course of development. Therefore, a more detailed study regarding the involvement of the *KPI* gene family during plant growth and development is required which will further aid our current understanding about their involvement during plant growth and development.

In this study, the expression of the four KPI genes was examined in different tissues, and at different developmental stages. Transcripts of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* were observed in all parts of the plant whereas *Tr-KPI4* is found to be more tissue-specific in terms of being expressed only in mature seed, stem and root (Figure 3.7). Therefore this spatial expression pattern of *Tr-KPIs* in white clover indicates that the members of the gene family are expressed throughout plant growth and development but vary in terms of transcript abundance which can point towards their diversified functions in the plant life cycle. To further characterize these four genes in white clover, different plant growth developmental stages were investigated.

For leaf development, the expression of *Tr-KPI2* and *Tr-KPI5* was found to be negatively related with age as the expression was more prominent in the youngest tissue, i.e. the apex followed by the first emerged and then the first fully expanded leaves. In contrast, expression of *Tr-KPI1*

was broadly similar over the same developmental stages (Figure 3.8). A similar level of expression was observed for*Tr-KPI1* and *Tr-KPI2* transcript abundance in the three different root developmental zones studied (Figure 3.9), while the transcript abundance of *Tr-KPI5* in the different root developmental zones followed the same pattern as in the leaf developmental stage, i.e. a lower transcript abundance in the mature tissue. *TrKPI4* showed a higher transcript abundance especially in the visible lateral root region which might be an indication of the function of this specific member in terms of regulating lateral root development as this region contains the lateral root primordia. For different nodule developmental stages, a higher transcript abundance of *Tr-KPI1* and *Tr-KPI5* was observed which can indicate a defense response against rhizobium infection (Figure 3.10). Overall, both the leaf and root developmental study suggests that *Tr-KPI5* and *Tr-KPI2* might be more responsive to developmental cues among the four genes studied.

Usually the apical parts of the plants are preferentially protected when compared with mature and senescing parts in the sense that this tissue provides the founder cells for the development of new organs and also supports indeterminate growth. Thus damage to the apical parts compromises continued development. Microarray analysis of apical shoots of Stika spruce (*Picea sitchensis*) showed preferential expression of different classes of proteinases and defense-related proteins including proteinase inhibitors which indicates a commitment to constitutive defense in the apical shoot (Friedmann *et al.,* 2007). Therefore, the higher transcript abundance of *Tr-KPIs* in the apical tissues highlights the probable role of inhibitors in protecting the growing parts from the degradation of key proteins by proteolytic enzymes. As well, the apical parts of the plant are a metabolically active developing tissue, and so a high transcript abundance of *Tr-KPIs* in the apical parts could also indicate high proteinase activity and the regulation of a balance between protein synthesis and degradation as a result of cell division. The comparatively lower transcript abundance of *Tr-KPIs* in the mature tissue (root and leaf) and in senescing tissue (leaf) indicates their importance during controlled senescence processes under normal conditions. Senescence which is the final phase of life, involves protein and macromolecule cleavage and the reallocation of nutrients in the reproductive and developing plant parts (Gan and Amsino, 1997, Mosolov and Valueva, 2011). Therefore, the down-regulation of *Tr-KPIs* in the senescing tissues under normal condition may indicate that the profile of proteolytic enzymes might have altered and other proteinases might now be more active compared with the target serine proteinases for Tr-KPIs.

It has been mentioned previously that the Kunitz proteinase inhibitors in white clover display the conserved features of the soybean trypsin inhibitor (Kunitz) family since they have two pairs of cysteines, which is considered to be a special characteristic of seed Kunitz PI (Spencer and Hodge, 1991). The observation that transcripts of *Tr-KPIs* have been detected in dry seeds (Figure 3.11) may indicate their existence as stored mRNAs and also signifies that *Tr-KPI* mRNAs accumulate during seed development in common with other legumes such as soybean and pea (Jofuku and Goldberg, 1989; Whelham *et al.,* 1998). These mRNA transcripts in the dry seed are thought to be required during late embryogenesis and during the early stages of germination (Hernandez-Nistal *et al.*, 2009). The developmental phases that occur during late embryogenesis and subsequent germination are characterized by spatial and temporal patterns of gene expression. However, there are other metabolic genes involved in seed germination whose induction starts after imbibition (Nakabayashi *et al.,* 2005), including those involved in the cell cycle, DNA processing, transcription and protein synthesis. The transcription of *Tr-KPI1* and *Tr-KPI5* was up-regulated at 24 h in the embryonic axis and increased gradually up to 72 h over the germination time course. *Tr-KPI2* transcripts were detected after 48 h and *Tr-KPI4* was expressed basally during the time course study. Here, the increasing abundance of *Tr-KPI1* and *Tr-KPI5* after 24 h of imbibition compared to *Tr-KPI2* and *Tr-KPI4* can be considered as an early induction. However, the function of these genes may be beyond one of storage as the transcript level increased from 24 h onward rather than decreasing. It was found in chickpea using both northern and western blot analysis that out of the two identified Kunitz trypsin inhibitor, CaTPI-1 served as a storage protein whereas CaTPI-2 was mainly active following germination especially during elongation of the embryonic axes (Hernández-Nistal *et al.,* 2009). Western blot analysis with mature and dry seed tissue from white clover using specific antibodies will further indicate whether the gene products serve as storage proteins or not and further research is required with other Tr-KPIs from white clover to determine if other members serve as storage proteins. The transcriptional study suggests the involvement of *Tr-KPIs* in some unknown process that leads the seed to germinate and in the ensuing growth of the embryonic axis. It should be noted that seeds of white clover are small when compared with other legumes such as soybean or chickpea, and also the storage organ, (i.e. the cotyledon), is very small compared to the embryonic axis. Therefore, from the transcriptional level study, it can be speculated that some of the members of these gene family other than the four identified ones, may play quite specific roles in reserve protein storage at seed maturity and *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* are more involved during the post germination process.

4.3 Tr-KPI genes and biotic stress

The physiology of tissues undergoes a variety of changes in response to artificial wounding or by the attack of insect pests. Localized damage can trigger the defense system throughout the plant including the tissue that is damaged directly (known as the local response) and in the undamaged part of the plant (known as the systemic response). The transcription of specific genes is activated either at the wounded site or in a distant tissue within minutes or several hours after wounding depending on the role they play, for example in repairing damaged tissue, the activation of defense signaling pathways, metabolic adjustment and producing toxic inhibitory substances aganist the predator (Leon *et al.*, 2001). Even though the vital function of jasmonic acid (JA) in plant responses to wounding is well established (Ellis *et al.*, 2002; Halitschke and Baldwin 2004; Abe *et al*., 2009; Landgraf *et al.*, 2012), the transient rise and involvement of other compounds such as systemin (which is a oligopeptide), oligosaccharides, and other phytohormones such as ABA (Pena-Cortes *et al.*, 1989) and ET (O'Donnell *et al.*, 1996) has also been identified to play regulatory role to wound signaling. In Solanaceous plants ,it has been documented that systemin, which is a 18 amino acid peptide, and is released at the time of wounding, plays the primary role as a long distant transmittable signal (Ryan, 2000). However Li *et al*., (2002) used two tomato mutants that are defective in the wound responsive pathway *spr-2* (responsive to JA/MEJA and insensitive to systemin) and *jai-1* (insensitive to both JA/MeJA and systemin) and showed that JA itself or its derivatives can also provide the signal for wounding in remote tissue.

In their finding, Green and Ryan (1972) proposed that mechanical wounding as well as wounding by the colorado beetle in the leaves of tomato and potato induces the expression of proteinase inhibitor genes throughout the plant body. Is has been proposed by many researchers, the transcription of proteinase inhibitor genes in different plant species (Solanaceae, Leguminosae, Poaceae, Brassicaceae, Salicaceae) is activated by biotic stresses. One of the aims of the present study is to characterize the transcriptional regulation of *Tr-KPI* genes in white clover in response to biotic stress.

To address this aim, the response of *Tr-KPIs* was examined using mechanical wounding to the first fully expanded leaves (FFE) (Figure 3.12). The results in this experimentation suggest that unlike many other *PI* genes from different plant species, *Tr-KPI* genes are also responsive to wounding. By 3 h of mechanical wounding, *Tr-KPI1* showed a 45-fold increase in expression in the wounded site, which then decreased by 12 h when compared with the control (Figure

3.13). The expression of *Tr-KPI2* and *Tr-KPI5* did not decrease by 12 h as was observed for *Tr-KPI1* but maintained a higher level of expression in the wounded tissue up to 24 h. These differences in expression of *Tr-KPI1* with *Tr-KPI2* and *Tr-KPI5* at the wound site indicate that *Tr-KPI1* might be the potential candidate as a wound responsive gene in terms of high transcript abundance at an earlier stage (3 h). In the systemic tissue (root) subtending from the third node, the level of expression of *Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5* showed a significant increase from 6 h to 24 h (Figure 3.14). The roots arising from the (distal) 4th node showed a transient expression after 3 h of wounding for the four genes studied (Figure 3.15). Interestingly, wounding in the leaf tissue activated the expression of *Tr-KPI4* in the root tissue. This study suggests that the tissue in the closer proximity (3rd nodal root) to the wounded site received any wound-generated signal earlier and showed a higher and significant level of transcript upregulation compared to the more distant tissue (4th nodal root). This can be explained by the frequency of wounding in local tissue - the signaling system will alert the nearby tissue first and plants may then require more frequent wounding and/or long-lasting effect to alert the distal tissue (Landgraf *et al*., 2012). This observation may also support the connection of the wound signaling network involving MeJA and other hormonal crosstalk (Ryan, 2000; Li *et al*., 2002), but this would need to be examined more directly, perhaps by the application of these putative signals. Overall the result of mechanical wounding did show that the members of *Tr-KPI* gene family present in white clover are responsive to mechanical wounding of the transcriptional level and therefore can be members of the defense responsive genes array in this plant.

Expressions of plant proteinase inhibitors are triggered in response to insect attack and in the long run inhibit the digestive enzymes required for many vital processes in the plant pest and therefore negatively influence insect growth and development. The wounding study suggests that *Tr-KPIs* are activated by the biotic stress response pathway which can be a generalized wound response and therefore, does not fulfill the query whether they are involved in plant defense mechanism triggered by insect herbivory. Proteinase inhibitors belonging to the serine sub-family have been widely used to engineer genetically modified plants to be more resistant to insect pests (Luo *et al.,* 2009). The involvement of Kunitz proteinase inhibitors in plant defense against insect herbivores was first shown in over-expressed SKTI transgenic rice by Lee *et al*., (1999).

A generalist herbivore *S. litura* which has a wide host range (plants from Solanaceae, Leguminosae, Malvaceae) was used to characterize the induction of four *Tr-KPI* genes at the transcriptional level. Local (FFE), basipetal (root) and acropetal (apical tissue) tissues were used

to analyze local and systemic response of the *Tr-KPIs.* The expression results in response to herbivore attack suggest that *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* were induced in the local tissue and in systemic tissue (apex) but varied in their early response and transcript abundance (Figure 3.17 and 3.18). Among the four genes studied in the local tissue, *Tr-KPI1* is the most highly induced gene (ca. 1000-fold at 24 h) followed by *Tr-KPI2* and *Tr-KPI5.* Herbivore insect feeding involves mechanical wounding (arising from the mouth parts during chewing and maceration of the leaf surface by leg movement and deposition of oral secretions, e.g. saliva and regurgitant) and also the combined effect increases the level of expression of these genes over the 24 h time course study. In the apical tissue *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* transcript levels started to increase by 6 h and were significantly upregulated by 24 h. Interestingly, *Tr-KPI4* and *Tr-KPI5* were induced after one hour of the attack in the root tissue and by 24 h, the transcript level were suppressed when compared with the control, and *Tr-KPI1* and *Tr-KPI2* transcript did not vary at all when compared with the control. Therefore a prolonged systemic defense response was observed only for the apical parts of the plant. The results of *Tr-KPIs* expression in the local and apical tissue also suggests that a tiered response occurred due to continual feeding where *Tr-KPI1* was activated initially followed by *Tr-KPI2* and *Tr-KPI5.* Constitutive expression of these proteinase inhibitor genes may specify their endogenous functions but the marked increase in transcript abundance in response to insect herbivory indicates a further role in plant defense. The variation in the acropetal and basipetal systemic response is indicative of selection of an essential to preferentially protect under herbivory attack. Plants are adapted to focus on the protection of key tissues as induced defense is costly for plants to implement (Frost *et al*., 2008).

Mechanical wounding and insect herbivory in the above ground parts of the plant triggered a systemic response that was perceived in below ground parts. Therefore, the question was asked as to what will occur when plants are challenged against a root feeding herbivore. Plant parasitic root nematodes attack the root tissue, puncturing the cells using a stylet, and then withdrawing necessary nutrients and secreting protein and metabolites to parasitize the plant. A trypsin proteinase inhibitor in cowpea (*CpTI*) was the first *KPI* studied in anti-nematode defense, and its over-expression was shown to reduce the fecundity of the female cyst nematode *Globodera pallida* (Urwin *et al.,* 1998). Later, the induction of proteinase inhibitors triggered by nematode invasion has been studied widely in sugarbeet (Cai *et al.,* 2003), potato (Luo *et al.,* 2009) and tomato (Fujimoto *et al.,* 2011) and it has been reported that the expression pattern of *KPI* genes is regulated depending on the type of tissue and also the plant genotype (Turra *et al.,* 2009). Here in this study, induction of the *Tr-KPIs* was studied in

response to cyst nematode *Heterodera trifolii* and root-knot nematode *Meloidogyne trifoliophila* invasion.

First, to characterize whether the identified *Tr-KPI* genes in white clover are responsive to nematode attack, roots of the white clover cultivar Huia were inoculated with eggs of the cyst nematode (*Heterodera trifolii*)and the root-knot nematode (*Meloidogyne trifoliophila*). Inoculation by cyst nematode triggered the expression of *Tr-KPI1, Tr-KPI4* and *Tr-KPI*5 in the root tissue at day 4 and the expression level of *Tr-KPI1*, *Tr-KPI4* and *Tr-KPI5* remained up-regulated in the root up to day 8 (Figure 3.20). A systemic response to nematode feeding was also observed in the leaf tissue for these genes at day 8 (Figure 3.20). After hatching from egg, the second stage juvenile had already established a feeding site in the root tissue by 4 day (observed by microscopy). Their continuous feeding on root tissue is known to activate the wound signaling pathway which later on triggers the expression of defense related gene (Gatehouse, 2002) and this include the *Tr-KPIs.* Out of four *Tr-KPI* genes, three of them (*Tr-KPI1*, *Tr-KPI2* and *Tr-KPI5*) were induced earlier (at day 4) compared to *Tr-KPI2*, the induction of which was observed at day 8 in response to the feeding cyst nematode. On the other hand, invasion by the RKN did not show any significant up-regulation for *Tr-KPI* gene expression in the root tissue at day 4 and day 8 (Figure 3.20). Some down regulation in the transcript level was observed for *Tr-KPI2, Tr-KPI4* and *Tr-KPI5* in the root tissue and the transcript level remain unchanged for *Tr-KPI1*. These findings suggest that the *Kunitz proteinase inhibitors* from white clover might be involved in defense against cyst nematode invasion but not by root-knot nematode. The early induction and longer effect in the root indicates more localized function of *Tr-KPIs* under cyst nematode attack. On the other hand, RKN nematode might have the ability to overcome the barrier of plant defense imposed by *Tr-KPIs*, including the suppression of *Tr-KPIs* by day 8 (Figure 3.20).

To further elucidate the involvement of *Tr-KPI* genes under cyst nematode attack, resistant and susceptible lines collected from the AgResearch Grasslands breeding programme were examined. In the resistant line 17R, all four *Tr-KPI* genes were significantly expressed at day 4 and continued to day 8 after inoculating the root with egg mass (Figure 3.23). In the susceptible line 23S, high transcript abundance of *Tr-KPI1* was observed at day 4 and day 8 and at day 4 for *Tr-KPI2* and *Tr-KPI5.* A higher level of transcript abundance was observed in the susceptible line 23S when compared with the resistant line 17R for *Tr-KPI2* and *Tr-KPI5* at day 4 (Figure 3.21). The transcript level of root specific *Tr-KPI4* was only found to be upregulated in the resistant interaction and remain unchanged in the susceptible interaction suggesting a

specific role of this member in nematode defense. In the resistant interaction, the transcript level of *Tr-KPI* genes decreased by day 8 compared to day 4 which might be an indication that after invading the root cortex, the number of viable nematodes decreased and they were no longer feeding at a higher rate as they were on day 4. For the susceptible interaction, a higher transcript level at day 4 and lower at day 8 for three of the *Tr-KPIs* indicates that the viable nematodes already established in the root by day 4 and their vigorous feeding triggered the expression of *Tr-KPIs*, but the level of expression was not high enough to deter the invasion. In this study, the aim was to explore the activation of *Tr-KPIs* in response to nematode infestation. The four *Tr-KPI* genes showed relatively high expression in the resistant interaction when compared with the susceptible interaction for the cyst nematode. Therefore it can be proposed that *Tr-KPIs* in white clover are important in defense against white clover cyst nematode in combination with other defense genes.

To summarize the biotic stress response studies, both basipetal (downward) and acropetal (upward) induction of *Tr-KPIs* using two different groups of plant herbivores (*S. litura* and *H. trifolli*) confirms the probable role of this gene family in white clover defense mechanisms. The relative transcript abundance in response to insect herbivory (about 1000- fold for *Tr-KPI1* after 24 h) and in the nematode experiment (2.5 fold increase for *Tr-KPI1* in 17R at day 4) can not be compared directly. For insect herbivory, the 0 h expression level was used as a calibrator according to Pfaffl equation (2001) while for the nematode experiment, the 0 h time point was not used. However, the high abundance of *Tr-KPIs* induction in response to insect herbivory does suggest that the *Tr-KPIs* could be more involved in defense against chewing insects rather than the herbivore that directly sucks the cell sap from the vascular bundle. In general, the results also highlight that a systemic wound signal moves downward and upward to alert the distant plant parts. Current understanding of such systemic signals suggests that JA or JA derivatives, ABA or ethylene might be the potential candidate as they are induced by herbivory, they are mobile inside the plant via phloem or xylem and they act as a regulator of resistance (Soler *et al.,* 2012). However, further research is required to confirm these factors as activating transcriptional cues.

4.4 Tr-KPI genes under abiotic stress

To address the question as to whether *Tr-KPIs* are also involved in abiotic stress responses, two different kinds of stress were imposed: osmotic (water stress) and nutrient stress (Pi-deficiency).

The role of plant proteinase inhibitors under water deficit conditions is still unclear although some work with cystatin (cysteine proteinase inhibitor) in over-expressed transgenic *Arabidopsis* lines showed increasing tolerance to high salt, drought, oxidative and cold stresses (Zhang *et al.,* 2008) which could be a pleiotropic effect because of over-expression. Though Downing *et al*., (1992) in *B. napus* and Kang *et al.*, (2001) in potato showed that *KPIs* are induced by water deficiency; they did not clarify the question as to whether this response is transient or an adaptive feature. Among different abiotic stresses imposed on plants, water deficiency is one of the most frequent and serious natural drawbacks. The morphological adaptation features to drought involve a deep root system, production of more root hairs and leaf rolling whereas the physiological features include alteration of photosynthesis, osmotic adjustment and a high rate of water use efficiency (Stolf-Moreira *et al*., 2010). However, the perception of water deficiency is triggered at the molecular level before any visible morphological or physiological responses are detected. Each of these early molecular responses is triggered by a different stress signaling pathway which can be classified into ABA-dependent or ABA-independent pathways (Shinozaki and Yamaguchi-Shinozaki, 1997). During the early commencement of drought, several transcription factors get activated by either of these pathways. The common transcription factors in response to drought or water deficiency identified in different plant species include different members of the DRE-binding protein (DREB) family, ABA-binding factor (ABF), C-repeat-binding factor (CBF), MYC and MYB transcription factors (Abe *et al.,* 1997; Sakuma *et al.*, 2006 ; Harb *et al*., 2010). These transcription factors then activate the downstream stress response genes which are needed for the survival of the plants by modifying their adaptation features that guide drought tolerance.

Being a temperate pasture legume, white clover shows optimum growth during the late spring and early summer which is the period of drought (Hutchinson *et al.,* 1994 and Hunt *et al.,* 2002). Therefore, a foremost objective in this study was to investigate the response of *Tr-KPI* genes in response to a water deficit. Two sets of treatments were used in which one group of plants was primed to a water deficiency (PS) for a week to observe whether the primed sets of plants can display an earlier or enhanced response in terms of the accumulation of *Tr-KPI* genes compared to the non primed plants (NPS), and also to see whether water stress can induce the expression of *Tr-KPI* genes. To prove that water stress is imposed, the expression level of a homologue of *NCED1* from tomato, which is a major player in the ABA biosynthesis pathway, was also determined. Under normal conditions, the level of ABA is maintained at a minimal concentration for normal plant growth. ABA triggers plants responses to adverse

condition such as water deficit, salt or cold stress by increasing its endogenous level (Zeevart, 1998). Zeaxanthin epoxidase (ZEP), 9-cis-epoxycarotenoid dioxygenase (NCED) and abscisic aldehyde oxidase are the three main enzymes in the ABA biosynthetic pathway (Marin, *et al.*, 1996; Schwartz *et al*., 1997 and Seo *et al*., 2000, Shinozaki *et al.*, 2001). The first devoted step in ABA biosynthesis pathway is considered to be the oxidative cleavage of epoxycarotenoid 9-*cis*-neoxanthin to xanthoxin by NCED (Schwartz *et al.*, 1997). It has been reported by Shinozaki *et al.*, (2001) that the elevated level of endogenous ABA production under drought conditions is dependent on *AtNECD3*, which is induced by drought.

In the experiments reported here, *Tr-NCED1* level was up-regulated in the NPS treatment when compared with the initial moisture content (Figure 3.26). Under the NPS treatment, the transcription level of *Tr-KPI1* and *Tr-KPI5* were induced significantly in the leaf tissue when compared with the control whereas the transcription level of *Tr-KPI2* remained unchanged (Figure 3.27). Interestingly the one week pre-stress treatment triggered the expression of all three genes studied which were significantly higher when compared to the expression level under the NPS treatment (Figure 3.27). In root tissue, only *Tr-KPI5* responded to the PS treatment (Figure 3.28). The up-regulation of *Tr-NCED1*, *TR-KPI1* and *Tr-KPI5* under the NPS treatment suggests a connection between the ABA-dependent pathway and *Tr-KPIs* induction. Whereas under the PS treatment, the involvement of the ABA-independent pathway is more prominent as the *Tr-NCED1* transcript did not show significantly higher abundance and the *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* transcripts showed a higher and earlier induction when compared with the NPS treatment.

Priming of plants to a specific stress is proposed to make them more resistant to a future exposure of the same stress (Bruce *et al*., 2007). It can be speculated that the plants under the PS treatment showed the enhanced protection provided by *Tr-KPIs* at the transcriptional level without the cost associated with the activation of different early stress induced genes such as *Tr-NCED1.* Though hormonal crosstalk and the possible triggering mechanism for these *Tr-KPIs* is unknown, this result clearly implies the involvement of *Tr-KPIs* under water stress and provides evidence again that the function of the *KPIs* is not solely restricted to one of storage protein in seed tissue or in response to biotic stress. As the transcription of *Tr-KPIs* was up-regulated, therefore, it is possible that protecting storage proteins from proteolysis by proteinases under water stress in the above ground parts might be one of the possible functions of *Tr-KPIs*. The non-specific expression of *Tr-KPIs* in the root tissue except *Tr-KPI5* under both the NPS and the PS treatment did not indicate any clear pattern of transcriptional

regulation in the below ground parts under water stress. Because white clover plants have the ability to vegetatively propagate through stolon branching, thereby there may be preferential protection in the shoot then compared with the roots.

Therefore, based on the transcriptional study, it can be proposed that the activation of *Tr-KPIs* (*Tr-KPI1* and *Tr-KPI5*) expression in the NPS and PS treated FFE leaves is a reflection of the regulation of proteolytic processes in the white clover leaves as part of a response to water deficit.

To further characterize the role of *Tr-KPIs* under water stress, a drought tolerant ecotype Tiesnshan and a drought susceptible cultivar kopu was compared. A clear up-regulation of *Tr-KPI1* in Tienshan and *Tr-KPI5* in cv. Kopu at the transcriptional level was observed under the PS treatment (Figure 3.28 and 3.29). The higher transcript level of *Tr-KPI1* could indicate its contribution in Tienshan together with other abiotic stress response genes in the adaptive evolution of a drought resistant ecotype.

Another endeavor in the present study was to investigate the possible involvement of *Tr-KPI* genes nutrient stress at the transcriptional level. The macro-nutrient phosphorus was chosen in this study, as under phosphorus starvation, root system architecture (RSA) is greatly altered. Here the length of the primary root is reduced and long and dense lateral roots are formed with dense root hairs (Pe´ret *et al.*, 2011) and the expression of *Tr-KPIs* were found to be developmentally regulated in root tissues. Phosphorus (P) as an essential macro-nutrient, is required by the plants not only for different physiological functions including photosynthesis, the conversion of energy, mobility of nutrients within the plant, metabolism of starch but also for transferring hereditary characteristics from one generation to another (Sharpley *et al*., 1999). Adequate Pi supply to the plant helps to maintain normal growth and development, while prolonged Pi deficiency may result in visual symptoms including premature senescence, extended seed dormancy, stunted growth, abnormal tillering, and reduced flower production (Mengel and Kirkby, 2001). It has been hypothesized that phosphorus deficiency may also affect the ability of the plant to uptake other nutrients and thus, in correcting the phosphorus deficiency, other nutrient deficiency problems may also disappear. Under Pi-starved conditions, the response of plants can be divided into two categories. The first category involves biochemical and developmental adaptations to improve Pi acquirement and recycling and second category involve protection against the stress imposed by Pi starvation condition (Franco-Zorrilla, *et al.,* 2004). Therefore, transcriptional control of the genes responsive to Pi starvation is vital. Current understanding of the Pi-starvation changes in plants suggests a

complex hormonal signaling network involving ethylene and auxin (Dinh *et al.,* 2012), and involved up-regulation of Pi-starvation associated transcription factors, the Pi status of the whole plant, systemically controlled long-distance response (late Pi-starvation associated genes; Franco-Zorrilla, *et al.,* 2004) and local Pi controlled response (controlled by the external Pi concentration; Wu *et al.,* 2003).

In a split root experiment with *Arabidopsis thaliana,* Thibaud *et al*., (2010) found two proteinase inhibitor genes (*At1G73260* and *At1G73330*), belonging to the Kunitz sub-family of serine proteinase inhibitors were locally induced in the Pi starved roots. In Pi starved potato leaves, higher transcript abundance was observed of inhibitor genes belonging to the cysteine and aspartic proteinase inhibitor groups within 17 days of Pi starvation (Hammond *et al*., 2011). Though the transcript level of *PIs* was upregulated in both *Arabidopsis* and potato, further analysis as to whether their expression is just a common Pi stress response or a signaling response (i.e. involved in Pi stress survival machinery) was undecided. In both of the microarray experiments to induce Pi deficiency, phosphorus was completely removed from the media. Complete withdrawing of Pi from the media can totally alter plant behaviour mainly at the transcriptional level. Though phosphorus is one of the most limiting and inaccessible nutrients (Hammond *et al*., 2011), plants can always absorb a sufficient amount of Pi from the soil. Therefore, in this experiment, a hydroponic media consisting of 5µM KH_2PO_4 was used as a source of Pi to induce phosphorus deficiency in the plant. An effort has been undertaken to study the expression pattern of *Kunitz proteinase inhibitor* genes in high Pi and low Pi media to see whether the change in Pi concentration can regulate the expression of *Tr-KPI* genes in different root parts.

As *Tr-KPI* genes showed differential expression in different root zones, i.e. elongation zone (EZ), visible lateral root zone (VL) and main root zone without any lateral root (MR), the roots used for the Pi-deficiency experiment were also harvested in different root zones and analyzed separately. A higher level of *Tr-KPI expression* has been observed in the growing zone rather than mature root zone (Figure 3.9). This differential expression in transcriptional level may indicate that *Tr-KPI*s are playing an active role in protecting the young growing tissues by acting as an active regulator of proteinases during root growth and development. The up-regulation of *Tr-KPI* genes in the elongation zone due to phosphorus deficiency may support the protective role of *Tr-KPI*s at the growing point. The result also suggests that the induction of *Tr-KPI* genes might not just be a general response in the EZ zone as the high level of transcripts were observed after 24 h onward (Figure 3.32). Though all four *Tr-KPIs* was up-

regulated in the EZ region, only *Tr-KPI2* and *Tr-KPI4* showed an extended level of expression in the visible lateral root zone suggesting their involvement in lateral root formation as Pi-deficiency condition induces a higher number of lateral root primordia (Guo *et al.,* 2011 and Dinh *et al.,* 2012) (Figure 3.33). The effect of Pi-deficiency on the expression of *Tr-KPI1* and *Tr-KPI5* from 6 h to 24 h indicates only transient expression in the visible lateral root zone. In the mature root zone, the transcript level of *Tr-KPI1, Tr-KPI2* and *Tr-KPI4* showed an earlier induction at 1 h of the Pi-deficiency treatment, further indicating transient or local expression (Figure 3.34). In the leaf tissue, the down regulation of *Tr-KPIs* has been observed up to 12 h of the treatment and the transcript level started to increase from 24 h onwards (Figure 3.35). This result indicates that it took at least 12 h to systematically induce the response of *Tr-KPIs* in the leaf tissue. These further support the view that *Tr-KPIs* are not the early responsive gene under Pi deficiency condition in the leaf tissue. In addition to the local response, the *Tr-KPI* genes may also be part of the systemic response. They may play a role in cell protection as cell metabolism becomes compromised due to the prolonged Pi stress.

The relative transcript abundance for abiotic experiments, i.e. osmotic stress and Pi deficiency, should not be compared with the herbivory experiments. If the *Tr-KPIs* are targeted to the vacuole (Table 3.1), it can be speculated that chewing of the leaves by *S. litura* ruptures the vacuole releasing *Tr-KPIs* together with other defense molecules which in turn trigger the biotic defense response pathway. However, localization and activities of both proteinases and their inhibitors indicates that many physiological processes depend on their delicate balance which is essential for adaptation to environmental conditions (Brzin and Kidric, 1995). Research concerning the involvement of *KPIs* in response to environmental stress other than biotic stress is rare. The change in the transcript abundance of *Tr-KPIs* under abiotic stress is only the first step and therefore, functional analysis of the gene products under such conditions is crucial to define their possible involvement. If *Tr-KPIs* are indispensable under abiotic stress conditions, and as the research suggests that *KPIs* have inhibitory activities against trypsin, chymotrypsin, elastase, papain and subtilisin (Brzin and Kidric, 1995, Birk, 2003, Major and Constabel, 2008) *in vitro*, it would be very worthwhile to investigate the involvement of proteinases and their inhibitors under such conditions.

4.5 Cis-Regulatory elements of Tr-KPIs

The findings in this study suggest that the expression of the *Tr-KPI* genes in white clover is under the control of different transcription factors which can act as activators or repressors for these genes. The bioinformatics-based analysis used in this study gives an indication of the nature of the TF binding sites in the promoter region. A number of growth and development related transcription factor binding sites such as AREF (*Tr-KPI1* and *Tr-KPI2*), ASRC (*Tr-KPI2* and *Tr-KPI4*), LFY (*Tr-KPI1, Tr-KPI2* and *Tr-KPI4*), MADS (*Tr-KPI1, Tr-KPI2, Tr-KPI4* and *Tr-KPI5*) are observed in the promoter regions of the identified *Tr-KPI* genes. This supports their possible involvement during growth and development as it has been found in this study that *Tr-KPI* genes are developmentally regulated in leaf and root tissue and also during the germination time course.

As an example, some of the AREF (auxin response factors) transcription factors contain transcriptional activation domains, while others contain repression domains regulating the expression of auxin response genes by binding to the auxin response element binding factors (AREF) in the promoter region in an auxin-regulated manner (Guilfoyle, *et al*., 1998). The presence of AREF in the promoter region of *Tr-KPI1* and *Tr-KPI2* indicate the involvement of these two *KPIs* in plant growth and development in an auxin-dependent manner either by activating or suppressing the transcription of these genes. LFY, which is the meristem identity gene, is required for the switch from vegetative to reproductive development (Parcy *et al*., 1998), and LFY binding elements has been observed in the *Tr-KPI1, Tr-KPI2* and *Tr-KPI4* promoter regions indicating their involvement in flower development in a yet unknown way. A number of ASRC (AS1/AS2 repressor complex) binding sites in the *Tr-KPI4* promoter region correlates with the non-detection of this particular transcript in the leaf tissue, and therefore it can be speculated that the repression of *Tr-KPI4* is important for leaf formation. Guo *et al.,* (2008) determined that AS1 and AS2 form a repressor complex that binds directly to the regulatory motif in the promoter region of the Knotted1-like homeobox (*KNOX*) genes which promotes stem cell activity and must be repressed to form determinate lateral organs, that is the leaf. ASRC binding sites has also been observed in the *Tr-KPI2* promoter region which may correlate with the lower transcript abundance in all leaf tissues when compared with *Tr-KPI1* and *Tr-KPI5*. Though most of the members of the MADS transcription factor family are involved in the transition to the reproductive stage of plant, their involvement during vegetative development such as in the embryo, root and leaf has also been shown (Alvarez-Buylla *et al.,*2000). MADS binding sites are observed in all of the four genes and it has been found in this

study that *Tr-KPI2* and *Tr-KPI5* are developmentally regulated in the leaf and root tissues. The transcriptional study regarding flower development has not been done in this study but high transcript abundance of *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* found in the inflorescence tissue indicates their possible involvement in flower and fruit development (Figure 3.7B).

Further, transcription factors responsive to water deficit such as the ABA responsive element (ABRE) were observed in *Tr-KPI1* and *Tr-KPI5*, the dehydration stress responsive element (DREB) in *Tr-KPI2* and stress activated VIP1 binding site (VRES) and MYBs (Jin and Martin, 1999) in *Tr-KPI1* and *Tr-KPI2* and calcium regulated NAC factor (CNAC) were found in the promoter region of *Tr-KPI5*. The presence of these TFBS in the promoter region highlights the involvement of *Tr-KPIs* during water deficiency as high transcript abundance was observed for *Tr-KPI1, Tr-KPI2* and *Tr-KPI5* under the NPS and PS treatments. For example, the DREB which is known as an important cis-acting element in response to drought in an ABA-independent manner (Zhang *et al*., 2008; Liu *et al*., 1998) has been observed in the *Tr-KPI2* promoter region. The presence of DREB positively correlates with the expression pattern of *Tr-KPI2* under the PS treated plant but not in NPS treated plant (Figure 3.26 and 3.27). On the other hand, the presence of ABRE binding sites in *Tr-KPI1* and *Tr-KPI5* positively correlates with the high transcript abundance of these two genes under the NPS treatment in an ABA dependent manner. It can also be speculated that under the PS treatment, activation of *Tr-KPI1* and *Tr-KPI5* transcript is under the control of MYB, MYC and NAC transcription factors. The MYB and MADS group binding sites that are involved in Pi-deficiency were observed in the promoter region of all four genes (Thibaud *et al.*, 2010) highlighting their involvement during Pi limitation. TFBS for MYB, NAC and MYBL, which are involved in wounding response, were found in all four promoter region indicating their responsiveness to herbivory attack. The presence of these TFBS also indicates that either these genes are activated or repressed by these transcription factors depending on the type of herbivory attack. Different groups of *PIs* can be responsive to herbivory attack (Hartl *et al.,* 2010) and all four *Tr-KPIs* were not responsive to the same level under *Spodoptera* or cyst nematode attack.

Both different biotic and abiotic stress responses of this gene family in white clover therefore can correlate with the presence of these TFBS. The analysis of the TFB sites was very stringent and the findings give a broad idea of the regulatory factors involved in the expression of *Tr-KPIs* in white clover. However, more direct promoter activity analysis *in vivo* is required, including for example, deletion of the potential binding sequence and fusing the modified promoter with a reporter gene (β-glucuronidase) to help to confirm any of these putative

domains (Liu *et al.,* 2006) and to more direct evidence of the activation or repression of the identified *Tr-KPI* genes.

4.6 Tr-KPIs RNAi lines

Tr-KPIs knockdown RNAi lines were classified depending on the relative transcript abundance, into three categories: high, medium and low (section 3.6). Surprisingly, knockdown of the target gene showed some effect on the non-target *Tr-KPI* genes. For example, *Tr-KPI1* was found to be highly expressed in three *Tr-KPI2* low category KD lines and showed the same level of expression in the *Tr-KPI5* lines when compared with the wild type and negative controls. The expression of *Tr-KPI2* was found to be much lower when compared with the wild type and negative control in the *Tr-KPI1* low category KD lines, whereas, in the *Tr-KPI5* lines, *Tr-KPI2* maintained a low level of expression. The expression of *Tr-KPI5* in the *Tr-KPI1* lines was found to be similar when compared with wild type and lower when compared with the negative controls. The RNAi constructs were designed in the non-overlapping region of the *Tr-KPIs* gene sequence, and so any increase in transcript abundance of the non-target *Tr-KPIs* could indicate a compensation effect. However, no statistical analysis has been performed with the RNAi lines as this study only compares the level of expression in different RNAi lines. The discrepancies in the transcript level need to be clarified by western blot analysis with specific antibodies for each of the four genes.

4.7 Medicago vs. white clover Kunitz proteinase inhibitors

To further explore the functions of the white clover *Kunitz* family members, a *Medicago truncatula* microarray data from the JCVI database (http://mtgea.noble.org/v2/) for 12 *KPI* genes was used to elucidate whether the members of *KPI* gene family are expressed in the same way as they are in white clover, including for example during development (Figure 4.1). The heat map data suggested a similar kind of expression of the *Tr-KPIs* orthologs in *Medicago* except *Tr-KPI5* (63% indentity with the closest member Medtr6g009650). The *Tr-KPI1* and *Tr-KPI2* orthologs *Medtr6g50730.1* (81% identity with Tr-KPI1) and *Medtr6g09680* (72% identity with Tr-KPI2) respectively, were found to be expressed in all vegetative tissue studied in *Medicago* which is similar to the expression of *Tr-KPI1* and *Tr-KPI2*. The *Tr-KPI4* ortholog *Medtr6g059650* (72% identity with Tr-KPI4) showed more tissue specific expression in total root, root tip and nodule tissue. Though expression data for different leaf and root developmental stages are absent in the *Medicago* microarray study, a similar expression pattern was observed for Tr-*KPI1, Tr-KPI2* and *Tr-KPI4* orthologs in the *Medicago* heat map for

the root tip tissue and in the nodule tissues (Figure 3.10). Under osmotic stress, though NaCl was used for the microarray study, the *Tr-KPI1, Tr-KPI2* and *Tr-KPI4* orthologs were induced in root tissue suggesting their involvement to the abiotic stress response. Methyl jasmonate (MeJA) application to the leaf tissue induced the expression of *Tr-KPI1, Tr-KPI2* and *Tr-KPI4* orthologs *Medtr6g50730.1, Medtr6g09680* and *Medtr6g059650* respectively suggesting their contribution to the biotic stress response induced by the JA mediated pathway. Though MeJA treatment was not performed in this study to provide evidence that *Tr-KPIs* are regulated by the JA mediated pathway, herbivory induction of *Tr-KPIs* in this study could point towards the involvement of JA in the induction of *Tr-KPIs* under biotic stress. The similarity in the expression pattern of *Tr-KPI1* with *Medtr6g50730.1, Tr-KPI2* with Medtr6g09680 and *Tr-KPI4* with *Medtr6g059650* suggests that they perform similar functions in both species and so may have evolved under the same evolutionary pressure in both species. *Tr-KPI5* in white clover was found to be developmentally regulated in the leaf and root tissue and during germination and is responsive to both biotic and abiotic stresses, whereas the ortholog *Medtr6g009650* in *Medicago* was found to be expressed in flower and during early seed developmental stages only. This discrepancy in expression of *Tr-KPI5* with the closest member *Medtr6g009650* suggests that this member in perennial white clover may have been evolved to perform additional roles beyond only those involved in seed development.

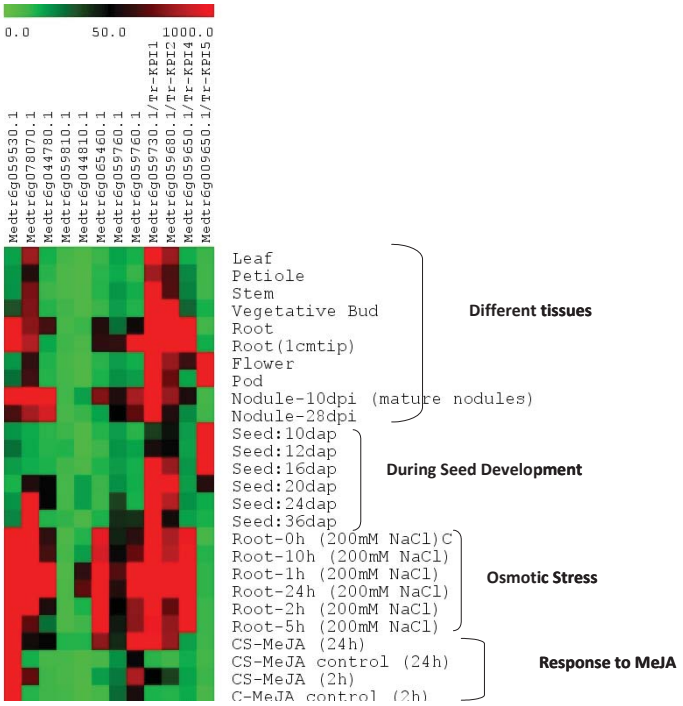


Figure 4.1: *KPI* gene expression data (Affymetrix Medicago Gene Chip ®) in *M. truncatula* (Jemalong A17). The JCVI database (http://mtgea.noble.org/v2/) was queried for the *KPI* genes and their expression level. Medicago orthologs of *Tr-KPI1, Tr-KPI2, Tr-KPI4*and *Tr-KPI5* are also indicated.

4.8 Summary and Conclusion: The Bigger Picture

This thesis is centered around the *Kunitz proteinase inhibitor* genes in white clover. The results of the transcriptional study of the *Tr-KPI* genes suggest that the functions of members of this gene family are not restricted simply to seed storage. All of the four genes investigated in this study at the transcriptional level showed differential expression in different plant parts highlighting that they are also developmentally regulated, particularly *Tr-KPI2* and *Tr-KPI5*. As *Tr-KPI1* and *Tr-KPI5* showed relatively high expression in some of the low expressing *Tr-KPI2* RNAi lines, it will be quite interesting to see whether there is any morphological differences present in these KD lines. Therefore, the low expressing RNAi KD lines can help in investigating their involvement in growth and development.

It has been observed that *Tr-KPI* genes in white clover are induced by both biotic and abiotic stress (Table 4.1). *Tr-KPI1* is always the early induced gene which has an extended expression level over all the biotic stress experiments in the local and systemic tissue. The closest member to *Tr-KPI1* among the four identified genes is *Tr-KPI2* (74% identity at the amino acid level) which behaved the same way as *Tr-KPI1*, but showed a later induction and less transcript abundance. On the other hand, *Tr-KPI4,* is a very distinct member of Tr-*KPI* gene family as it is expressed prominently in root tissue, but is also a biotic stress responsive gene. Considering the earlier response to biotic stress, *Tr-KPI5* falls into the intermediate category.

In view of the abiotic stress responses, *Tr-KPI1* and *Tr-KPI5* behaved the same way under the water deficit study for both the NPS and the PS treatment in the leaf tissue. The induction of *Tr-KPI2* under the PS treatment suggests that priming is more important for this particular member to be induced in response to a water deficit. Surprisingly in the root tissue, only *Tr-KPI5* showed a significant increase in the transcript level in primed plants where the levels for other members of *Tr-KPIs* were inconsistent suggesting an as yet unknown role for *Tr-KPI5* in root tissue under water deficiency. Therefore, *Tr-KPI1* and *Tr-KPI5* could be the most important *KPIs* members in plant defense in response to osmotic stress conditions. The macro-nutrient Pi-deficiency study suggests that *Tr-KPI4* could be the most important gene followed by *Tr-KPI2* among the four *Tr-KPIs* as they showed earlier and extended stable induction in the elongation and visible lateral root zone. This suggests their involvement in changing root system architecture that occurs in response to Pi-deficiency. The low category RNAi KD lines will be an important tool to further categorize the function of this gene family.

The constitutive expression of *Tr-KPIs* (*Tr-KPI1, Tr-KPI2* and *Tr-KPI5*) in different plant parts indicates their involvement against their target proteinases throughout the life cycle of the plant depending on the stage of development. Expression of *Tr-KPI4* in the root tissue might indicate a specific role in root tissue. Induced expression of the same *Tr-KPIs* under biotic and abiotic stress highlights their involvement in plant defense system. All these results suggest functional diversification of the members of *Tr-KPI* gene family in white clover.

KPIs form a multigene family not only in the legume species soybean and *Medicago* but also in *Arabidopsis*, poplar and potato. The presence of multigene families in higher plants is quite common such as the *PR10* gene family (Lebel *et al.,* 2010). Tandem duplication, which is the possible cause of evolution of multigene families, is quite common in plants. For example, duplication represents about 16% of *Arabidopsis* genes (Rizzon *et al.,* 2006). Depending on the type of stress and growth conditions, the new copies of a gene following duplication may then undergo alterations which allow functional diversification (Lebel *et al.,* 2010). Therefore, it can be speculated that the presence of *Tr-KPI* multigene family in white clover provides significant adaptive advantages by permitting a differential regulation of each member of the family in response to different endogenous (developmental) cues and exogenous, that is environmental stimuli. The results presented in this study can contribute important information to our current knowledge about the *KPI* gene family and their function in plants and highlights the need for continuous investigation of the *Tr-KPI* members in white clover.

Therefore, depending on the results in this study it can be proposed that *Tr-KPIs* are tri-functional in white clover (Figure 4.2) and possibly also in *M. truncatula*. That is they are involved in plant growth and development, act as defense responsive genes under herbivore attack and regulate the elevated level of proteinase activity under abiotic stress.

Table 4.1: Grouping of *Tr-KPIs* based on the expression in growth and development; biotic and abiotic stress response at the transcriptional level

Gene	Develop-mental cue	Biotic Stress							Abiotic stress							
		Mechanical wounding		Insect herbivory			Nematode herbivory		Water deficiency				Nutrient Stress			
		Local	Systemic (Root)	Local	Systemic (Apex)	Systemic (Root)	Local	Systemic (Leaf)	NPS		PS		EZ	VL	MR	Leaf
									Leaf	Root	Leaf	Root				
Tr-KPI1																
Tr-KPI2																
Tr-KPI4																
Tr-KPI5																

Higher and extended level of expression

Intermediate level of expression

Not detected/Lower level of expression

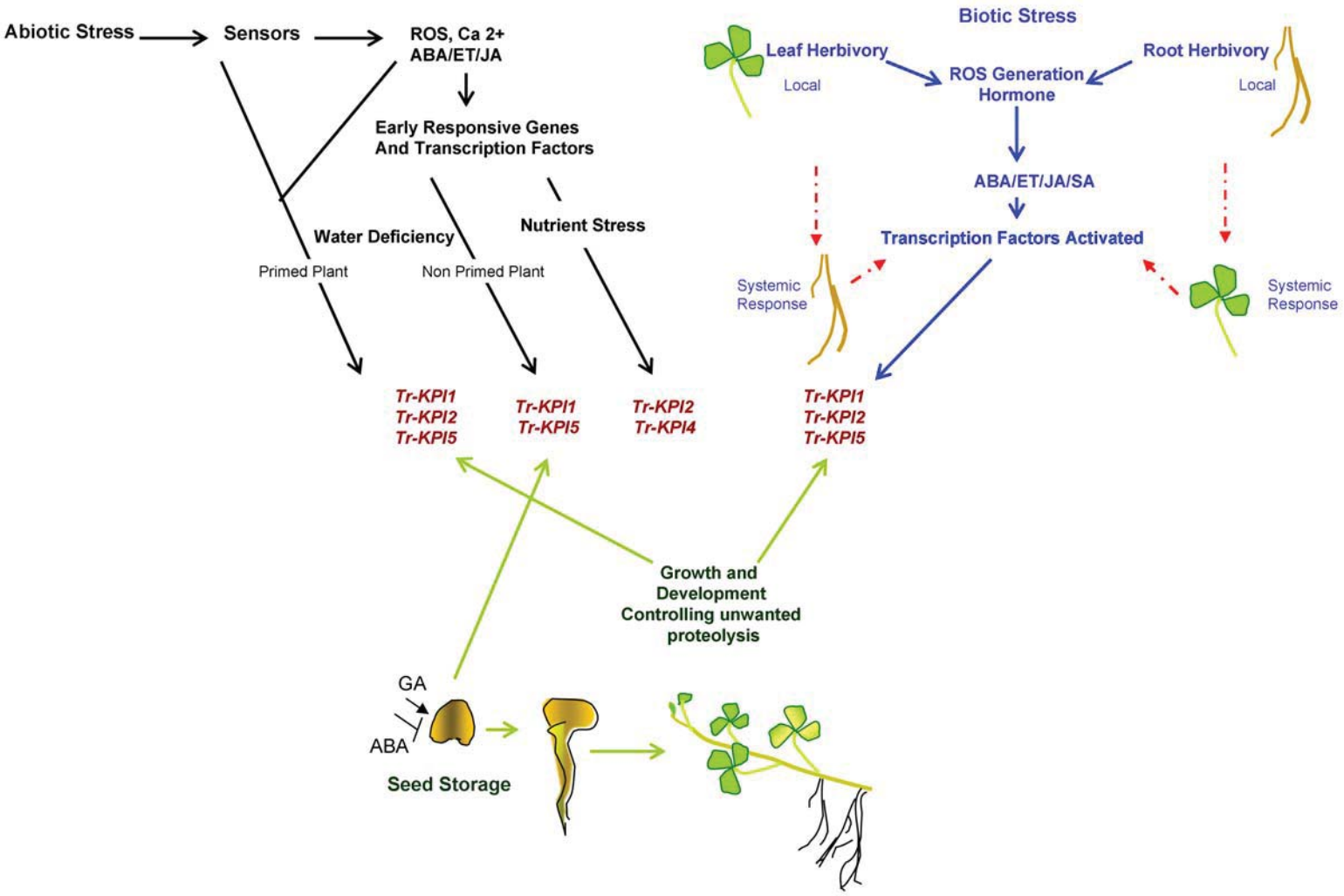


Figure 4.2: Proposed tri-functional role of *Kunitz Proteinase Inhibitors* in white clover under abiotic (black), biotic (blue) stress responses and under normal growth and development (green).

4.9 Future Directions

A role in growth and development and in stress responses for the *Tr-KPIs* represents an extensive involvement of this gene family in white clover. However many aspects of the functionality, such as activity against proteinases of *Tr-KPIs* remained undiscovered. Therefore, the following experiments can help in answering a whole set of questions concerning this gene family.

4.9.1 Growth and development analysis of Tr-KPIs RNAi knockdown lines

RNAi knockdown lines for each of the four *Tr-KPI* genes have been developed in white clover. Preliminary analysis has shown that the lines expressing a lower abundance of *Tr-KPIs* show altered development when compared with the wild type plants. This includes a shorter stolon and increased primary root length (data not shown). Measurements of different growth parameters for both above and below ground parts of the plants are needed to further characterize these plants. Any difference in the growth parameters of the knockdown *Tr-KPI* lines will clearly indicate a more direct involvement of the expression of the *Tr-KPI* gene family in endogenous functions and their importance to the plant under normal growth conditions.

The same abiotic experimental setup reported in this thesis can also be used with the RNAi knockdown lines. The performance of Tr-KPI knockdown lines under these stresses will further clarify their direct functional involvement. The water deficiency study from the NPS and PS treated plants suggested the involvement of *Tr-KPI1* and *Tr-KPI5.* Therefore, the result with the RNAi knockdown lines may show two distinct outcomes. The first might be that the knockdown plants display a reduced tolerance due to the lack of *Tr-KPI* expression when compared with the control indicating the direct involvement of *Tr-KPI* expression under abiotic stress response. Alternatively, an increased tolerance to water deficiency may be observed as other members of the *Tr-KPI* family (or other stress-associated genes) may be activated to protect the plant. Measurement of fresh and dry weight of the above and below ground parts of the plant and other parameters such as leaf number, petiole length, stolon length and branch number will provide evidence for these two speculated outcomes. The macro-nutrient Pi-deficiency study suggested that *Tr-KPI4* followed by *Tr-KPI2* could be the most important genes among the four *Tr-KPIs* as they showed earlier and extended stable induction of expression in the elongation and visible lateral root zone. Therefore, a Pi deficiency study with

the knockdown lines could also yield two different outcomes based around an exaggerated change in root system architecture to that normally observed in response to Pi deficiency (increase in primary root elongation; increase in lateral root number) such that either a reduced primary root length and lower rate of lateral root formation or a longer primary root length and higher rate of lateral root formation is observed. Again, a reduced primary root length and lower lateral root formation may suggest a direct involvement of *Tr-KPI4* and *Tr-KPI2* in root development under Pi deficiency while an extended primary root length and higher rate of lateral root formation will provide evidence the involvement of other yet unidentified *Tr-KPI* members in white clover (or that *Tr-KPI4* and *Tr-KPI2* serve to repress root growth). Further anatomical study of the root elongation zone of these knockdown lines will also show whether there is any alteration in root development due to a lack of *Tr-KPI* transcription under Pi deficiency.

4.9.2 Analysis of Tr-KPIs inhibitory activity against different proteinases

Tr-KPIs have been over-expressed under the CaMV35S promoter in tobacco plants. The plants have already been screened for protein accumulation, and purification of Tr-KPI1, Tr-KPI2, Tr-KPI4 and Tr-KPI5 is under way using G75 column and ion exchange chromatography, in collaboration with Dr. William Laing (Plant and Food Research, Auckland). So far, the preliminary inhibitory data (not shown) using purified Tr-KPI1 suggests that the over-expressed lines have a higher inhibitory activity against trypsin compared with the wild type control. Further analysis will be carried out to check their inhibitory activity using different commercially available proteinases. These data will assist with characterizing the biochemical activity of *Tr-KPIs* in terms of activity against putative target proteinases, or whether some or all of the *Tr-KPI* gene products display other cellular functions. Such information is important in terms of interpreting the cellular mechanisms by which any developmental or stress-associated changes observed in the RNAi knockdown lines are mediated.

Further, *Spodoptera litura* will be fed on the leaf material of over-expressed and RNAi knockdown lines. Their performance (eg. growth rate) will indicate whether the expression of the *Tr-KPI* genes has any influence on the development of feeding insect herbivores. The expected outcome of the study will be the loss in weight of the larvae feeding on tissue from the over-expressed tobacco lines and gain in weight of the larvae fed tissue from the RNAi knockdown white clover lines. However, there is also a possibility of seeing the converse result in which a loss in weight in insects fed tissue from the knockdown lines. This will indicate the

activation of other defense-related genes due to the RNAi knockdown of the target *Tr-KPI* genes leading to leaf material with an increased degree of toxicity to feeding insects.

4.9.3 Identification of other Tr-KPIs members

Here in this thesis, four full length and four partial sequences were identified and focus given to the characterization of four full length genes during growth and development and in response to biotic and abiotic stresses. As this study suggests that these gene family members can be multifunctional, it is important to identify and characterize other possible members if they exist. A full transcriptome study using tissues from different developmental stages, and different plant material exposed to abiotic and biotic stress will help to give necessary information about additional family members.

4.9.4 Growth and development analysis of the Medicago knockout lines

The *Medicago* knockout lines for each of the orthologs of the four *Tr-KPI* genes have been received from the Samuel Roberts Noble Foundation and the plants were grown for seed collection for different sets of experiments. Growth and developmental parameters with these knockout lines will provide information about the function of these *KPIs* in the model plant species. The correlation of the results of white clover RNAi knockdown lines and *Medicago* knockout lines, particularly in terms of the most identical PIs (Figure 4.1) in the two species will also provide clues as to the role of these KPIs in white clover (and *Medicago*).

References

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D. and Shinozaki, K. 1997. Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell, 9:1859–1868.

Abe H., Shimoda T., Ohnishi J., Kugimiya S., Narusaka M., Seo S., Narusaka Y., Tsuda S. and Kobayashi M. 2009. Jasmonate dependent plant defense restricts thrips performance and preference. BMC Plant Biology, 9(97):1-12.

Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F. and Yanofsky, M.F. 2000. MADS-box gene evolution beyond ¯owers: expression in pollen, endosperm, guard cells, roots and trichomes. The Plant Journal, 24(4), 457-466.

Araujo, A.P., Hansen, D., Vieira, D.F., Oliveira, C., Santana, L.A., Beltramini, L.M., Sampaio, C.A., Sampaio, M.U. and Oliva, M.L. 2005. Kunitz-type *Bauhinia bauhinioides* inhibitors devoid of disulfide bridges: isolation of the cDNAs, heterologous expression and structural studies. Biological Chemistry, 386: 561–568.

Atwood, S.S. and Hill, H.D .1940. The regularity of meiosis in microsporocytes of *Trifolium repens*. American Journal of Botany, 27:730–735.

Bailey-Serres, J. 1999. Selective translation of cytoplasmic mRNAs in plants. Trends in Plant Science, 4(4): 142-148.

Barrett, A.J. 1986. The classes of proteolytic enzymes. *In* M.J. DaUing (ed.), Plant Proteolytic Enzymes. Boca Raton, CRC Press, pp. 1-16**.**

Baumlein, H., Nagy, I., Villaroel, R., Inzb, D., and Wobus, U. 1992. Cis-analysis of a seed protein gene promoter: The conservative RY repeat CATGCATG within the legumin box is essential for tissue- specific expression of a legumin gene. The Plant Journal, 2:233-239.

Bauw, G., Nielsen, H. V., Emmersen, J., Nielsen, K. L., Jørgensen, M and Welinder, K. G. 2006. Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. FEBS Journal, 273:3569–3584.

Birk, Y. 2003. In: Plant Protease Inhibitors: Significance in Nutrition, Plant Protection, Cancer Preventionand Genetic Engineering. Ed. (Berlin Heidelberg: Springer-Verlag) p. 1-126

Bode, W. and Huber, R. 2000. Structural basis of the endoproteinase-protein inhibitor interaction. Biochimica et Biophysica Acta, 1477:241-252.

Bohac, J.R. 1991. A modified method to purify patatin from potato tubers. Journal of Agriculture and Food Chemistry, 39:1411–1415.

Boigegrain R. A. Mattras, H., Brehelin, M. and Coletti-Previer, M.A. 1994. Invertebrate proteinase inhibitors. Pure and Applied Chemistry, 66(1):1-7.

Boter, M., Ruı´z-Rivero, O., Abdeen, A., and Prat,S. 2004. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis.* Genes and Development, 18:1577–1591.

Bruce, T.J.A., MAtthes, M.C., Napier, J.A. and Pickett, J.A. 2007. Stressful ‘‘memories’’ of plants: Evidence and possible mechanisms. Plant Science, 173:603–608.

Brzin, J. and Kidric, M. 1995. Proteinases and their inhibitors in plants: role in normal growth and in response to various stress conditions. Biotechnology and Genetic Engineering Review, 13: 420-467.

Cai, D., Thurau, T., Tian, Y., Lange, T., Yeh, K.W. and Jung, C. 2003. Sporamin-mediated resistance to beet cyst nematodes (*Heterodera schachtii* Schm.) is dependent on trypsin inhibitory activity in sugar beet (*Beta vulgaris* L.) hairy roots. Plant Molecular Biology, 51:839–49.

Cartharius, K., Frech, K., Klocke, B., Haltmeier, M., Klkngenhoff, A., Frisch, M., Bayerlein, M. and Werner, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics, 21:2933-2942.

Cathcart, E. P. 1904. On The Antitryptic Action of Normal Serum. The Journal of Physiology, 31(6): 497-506.

Chrispeels, M.J. and B. Baumgartner. 1978. Trypsin inhibitor in mung bean cotyledons: purification, characteristics, subcellular localization, and metabolism. Plant Physiology, 61: 617-623.

Chye, M.L., Sin, S.F., Xu, Z.F. and Yeun, E. C. 2006. Serine proteinase inhibitor proteins: exogenous and endogenous functions. InVitro Cellular and Developmental Biology-Plant, 42:100–108.

de Castro, L.A.B., Carneiro, M., Monte-Neshich, D.C., de Paiva, G.R. 1992**.** Spatial and temporal gene expression patterns occur during corm development. The Plant Cell, 4: 1549–1559.

DiBella, F.P. and Liener, I.E.1969. Soybean trypsin inhibitor: cleavage and identification of a disulfide bridge not essential for activity. Journal of Biological Chemistry, 244: 2824–2829.

Dinh, P., Roldan, M., Leung, S. and McManus, M.T. 2012. Regulation of root growth by auxin and ethylene is influenced by phosphate supply in white clover (*Trifolium repens* L.). Plant Growth Regulation, 66: 179-190.

do Socorro,M., Cavalcanti,M., Fritz,H., Jochum,M., Mentele,R., Sampaio,M., Coelho,L.C., Batista,I.F. & Sampaio,C.A. 2002. Characterization of a Kunitz trypsin inhibitor with one disulfide bridge purified from *Swartzia pickellii*. Biochemical and Biophysical Research Communications, 291: 635-639.

Downing, W.L., Mauxion, F., Fauvarque, M.O., Reviron, M.P., DeVienne, D., Vartanian, N., and Giraudat, J., 1992. A *Brassica napus* transcript encoding a protein related to the Künitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. The Plant Jpurnal, 2(5):685–693.

Duthie, E. S. and Lorenz, L.1949. Assay and nature of serum antiprotease. Biochemistry Journal, 44(2): 167-173.

Earnshaw, W.C. 1995. Nuclear changes in apoptosis. Current Opinion in Cell Biology, 7:337-343.

Ellis C., Karafyllidis I. and Turner J.G. 2002. Constitutive activation of jasmonate signaling in an *Arabidopsis* mutant correlates with enhanced resistance to *Erysiphe cichoracearum*, *Pseudomonas syringae*, and *Myzus persicae*. Molecular Plant-Microbe Interaction, 15:1025–1030.

Ellison, N.W., Liston, A., Steiner, J.J., Williams, W.M. and Taylor, N.L. 2006. Molecular phylogenetics of the clover genus (*Trifolium*-Leguminosae). Molecular Phylogenetics and Evolution, 39:688–705.

Emanuelson, O., Nielson, H., Brunak, S. and von Heijne, G. 2000. Predicting subcellular localization of proteons based on their N-terminal animo acid sequence. Journal of Molecular Biology, 300: 1005-1016.

Fan, S.G. and Wu, G.J. 2005. Characteristics of plant proteinase inhibitors and their applications in combating phytophagous insects. Botanical Bulletin of Academia Sinica, 46:273-292.

Fritz, H. 2000. Foreword. *In* K. von der Helm, B.D. Korant, and J.C. Cheronis (eds.), Proteases as Targets for Therapy. Berlin: Springer-Verlag, pp. V-VI.

Flexner, C. 1998. HIV-Protease Inhibitors. The New England Journal of Medicine. 338:1281-1293.

Franco-Zorrilla, J.A., GonzaÂ lez, E., Bustos, R., Linhares, F., Leyva, A. and Paz-Ares, J. 2004. The transcriptional control of plant responses to phosphate limitation. Journal of Experimental Botany, 55(396):285-293.

Friedmann, M., Ralph, S.G., Aeschliman, D., Zhuang, J., Ritland, K., Ellis, B.E., Bohlmann, J. and Douglas, C.J. 2007.Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. Journal of Experimental Botany, 58(3):593-614.

Frost, C.J., Mescher, M.C., Carlson, J.E. and De Morae, C.M. 2008. Plant Defense Priming against Herbivores: Getting Ready for a Different Battle. Plant Physiology, 146: 818–824.

Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current Opinion in Plant Biology, 9:436–442.

Fujimoto, T., Tomitaka, Y., Abe, H., Tsuda,S., Futai,K., and Mizukubo,T. 2011. Expression profile of jasmonic acid-induced genes and the induced resistance against the root-knot nematode *(Meloidogyne incognita*) in tomato plants *(Solanum lycopersicum*) after foliar treatment with methyl jasmonate. Journal of Plant Physiology, 168:1084–1097.

Gan, S.and Amasino, R.M. 1997. Making Sense of Senescence: Molecular Genetic Regulation and Manipulation of Leaf Senescence. Plant Physiology, 113: 313-319.

Gatehouse, J.A. 2002. Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist, 156: 145–169.

Green, T.R. and Ryan, C.A. 1972. Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. Scienc*e*, 175: 776-777.

Glaczinski, H., Heibges, A., Salamini, R. and Gebhardt, C. 2002. Members of the Kunitz-type protease inhibitor gene family of potato inhibit soluble tuber invertase in vitro. Potato Research, 45: 163–176.

Guan, L.M., Zhao, J. and Scandalios, J.G. 2000. Cis-elements and trans-factors that regulate expression of the maize Cat1 antioxidant gene in response to ABA and osmotic stress: H2O2 is the likely intermediary signaling molecule for the response. The Plant Journal, 22(2): 87-95.

Guilfoyle, T.J., Ulmasov, T. and Hagen, G. 1998. The ARF family of transcription factors and their role in plant hormone-responsive transcription. Cellular and Molecular Life Sciences, 54(7):619-27.

Guo, W., Zhao, J., Li, X., Qin, L., Yan, X. and Liao, H. 2011. A soybean b-expansin gene GmEXPB2 intrinsically involved in root system architecture responses to abiotic stresses. The Plant Journal, 66: 541–552.

Guo, M., Thomas, J., Collins, G. and Timmermans, M.C. 2008. Direct repression of *KNOX* loci by the ASYMMETRIC LEAVES1 complex of *Arabidopsis*. The Plant Cell, 20(1):48-58.

Gutierrez-Campos, R., Torres-Acosta, J.A., Saucedo-Arias, L.J. and Gomez-Lim, M.A. 1999. The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. Nature Biotechnology, 17: 1223-1226.

Halitschke, R. and Baldwin, I.T. 2004. Jasmonates and related compounds in plant-insect interactions. Journal of Plant Growth Regulation, 23:238–245.

Hammond, J.P., Broadley, M.R., Bowen, H.C., Spracklen, W.P., Hayden, R.M. and White, P.J. 2011. Gene Expression Changes in Phosphorus Deficient Potato (*Solanum tuberosum* L.) Leaves and the Potential for Diagnostic Gene Expression Markers. PloS One, 6(9):1-15.

Haq, S.K. and Khan, R.H. 2003. Characterization of a proteinase inhibitor from *Cajanus cajan* (L.) Journal of Protein Chemistry, 22:543-554.

Haq, S.K., Atif,S.M. and Khan, RH. 2004. Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection.Archives of Biochemistry and Biophysiscs, 431: 145-159.

Harb, A., Krishnan, A., M.R. Ambavaram, M. and Pereira, A. 2010. Molecular and physiological analysis of drought stress in *Arabidopsis* reveals early responses leading to acclimation in plant growth. Plant Physiology, 154:1254-1271.

Hartl, M., Giri, A.P., Kaur, H.and Baldwin, I.T. 2010.Serine Protease Inhibitors specifically defend *Solanum nigrum* against generalist herbivores but do not influence plant growth and development. The Plant Cell, 22:4158–4175

Heibges, A., Glaczinski, H., Ballvora, A., Salamini, F. and Gebhardt, C. 2003a. Structural diversity and organization of three gene families for Kunitz type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). Molecular Genetics and Genomics, 269: 526–534.

Heibges, A., Salamini, F. and Gebhardt, C. 2003b. Functional comparison of homologous members of three groups of Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). Molecular Genetics and Genomics, 269: 535–541.

Hernandez-Nistal, J.H., Martı´n, I., Jime´nez, T., Dopico, B. and Labrador, E. 2009. Two cell wall Kunitz trypsin inhibitors in chickpea during seed germination and seedling growth. Plant Physiology and Biochemistry, 47:181–187.

Hibbetts K., Hines, B. and Williams, D. 1999. An overview of proteinase inhibitors. Journal of Veterinary Medicine, 13:302–308.

Hilder, V. A., Gatehouse, A. M. R., Sheerman, S. E., Barker, R. F. and Boulter, D. 1987. A novel mechanism of insect resistance engineered into tobacco. Nature, 330:160-163.

Hilder, V.A., A.M.R. Gatehouse, and D. Boulter. 1993. Transgenic plants conferring insect tolerance: protease inhibitor approach. *In* S.D. Kung and R. Wu (eds.), Transgenic Plants, vol. 1, Engineering and Utilization. New York: Academic Press, pp. 317-338.

Hollander-Czytko, H., Andersen, J. K. and Ryan, C.A. 1985.Vacuolar localization of wound-induced carboxypeptidase inhibitor in potato leaves. Plant Physiology, 78:76-79.

Hopkins, A., Matkin, E.A., Ellis, J.A. and Peel, S. 1985. South-west England grassland survey 1983. I. Age structure and sward composition of permanent and arable grassland and their relation to manageability, fertilizer nitrogen and other management factors. Grass and Forage Science, 40:349-359.

Horisberger, M. and Tacchini-Vonlanthen, M. 1983a. Ultrastructural localization of Kunitz inhibitor on thin sections of *Glycine max* (soybean) cv. Maple Arrow by the gold method. Histochemistry, 77: 37-50.

Horisberger, M. and Tacchini-Vonlanthen, M.1983b. Ultrastructural localization of Bowman-Birk inhibitor on thin sections of *Glycine max* (soybean) cv. Maple Arrow by the gold method. Histochemistry, 77: 313-321.

Hunt, J.E., Kelliher, F.M., McSEveny, T.M. and Byres, J.N. 2002. Evaporation and carbon dioxide exchange between atmosphere and tussock grass during summer drought. Agriculture and Forest Meteorology, 543:147-152.

Hunter, D. 1999. Characterisation of ACC Oxidase during leaf maturation and senescence in white clover (*Trifolium repens* L.). Thesis. Massey University.

Hunter, D.A. and Reid, M.S. 2001. A simple and rapid method for isolating high quality RNA from flower petals. Acta Horticulturae, 543: 147-152.

Hutchinson, K.J., King, K.L. and Wilkinson, D.R. 1994. Effects of rainfall, moisture stress and stocking rate on the persistence of white clover over 30 years. Australian Journal of Experimental Agriculture, 35:1039-1047.

Inoue,H., Nojima, H. and Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene, 96(1): 23-28

Ishikawa, A., Ohta, S., Matsuoka, K., Hattori, T., and Nakamura, K. 1994. A family of potato genes that encode Kunitz-type proteinase inhibitors: structural comparisons and differential expression. Plant and Cell Physiology, 35:303–312.

Jeong, M.J. and Shih, M.C. 2003. Interaction of a GATA factor with cis-acting elements involved in light regulation of nuclear genes encoding chloroplast lyceraldehyde-3-phosphate dehydrogenase in Arabidopsis. Biochemical and Biophysical Research Communications, 300:555–562.

Jime´neza, T., Martı´na, I., Herna´ndez-Nistal, J., Labrador, E., and Dopico, B. 2008. The accumulation of a Kunitz trypsin inhibitor from chickpea (TPI-2) located in cell walls is increased in wounded leaves and elongating epicotyls. Physiologia Plantarum, 132: 306–317.

Jin, H. and Martin, C. 1999. Multifunctionality and diversity within the plant MYB-gene family. Plant Molecular Biology, 41(5):577-85.

Jørgensen , J.E., Stougaard , J. and Marcker, K.A. 1991. A Two-Component Nodule-Specific Enhancer in the Soybean N23 Gene Promoter. The Plant Cell, 3: 819-827

Jørgensen, M., Stensballe, A and Welinder, K.G. 2011. Extensive post-translational processing of potato tuber storage proteins and vacuolar targeting. The FEBS Journal, 278:4070–4087.

Jofuku, K. D. and Goldberg, R. B. 1989. Knutiz trypsin inhibitors genes are differentially expressed during the soybean life cycles and in transformed tobacco plants. The Plant Cell, 1:1079-1093.

Johnson, R., Narvaez, J., An, G. and Ryan, C. 1989. Expression of proteinase inhibitor I and II in transgenic tobacco plant: effects on natural defense against *Manduca sexta* larvae. Proceedings of the National Academy of Sciences USA. 86(24): 9871-9875.

Jongsma, M.A. and Bolter, C.J. 1997. The adaptation of insects to plant proteinase inhibitors. Journal of Insect Physiology, 43: 885-896.

Kang, S., Choi, J. and Suh, S. 2001. A leaf-specific 27kda protein of potato kunitz-type proteinase inhibitor is induced in response to abscisic acid, ethylene, methyl jasmonate and water deficit. Molecules and Cells, 13(1):144-147.

Karrer, E.E., Beachy, R.N., and Holt, C.A. 1998. Cloning of tobacco genes that elicit the hypersensitive response. Plant Molecular Biology, 36: 681–690.

Kendall, K. A. 1951. Inhibition of the proteolytic activity of trypsin by green plant extracts. Journal of Dairy Science, 34:499-500.

Kim, J.Y., Park, S.C., Kim, M.H., Lim, H.T., Park, Y. and Hahm, K.S. 2005. Antimicrobial activity studies on a trypsin-chymotrypsin protease inhibitor obtained from potato. Biochem Biophys Res Commun, 330: 921–927

Kim, J., Baek, S. A. and Im K. H. 2009. Overexpression of a Kunitz-type trypsin inhibitor (AtKTI1) causes early flowering in Arabidopsis. Plant Growth Regulator, 59:75–81.

Koiwa, H., Bressan R. A. and Hasegawa, P. M. 1997. Regulation of protease inhibitors and plant defense. Trends in Plant Science, 2(10):379-384.

Kovalchuk, N., Smith, J., Bazanova, N., Pyvovarenko, T., Singh, R., Shirley, N., Ismagul, A., Johnson, A., Milligan, A.S., Hrmova, M., Langridge, P. and Lopato, S. 2012. Characterization of the wheat gene encoding a grain-specific lipid transfer protein TdPR61, and promoter activity in wheat, barley and rice. Journal of Experimental Botany, 63(5):2025-2040.

Kucera, B., Cohn, M.A. and Leubner-Metzger, G .2005. Plant hormone interactions during seed dormancy release and germination. Seed Science Research. 15(04):281–307

Kunitz, M. 1946. Crystalline Soybean Trypsin Inhibitor. The Journal of General Physiology, 29(3): 149–154.

Laluk, K. and Mengiste, T. 2011. The Arabidopsis extracellular UNUSUAL SERINE PROTEASE INHIBITOR functions in resistance to necrotrophic fungi and insect herbivory. The Plant Journal, 68(3): 480-494.

Landgraf, R., Schaarschmid, S. and Hause, B. 2012. Repeated leaf wounding alters the colonization of *Medicago truncatula* roots by beneficial and pathogenic microorganisms. Plant, Cell and Environment, 35(7): 1344-1357.

Laskowski, M. and Kato, I. 1980. Protein inhibitors of proteinases. Annual Review of Biochemistry, 49:593-626.

Lebel,S., Schellenbaum, P., Walter, B. and Maillot, P. 2010. Characterisation of the *Vitis vinifera* PR10 multigene family. *BMC Plant Biology,* 10:184.

Lee, S.I., Lee, S.H., Koo, J.C., Chun, H.J., Lim, C.O., Mun, J.H., Song, Y.H. and Cho, M.J. 1999. Soybean Kunitz trypsin inhibitor (SKTI) confers resistance to the brown plant hopper (*Nilaparvata lugens* Stal) in transgenic rice. Molecular Breeding, 5: 1–9.

Leon, J., Rojo, E. and Sanchez-Serrano, J. 2001. Wound Signalling in plants. Journal of Experimental Botany, 52(354):1-9.

Li, J., Brader, G. and Palva, E. T. 2008. Kunitz Trypsin Inhibitor: An antagonist of cell death triggered by phytopathogens and fumonisin B1 in Arabidopsis. Molecular Plant, 1(3): 482–495.

Li, L., Li, C., Lee, G.I. and Howe, G.A. 2002. Distinct roles for jasmonate synthesis and action in the systemic wound response of tomato. Proceedings of the National Academy of Sciences USA**.** 99(2). 6416-6421.

Liener, I. E. and Kakade, M. L. 1980. Protease inhibitors. In: I. E. Liener, ed. Toxic constituents of plant foodstuffs, 2 ed., New York.

Lievens, S., Goormachtig, S. and Holsters, M. 2004.Nodule-enhanced protease inhibitor gene: emerging patterns of gene expression in nodule development on *Sesbania rostrata*. Journal of Experimental Botany, 55(394):89-97.

Lima, T.B., Silva, O.N., Migliolo, L., Souza-Filho, C., Goncalves, E.G., Vasconcelos, I.M., Oliveira, J.T.A., Amaral, A.C. and France, O.L. 2011. A Kunitz proteinase inhibitor from corms of *Xanthosoma blandum* with bactericidal activity. Journal of Natural Products, 74(5): 969-975.

Linkies, A. and Leubner-Metzger, G. 2012. Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. The Plant Cell Rep,31:253–270.

Lipke, H., Fraenkel, G.S. and Liener, I.E. 1954. Effect of soybean inhibitors on growth of *Tribolium confusum*. Agricultural Food Chemistry, 2: 410–414.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K.and Shinozaki, K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature responsive gene expression, respectively, in *Arabidopsis*. The Plant Cell, 10:1391-1406.

Liu, J., Xia, K., Zhu, J., Deng, Y., Huang, X., Hu, B., Xu, X. and Xu, Z. 2006. The nightshade proteinase inhibitor IIb gene is constitutively expressed in glandular trichomes. Plant Cell and Physiology, 47(9): 1274–1284.

Lorberth, R., Dammann, C., Ebneth, M., Amati, S. and Sanchez-Serrano,J. J. 1992. Promoter elements involved in environmental and developmental control of potato proteinase inhibitor **II** expression. The Plant Journal, 2 (4):477-486.

Luo, M., Wang, Z., Li, H., Xia, K., Cai, Y. and Xu, Z. 2009. Overexpression of a weed (*Solanum americanum*) proteinase inhibitor in transgenic tobacco results in increased glandular trichome density and enhanced resistance to *Helicoverpa armigera* and *Spodoptera litura.* International Journal of Molecular Sciences, 10:1896-1910.

Macedo, M.L., Garcia, V.A., Freire, M.G. and Richardson, M. 2007. Characterization of a Kunitz trypsin inhibitor with a single disulfide bridge from seeds of Inga laurina (SW.) Willd. Phytochemistry, 68: 1104–1111.

Macedo, M.L.R., Freire, M.G.M., Franco, O.L., Migliolo, L. and de Oliveira, C.F.R. 2011.Practical and theoretical characterization of *Inga laurina* Kunitz inhibitor on the control of *Homalinotus coriaceus.* Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 158(2): 164–172.

Major, I.T. and Constabel, C.P. 2008. Functional analysis of the Kunitz Trypsin Inhibitor family in Poplar reveals biochemical diversity and multiplicity in defense against herbivores. Plant Physiology, 146: 888-903.

Manen, J.F., Simon,P., Van Slooten, J.C. Osteras,M., Frutiger, S. and Huhges, G.J. 1991. A nodulin specifically expressed in senescent nodules of winged bean is a protease inhibitor. The Plant Cell, 3:259-270.

Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., Frey, A. and Marion-Poll, A. 1996. Molecular identi®cation of zeaxanthin epoxidase of Nicotiana plumbaginifolia, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. The EMBO Journal, 15: 2331-2342.

McCoy, A.J. and Kortt, A.A. 1997. The 1.8 A crystal structure of winged bean albumin 1, the major albumin from *Psophocarpus tetragonolobus* (L.) DC. Journal of Molecular Biology, 269: 881–891.

Mengel K. and Kirkby, E.A .2001. Principles of Plant Nutrition. Fifth Edition. Dordrecht: Kluwer.

Mello, G.C., Oliva, M.L.V., Sumikawa, J.T., Machado, O.L., Marangoni,T. S., Novello, J.C. and Macedo, M.L.R. 2002. Purification and characterization of a new trypsin inhibitor from *Dimorphandra mollis* seeds. Journal of Protein Chemistry, 20: 625-632.

Monte-Neshich, D.C., Rocha, T.L., Guimarães, R.L., Santana, E.F., Loureiro, M.E., Valle, M and Grossi, de Sá M.F. 1995**.** Characterization and spatial localization of the major globulin families of taro (*Colocasia esculenta* L. Schott) tubers. Plant Science, 112: 149–159.

Moser C, Gatto P, Moser M, PindoM and Velaso R. 2004. Isolation of functional RNA from small amounts of different grape and apple tissues. Molecular Biotechnology, 26 (2): 95-99.

Mosolov, V.V. and Valueva, T.A. 2011. Inhibitors of proteolytic enzymes under abiotic stresses in plants. Applied Biochemistry and Microbiology, 47: 453-459.

Nakai K, **Kanehisa M**. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. Proteins, 11(2):95-110.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kaniya, Y. and Nambara, E. 2005. Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seeds. The Plant Journal, 41:697–709.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2012. NAC transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta 1819:97–103.

Nikmatullah, A. 2009. Regulation of ethylene biosynthesis in vegetative tissues of white clover (*Trifolium repens* L.) during water deficit. Thesis. Massey University.

Norton, G. 1991. Proteinase Inhibitors. *In* J.P.F. D'Mello, C. M. Duffus, and J.H. Duffus (eds.), Toxic Substances in Crop Plants. The Royal Society of Chemistry, pp. 68-106.

Orf, J. H. and Hymowitz, T. 1979. Inheritance of the absence of the Kunitz trypsin inhibitor in seed protein of soybeans. Crop Science*,* 19:107-109.

O'Donnell, P.J., Calvert, C., Atzorn, R., Wasternack, C., Leyser, H.M.O. and Bowles, D.J. 1996. Ethylene as a Signal Mediating the Wound Response of Tomato Plants. Science, 274(5294):1914-7.

Oliva, M.L.V., Ferreira, R.S., Ferreira, J.G., de Paula, C.A.A., Salas,C.S. and Sampaio, M.U. 2011. Structural and functional properties of Kunitz proteinase inhibitors from Leguminosae: A Mini Review. Current Protein and Peptide Science, 12:348-357.

Parcy, F., Nilsson, O., Busch, M.A.,Lee, I. and Weigel, D. 1998. A genetic framework for floral patterning. Nature, 395(6702):561-566.

Park, S.C., Kwon, H.B and Shih, M.C. 1996. Cis-acting elements essential for light regulation of the nuclear gene encoding the a subunit of chloroplast clyceraldehyde 3-phosphate dehydrogenase in *Arabidopsis thaliana*. Plant Physiology, 112:1563-1 571.

Park, D.S., Graham, M.Y., and Graham, T.L. 2001. Identification of soybean elicitation competency factor CF-1 as soybean Kunitz trypsin inhibitor. Physiological and Molecular Plant Pathology, 59:265–273.

Park, Y., Choi, B.H., Kwak, J.S., Kang, C.W., Lim, H.T., Cheong, H.S. and Hahm, K.S. 2005. Kunitz-type serine protease inhibitor from potato (*Solanum tuberosum* L. cv. Jopung). Journal of Agricultural and Food Chemistry, 53: 6491–6496.

Parsons, A. J., Edwards, G. R., Newton, P. C. D., Chapman, D. F., Caradus, J. R., Rasmussen, S. and Rowarth, J. S. 2011. Past lessons and future prospects: plant breeding for yield and persistence in cool-temperate pastures. Grass and Forage Science, 66: 153–172.

Pena-Cortes, H., Sanchez-Serrano, J.J., Mertens, R., Willmitzer, L. and Part, S. 1989. Abscisic acid is involved in the wound-induced expression of the proteinase inhibitor II gene in potato and tomato. Proceedings of the National Academy of Sciences USA, 86(24): 9851-9855.

Pena-Cortes, H., Prat, S., Atzorn, R., Wasternack, C. and Willmitzer, L. 1995. Abscisic acid-deficient plants do not accumulate proteinase nhibitor II following systemin treatment. Planta, 198:447-451.

Pe´ret, B., Cle´ment,M., Nussaume. L and Desnos, T. 2011. Root developmental adaptation to phosphate starvation: better safe than sorry. Trends in Plant Science, 16(8):442-450.

Pereira, K.R.B., Botelho-Júnior, S., Domingues, D.P., Machado, O.L.T., Oliveira, A.E.A, Fernandes, K.V.S., Madureira, H.C., Pereira, T.N.S. and Jacinto, S. 2011. Passion fruit flowers: Kunitz trypsin inhibitors and cystatin differentially accumulate in developing buds and floral tissues. Phytochemistry, 72:1955–1961.

Peterson, M.C., Heigaard, J., Thompson, G.A. and Schulz, A .2005.Cucurbit phloem serpins are graft-transmissible and appear to be resistant to turnover in the sieve element–companion cell complex. Journal of Experimental Botany, 56(422): 3111-3120.

Peterson, T.N., Brunak, S., von Heijne, G. and Nielson, H. 2011. SignalIP 4.0; discriminating signal peptides from transmembrane regions. Nature Methods, 8:785-786.

Pfaffl, M.W. 2001. A new mathematical model model for relative quantification in real-time RT-PCR. Nucleic Acid Research, 29(900): 2002-2007.

Philippe, R.N., Ralph, S.G., Ku¨lheim, C., Jancsik, S.I. and Bohlmann, J. 2009. Poplar defense against insects: genome analysis, full-length cDNA cloning, and transcriptome and protein analysis of the poplar Kunitz-type protease inhibitor family. New Phytologist, 184: 865–884.

Pitzschkea, A., Djameia, A., Teigea, M. and Hirta, H. 2009.VIP1 response elements mediate mitogen-activated protein kinase 3-induced stress gene expression. Proceedings of the National Academy of Sciences USA, 106(43):18414-18419.

Pusztai, A. 1972. Metabolism of trypsin inhibitory proteins in the germinating seeds of kidney bean (*Phaseolus vulgaris*). Planta, 107:21–129.

Qi, Rui-Feng, Song, Zhan-Wu and Chi, Cheng-Wu. 2005. Structural Features and Molecular Evolution of Bowman-Birk Protease Inhibitors and Their Potential Application. Acta Biochimica et Biophysica Sinica, 37(5): 283–292.

Rashed, N.A., McDonald, M.H. and Matthews, B.F. 2008. Protease inhibitor expression in soybean roots exhibiting susceptible and resistant interactions with soybean cyst nematode. Journal of Nematology, 40(2):138–146.

Rhodes, I., Collins, R.P. and Evans, D.R. 1994.Breeding white clover for tolerance to low temperature and grazing stress. Euphytica, 77(3): 239-242.

Richards, K.D., Snowden, K.C., Gardner, R.C.1994. *Wali6* and *wali7*: genes induced by aluminium in wheat (*Triticum aestivum* L.) roots. Plant Physiology. 105: 1455–1456.

Richardson, M. 1977. The protease inhibitors of plants and micro-organisms. Phytochemistry, 16: 159-169.

Richardson, M. 1991. Seed storage proteins: the enzyme inhibitors. In LJ Rogers, ed, Amino Acids, Proteins and Nucleic Acids, Vol 5. Academic Press, Toronto, pp 259–3.

Richardson, M., Campos, F. A. P., Xavier-Filho, J., Macedo, M. L. R., Maia, G. M. C. and Yarwood, A. 1986. The amino acid sequence and reactive (inhibitory) site of the major trypsin inhibitor (DE5) isolate from seeds of the Brazilian carolina tree (*Adenanthera pavonia* L.). Biochemistry et Biophysics Acta, 872:134-140.

Riechmann, J.L. and Meyerowitz, E.M. 1997. MADS domain proteins in plant development. Biological Chemistry, 378(10):1079-1101

Ritonja, A., Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttle, D.J., Barrett, A.J. and Turk, V. 1990. The amino acid sequence of a novel inhibitor of cathepsin D from potato. FEBS Letter, 267: 13–15.

Rizzon, C., Ponger, L. and Gaut, B.S. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Computational Biology*, 2(9):e115.

Roberts, T.H. and Heigaard, J. 2008. Serpins in plants and green algae. Functional and Integrative Genomics, 8(1):1-27.

Robinson, D.A. 1974. A note on the expression of soil moisture content. Area, 6(1):9-13.

Ruijter, J. M., Ramakers, C., Hoogaars, W. M. H., Karlen, Y., Bakker, O., van den Hoff, M. J. B. and Moorman, A. F. M. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Research, 37(6): e45.

Ryan, C.A. 1989. Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores. BioEssays, 10: 20-24.

Ryan, C.A. 1990. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annual Review of Phytopathology, 28: 425–449.

Ryan, C.A. 2000. The systemin signaling patway: differential activation of plant defense genes. Biochimica et Biophysica Acta, 1477(1): 112-121.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Vol I. Spring Harbour Laboratory Press, New York.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2006. Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. The Plant Cell, 18: 1292–1309.

Schirra, H.J., Guarino, R.F., Anderson, M.A. and Craik, D.J. 2009. Selective removal of individual disulfide bonds within a potato type ii serine proteinase inhibitor from *Nicotiana alata* reveals differential stabilization of the reactive-site loop. Journal of Molecular Biology, 395(2): 609-626.

Schwartz, S.H., Tan, B.C., Gage, D.A., Zeevaart, J.A. and McCarty, D.R. 1997. Specific oxidative cleavage of carotenoids by VP14 of maize. Science, 276:1872-1874.

Seo, M., Koiwai, H., Akaba, S., Komano, T., Oritani, T., Kamiya, Y. and Koshiba, T. 2000. Abscisic aldehyde oxidase in leaves of *Arabidopsis thaliana*. The Plant Journal, 23:481-488.

Shain, Y. and Mayer, A.M. 1965. Proteolytic enzymes and endogenous trypsin inhibitor in germinating lettuce seeds. Physiologia Plantarum, 18: 853-859.

Sharpley,N., Daniel,T. and Qright,R. 1999. Functions of phosphorus in lants, Better Crops, 83: 6-7.

Shee, C. and Sharma, A.K. 2008. Storage and affinity properties of *Murraya koenigii* trypsin inhibitor. Food Chemistry, 107(1):312-319.

Shewry, P.R., Napier, J.A. and Tatham, A.S. 1995. Seed Storage Proteins: Structures 'and Biosynthesis. The Plant Cell, 7:945-956.

Shi, Y. 2002. Mechanisms of caspase activation and inhibition during apoptosis. Molecular Cell, 9: 459–470.

Shinozaki, K. and Yamaguchi-Shinozaki, K. 1997. Gene expression and signal transduction in water-stress response. Plant Physiology, 115: 327–334.

Singh, K., Foley, R.C. and Onate-Sanchez, L. 2002.Transcription factors in plant defense and stress responses. Current Opinion in Plant Biology, 5(5): 430-436.

Sin, S.F. and Chye, M.L. 2004. Expression of proteinase inhibitor II proteins during floral development in *Solanum americanum*. Planta, 219: 1010-1022.

Solano, R., Stepanova, A., Chao, Q. and Ecker, J.R. 1998. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes Develeopment, 12(23):3703-14.

Soler, R., Erb, M. and Kaplan, I. Long distance root–shoot signalling in plant–insect community interactions. Trends in Plant science, 16 September, 2012.

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

Smith, S. and Smet, I.D. 2012. Root system architecture: insights from Arabidopsis and cereal crops. Philosophical Transactions of the Royal Socociety, B, 367: 1441–1452.

Snowden, K.C., Richards, K.D., and Gardner, R.C. 1995. Aluminum induced genes. Plant Physiology, 107(2): 341–348.

Spencer, M.E. and Hodge, R. 1991. Cloning and sequencing of the cDNA encoding the major albumin of *Theobroma cacao*. Planta, 183:528–535.

Speransky, A.S., Cimaglia, F., Krinitsina, A.A., Poltronieri, P., Fasano, P., Bogacheva, A.M., Valueva, T.A., Halterman, D., Shevelev, A.B. and Santino, A. 2007. Kunitz-type protease inhibitors group B from *Solanum palustre*. Biotechnol Journal. Epub ahead of print, PMID: 17639528.

Srinivasan, T., Kumar, K.R.R. and Kirti, P.B. 2009. Constitutive expression of a trypsin protease inhibitor confers multiple stress tolerance in transgenic tobacco. Plant and Cell Physiology, 50(3):541-553.

Stiekema, W.J., Heidekamp, F., Dirkse, W.G., Beckum, J.V., Haan, P.D., Bosch, C.T. and Louserse, J.D. 1988. Molecular cloning and analysis of four potato tuber mRNAs. Plant Molecular Biology, 11:255–269.

Stolf-Moreira, R., Medri, M.E., Neumaier, N., Lemos, N.G., Pimenta, J.A. , Tobita, S., Brogin, R.L. , Marcelino-Guimarães, F.C., Oliveira, M.C.N., Farias, J.R.B. , Abdelnoor, R.V. and Nepomuceno, A.L.2010**.** Soybean physiology and gene expression during drought. Genetics and Molecular Research 9, (4): 1946-1956.

Strickland, J.A, Orr, G.L, Walsh, T.A. 1995. Inhibition of larval growth by patatin, the lipid acylhydrolase from potato tubers. Plant Physiology, 109:667–674.

Tan-Wilson, A.L., Hartl, P.M., Delfel, N.E. and Wilson, K.A. 1985. Differential expression of kunitz and bowman-birk soybean proteinase inhibitors in plant and callus tissue. Plant Physiology, 78: 310-314.

Terada, S., Fujimura, S., Katayama, H., Nagasawa, M. and Kimoto, E. 1994. Purification and characterization of two Kunitz family subtilisin inhibitors from seeds of *Canavalia lineata*. The Journal of Biochemistry, 115: 392–396.

Thibaud, Marie-Christine, Arrighi Jean-Franc¸ois , Bayle1, V., Chiarenza, S., Creff, A., Bustos, R., Paz-Ares, J., Poirier, Y. and Nussaume, L. 2010. Dissection of local and systemic transcriptional responses to phosphate starvation in *Arabidopsis*. The Plant Journal, 64:775–789.

Todd, A.E., Orengo, C.A. and Thornton, J.M. 2001. Evolution of function in protein superfamilies, from a structural perspective. Journal of Molecular Biology, 307:1113-1143.

Turra`a, D., Bellin, D., Lorito, M. and Gebhardt, C. 2009. Genotype-dependent expression of specific members of potato protease inhibitor gene families in different tissues and in response to wounding and nematode infection. Journal of Plant Physiology, 166:762-774.

Urwin, P.E., H.J. Atkinson, D.A. Waller, and M.J. McPherson. 1995. Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida.* The Plant Journal, 8: 121-131.

van der Biezen, E.A., Brandwagt, B.F., van Leeuwen, W., Nijkamp, H.J.J. and Hille, J. 1996. Identification and isolation of the FEEBLY gene from tomato by transposon tagging. Molecular genetics and Genomics, 251:267–280.

Valueva, T.A., Revina, T.A., Mosolov, V.V. and Mentele, R. 2000. Primary structure of potato Kunitz-type serine proteinase inhibitor. Biological Chemistry, 381:1215–1221.

Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J., Rycke, R.D., Brackenier, A., Inze D., Harris, J.L. and Breusegem, F.V. 2006. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for Metacaspase 9. Journal of Molecular Biology, 364(4):625-636.

von Dahl, C.C. and Baldwin, I.T. 2007. deciphering the role of ethylene in plant–herbivore interactions. Journal Plant Growth Regulator, 26:201–209.

Vodkin, L. O. 1981. Isolation and characterization of messenger RNAs for seed lectin and Kunitz inhibitor in soybeans. Plant Physiology, 68:766-771.

Voisey, C.R., White, D.W.R., Dudas, B., Appleby, R.D., Ealing, P.M. and Scott, A.G. 1994. *Agrobacterium* mediated transformation of white clover using direct shoot organogenesis. Plant Cell Reports, 13: 309–314.

Watson, R.N., Dickie, G., Neville, F.J. and Bell, N.L.1996. Seedling growth of nine white clover cultivars after invasion by clover nematodes. Proceedings of the NZ Plant Protection Conference. 239-243.

Welham, T., O'Neill, M., Johnson, S., Wang, T.L. and Domoney, C. 1998. Expression patterns of genes encoding seed trypsin inhibitors in *Pisum sativum*. Plant Science, 131:13–24.

Woltering, E.J., van der Bent, A., and Hoeberichts, F.A. 2002. Do plant caspases exist? Plant Physiology, 130: 1764–1769.

Wu, P., Ma,L., Hou, X., Wang, M., Wu, Y., Liu, F. and Deng, X.W. 2003. Phosphate starvation triggers distinct alterations of genome expression in Arabidopsis roots and leaves. Plant Physiology, 132:1260-1271.

Xavier-Filho, J. 1992. The biological roles of serine and cysteine proteinase inhibitors in plants. Revista Brasileira de Fisiologia Vegetal, 4(1): 1-6.

Xavier-Filho, J. and Campos, F.A.P. 1989. Proteinase inhibitors. In: Cheeke, P. R. (ed.). Toxicants of Plant Origin. Boca Raton, CCR Press, 3:1-27.

Xu, Z.F., Qi,W.Q., Ouyang, X.Z., Yeung, E. and Chye, M.L. 2001. A proteinase inhibitor II of *Solanum americanum* is expressed in phloem. Plant Molecular Biology, 47: 727-738.

Yeu, S. Y., Park, B. S., Sang, W.G., Choi, Y. D., Kim, M.C., Song, J. T., Paek, N., Koh, H. and Seo, H. S. 2007. The serine proteinase inhibitor OsSerpin is a potent tillering regulator in rice. Journal of Plant Biology, 50(5):600-604.

Yousa, T., Ichimura, K., Mizoguchi, T. and Shinozaki, K. 2001. Oxidative Stress Activates ATMPK6, an *Arabidopsis* Homologue of MAP Kinase. Plant and Cell Physiology, 42(9):1012-1016.

Zeevaart, J. A.D. 1998. Metabolism and physiology of abscisic acid. Annual Review of Plant Physiology, 39:439-473.

Zhang. X., Liu, S. and Takano, T. 2008. Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. Plant Molecular Biology, 68:131–143.

Appendices

Appendix 1: Families of proteinase inhibitors from the angiosperms listed in the MEROPS database

FAMILY	SUBFAMILY	Clan	TYPE INHIBITOR
I3	I3A	IC	soybean Kunitz trypsin inhibitor (*Glycine max*)
I3	I3B	IC	proteinase inhibitor B (*Sagittaria sagittifolia*)
I6		IJ	ragi seed trypsin/alpha-amylase inhibitor (*Eleusine coracana*)
I7		IE	trypsin inhibitor MCTI-1 (*Momordica charantia*)
I12		IF	Bowman-Birk inhibitor unit 1 (*Glycine max*)
I18		JD	mustard trypsin inhibitor-2 (*Sinapis alba*)
I20		JO	potato peptidase inhibitor II inhibitor unit 1 (*S. tuberosum*)
I37		IE	potato metallocarboxypeptidase inhibitor (*S. tuberosum*)
I55		unassigned	squash aspartic peptidase inhibitor (*Cucumis sativus*)
I67		IF	bromein (*Ananas comosus*)
I73		JN	{Veronica} trypsin inhibitor (*Veronica hederifolia*)
I90		unassigned	trypsin inhibitor (*Mirabilis jalapa*)

Appendix 2: Sequences from the AgResearch white clover EST database that gave hits against the *M. truncatula KPI* genes, as indicated

Group A: Seven sequences showed 80% identity with *M. truncatula KPI* gene

>lcl|PGC.CTR0034079347-cF2_20040416 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.CTR0018077912-cF2_20030818 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.FTRC101582P22-c1PHRAP_20040720 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.CTR0019077959-cF2_20031229 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.CTR0033079234-cF2_20040328 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.CTR0036080049-cF2_20040726 11 to 769: Frame 2 253 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYH
VIFVDAASHEADQYIKSVV

>lcl|PGC.FTRC101582P22-b0FSP_20030715 11 to 682: Frame 2 224 aa

KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFVFITNLSLAFSNEDVEQVLDINGNAIFPG
GEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKW
LVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKR

Clustalw alignment of the seven sequences of *M. truncatula* comprising group A

```
lcl|PGC.CTR0034079347-cF2_2004      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.CTR0018077912-cF2_2003      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.FTRC101582P22-c1PHRAP_      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.CTR0019077959-cF2_2003      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.CTR0033079234-cF2_2004      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.FTRC101582P22-b0FSP_20      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
lcl|PGC.CTR0036080049-cF2_2004      KSSLIFPYKVTIIIYRSRCNSTKHKQISKIKSYLTMKHVSSLTLCFLLFV 50
                                    **************************************************

lcl|PGC.CTR0034079347-cF2_2004      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.CTR0018077912-cF2_2003      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.FTRC101582P22-c1PHRAP_      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.CTR0019077959-cF2_2003      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.CTR0033079234-cF2_2004      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.FTRC101582P22-b0FSP_20      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
lcl|PGC.CTR0036080049-cF2_2004      FITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKT 100
                                    **************************************************

lcl|PGC.CTR0034079347-cF2_2004      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.CTR0018077912-cF2_2003      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.FTRC101582P22-c1PHRAP_      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.CTR0019077959-cF2_2003      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.CTR0033079234-cF2_2004      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.FTRC101582P22-b0FSP_20      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
lcl|PGC.CTR0036080049-cF2_2004      GDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNC 150
                                    **************************************************

lcl|PGC.CTR0034079347-cF2_2004      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.CTR0018077912-cF2_2003      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.FTRC101582P22-c1PHRAP_      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.CTR0019077959-cF2_2003      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.CTR0033079234-cF2_2004      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.FTRC101582P22-b0FSP_20      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
lcl|PGC.CTR0036080049-cF2_2004      AKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKL 200
                                    **************************************************

lcl|PGC.CTR0034079347-cF2_2004      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
lcl|PGC.CTR0018077912-cF2_2003      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
lcl|PGC.FTRC101582P22-c1PHRAP_      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
lcl|PGC.CTR0019077959-cF2_2003      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
lcl|PGC.CTR0033079234-cF2_2004      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
lcl|PGC.FTRC101582P22-b0FSP_20      GFCIKGSPTCLDIGRYDNDEAGKR------------------------- 224
lcl|PGC.CTR0036080049-cF2_2004      GFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIK 250
                                    ************************

lcl|PGC.CTR0034079347-cF2_2004      SVV 253
lcl|PGC.CTR0018077912-cF2_2003      SVV 253
lcl|PGC.FTRC101582P22-c1PHRAP_      SVV 253
lcl|PGC.CTR0019077959-cF2_2003      SVV 253
lcl|PGC.CTR0033079234-cF2_2004      SVV 253
lcl|PGC.FTRC101582P22-b0FSP_20      ---
lcl|PGC.CTR0036080049-cF2_2004      SVV 253
```

Group B: Five sequences having 50% or less identity with *M. truncatula KPI* gene

>lcl|PGC.CTR0016005245-cE3_20030701 1 to 348: Frame 1 116 aa

TPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGR
YDNDEAGKRLNLTEHESYHVIFVDAASHEADQYTKSVV

>lcl|PGC.CTR0021001199-cE4_20040101 380 to 33: Frame -1 116 aa

TPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGR
YDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIKSVV

>lcl|PGC.FTRC101582P22-g0RSP_20030715 657 to 316: Frame -1 114 aa

VEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYD
NDEAGKRLNLTEHESYHVIFVDAASHEADQYIKSVV

>lcl|PGC.ETRS55OX05D05-g1M13RE_20030605 1 to 348: Frame 1 116 aa

TPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPSVQTLSGTFNIHKHESGFGYKLGFRIKGSPTCLDIGR
YDNDEAGKRLNLTEHESYHVIFVDAASHEADQYIKSVV

>lcl|PGC.ETRS55OX13D11-g1M13RE_20030625 335 to 18: Frame -1 106 aa

PNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRL
NLTEHESYHVIFVDAASHEADQYTKSVV

Clustalw alignment of the five sequences of *M. truncatula* comprising group B

```
lcl|PGC.CTR0016005245-cE3_2003      TPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTF 50
lcl|PGC.ETRS55OX13D11-g1M13RE_      ----------PNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTF 40
lcl|PGC.CTR0021001199-cE4_2004      TPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTF 50
lcl|PGC.FTRC101582P22-g0RSP_20      --VEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTF 48
lcl|PGC.ETRS55OX05D05-g1M13RE_      TPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPSVQTLSGTF 50
                                              ***************:********.*****.********

lcl|PGC.CTR0016005245-cE3_2003      NIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIF 100
lcl|PGC.ETRS55OX13D11-g1M13RE_      NIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIF 90
lcl|PGC.CTR0021001199-cE4_2004      NIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIF 100
lcl|PGC.FTRC101582P22-g0RSP_20      NIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIF 98
lcl|PGC.ETRS55OX05D05-g1M13RE_      NIHKHESGFGYKLGFRIKGSPTCLDIGRYDNDEAGKRLNLTEHESYHVIF 100
                                    ************** ********************************

lcl|PGC.CTR0016005245-cE3_2003      VDAASHEADQYTKSVV 116
lcl|PGC.ETRS55OX13D11-g1M13RE_      VDAASHEADQYTKSVV 106
lcl|PGC.CTR0021001199-cE4_2004      VDAASHEADQYIKSVV 116
lcl|PGC.FTRC101582P22-g0RSP_20      VDAASHEADQYIKSVV 114
lcl|PGC.ETRS55OX05D05-g1M13RE_      VDAASHEADQYIKSVV 116
                                    ********** ****
```

Group C: Partial 15 sequences with less than 50% identities with *M. truncatula KPI* sequence

>00112813WCB728B2E3 2 to 586: Frame 2 195 aa
EXCLXXPFXLFFVSXPIFPSILNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREV
KNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTFNIHKHES
GFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHES

>00112113WCEC26A1P5 3 to 584: Frame 3 194 aa
KXVXXXLXFSSLFPHQSFLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREV
KNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTFNIHKHES
GFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLNLTEHES

>01020709WCD81D1AX3 15 to 605: Frame 3 197 aa
LXNLXXCFNTXPFPXLLFVSITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGRVRIGKTGDLKCPVTV
LQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLSGTF
NIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKTFEFD

>00111013WCE62321C2 3 to 590: Frame 3 196 aa
QIXRXILSNXKXFILPFXSLLCFLTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKTGDLKCP
VTVLQDRREVKNGIPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKACVGIGGPENYPGVQTLS
GTFNIHKHESGFGYKLGFCIKGSPTCLDIGRYDNDEAGKR

>00112413WCF32831L1 34 to 600: Frame 1 189 aa
EAVYXXPLLPPLCFLTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPELRGPGDGGVRIGKTGDLKCPVTVLQDRSE
VKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHE
SGFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLN

>00100607WCG31672E1 35 to 598: Frame 2 188 aa
SCXYSPLLPPLCFLTNLSLXFSNEDVEQVLDINGNAIFPGGEYYILPELRGPGDGGVRIGKTGDLKCPVTVLQDRSEV
KNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHES
GFGYKLGFCIKGSPTCLDIGRYDNDEAGKRLN

>99090503WCAB0E1AXX 25 to 651: Frame 1 209 aa
YFLTIIIYRSRCNPPKHKQIIKIKSYLTMKHVSSLTLSILFFVSITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILP
ALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPIEIEFFKKPNCAKSSKWLVFVDNV
VKKACVGIGXPENYPGVQTLSGTFNIHKHESGFGYXLGFCIKXSPYLVGYWEX

>00100508WCCA1631T4 37 to 444: Frame 1 136 aa
SMFILPFPSLLCFLTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKTGDLKCPVTVLQDRREV
KNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKSLCWYWWS

>00110807WCB52273L2 30 to 584: Frame 3 185 aa
CLXXPLLPPLCFXTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPELRGPGDGGVRIGKTGDLKCPVTVLQDRSEVK
NGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESG
FGYKLGFCIKGSPTCLDIGRYDNDEAGKR

>00100607WCH81682B4 31 to 576: Frame 1 182 aa
CXSSPLLPPLCFXTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPELRGPGDGGVRIGKTGDLKCPVTVLQDRSEVK
NGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHESG
FGYKLGFCIKGSPTCLDIGRYDNDEA

>00102008WCA41BE3C1 7 to 444: Frame 1 146 aa
NLXDXSYLTVKXFXLPFXSLLCFLTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPALRGPGGGGVRIGKTGDLKCP
VTVLQDRREVKNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVVKKSLCWYWWS

>00092207WCH51211I2 35 to 556: Frame 2 174 aa
SCLYSPLLPPLCFLTNLSLAFSNEDVEQVLDINGNAIFPGGEYYILPELRGPGDGGVRIGKTGDLKCPVTVLQDRSEV
KNGLPVKFTIPDISTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNVIKKACVGIGSPENYPGVQTLSGTFNIHKHES
GFGYKLGFCIKGSPTCLD

>00112413WCB42841O1 36 to 602: Frame 3 189 aa

YFLTIIIYRSRCNPPKHKQIIKIKSYLTMKHVSSLTLSILFFVSITNLSQAFSNEDVEQVLDINGNAIFPGGEYYILP
ALRGPGGGGVRIGKTGDLKCPVTVLQDRREVKNGLPVKFTIPDIGTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNV
VKKACVGIGGPENYPGVQTLSGTFNIHKHESGF

>00092607WCGB1361N5 41 to 574: Frame 2 178 aa

YFLTIIIYRSRCNPPKHKQISKIKSYLTMKHVSSLTLSILFFVSITNLSLAFSNEDVEQVLDINGNAIFPGGEYYILP
ALRGPGGGGVRIGKTGDLKCPVTILQDRREVKNGLPVKFTIPDVSTGIIFTGTPVEIEFFKKPNCAKSSKWLVFVDNV
VKKACVGIGGPENYPGVQTLSG

>00112213WCF22711TF 36 to 596: Frame 3 187 aa

SVLPFCPQNTYKKHYLTMKPVLSLTFSFFLIVLITNISLALSYEAVEQVLDSLGNPIFPGGKYYIFPVSHDDTYGGGL
RLAKTGDSKCEVTALQDDNIVIDNIPVKFSIPGISPGIIFTGTPIEIEFTKKPSFVESSKWLIFVDDVIQKACVGIGG
PENYPHFKTLNGRFYIEKHESGFGYKLXYCV

Appendix 3: Sequence alignments for designing four different sets of degenerate primers

Group 1: **F:** GGAYAYAARTGGYAACCCCMTTWTYCC
R: CCARCRTRAYYTTCAGGACCACC

```
Medicago        ATGAAGCATTTTTTATCACTAACCCTTTCCTTCTTCATCTTTGTTTTCATCACCAATCTT 60
Soybean         ATGAAGCCTACCCTACTATTATCCCTTTCCTTCCT-GCCTCTCTTTGC-----------T 48
CaTPI1          ATGAAATCCATTGTATTCCTCCTATTTGCAATTCTAACCAATCCTTTA-----------T 49
                *****        **    *     ** *  *  *   *  *  **            *

Medicago        TCACTAGCTACCTCAA--ATGATGTTGAGCAAGTATTGGACATAAATGGTAACCCCATTT 118
Soybean         TTCCTGGCTCTTTCAG--AAGATGTTGAACAAGTTGTGGACATAAGTGGCAACCCCATTT 106
CaTPI1          TTGCATTCTCCAATAACAATGCTATTGAGCAAGTGTTGGATACAAATGGTAACCCCCTTA 109
                *  *   **     *   * * * **** *****  **** * ** *** ****** **

Medicago        TTCCAGGTGGTCAATACTACATTTTACCAGCACTTCGTGGCCCCGGAGGAGGAGGAGTAA 178
Soybean         TCCCAGGTGGCACATATTACATTATGCCATCAACTTGGGGCGCTGCCGGTGGTGGATTGA 166
CaTPI1          TCCCTGGTGATGAATACTACATTTTTCCAGCAAGTGATAACCCTAAAACTGGAGGACTAA 169
                * ** ****    *** ****** * *** **  *     * *       ** *** * *

Medicago        GATTAGGAAGAACCGGTGATTTAAAGTGTCCGGTTACCGTCCTACAAGATCGTAGAGAAG 238
Soybean         AACTAGGCCGGACAGGAAACTCAAACTGCCCAGTTACTGTTTTGCAAGATTACTCAGAAA 226
CaTPI1          CCCTAAACAAAATTAGTGATGCAGAGTGTCCTGTAACTGTCCTACAAAATAATGCGACAA 229
                   **      *   *  *   * * ** ** ** ** **  * *** **        *

Medicago        TAAAAAATGGTCTACCAGTGAAATTCACCATACCAGGAA----TAAGTCCTGGTAT--AA 292
Soybean         TCTTCCGTGGCACACCAGTCAAATTCAGCATACCTGGGA----TAAGCCCTGGAAT--CA 280
CaTPI1          ------GAGGTTTACCAGTTAAATTCACCTTATCAGGAAGTAATAACACTGGTAATAATA 283
                      **   ****** ******* * ** * ** *    ***  *  *  **   *

Medicago        TTTTCACTGGTACACCACTTGAGATCGAGTACACGAAAAAACCTAGTTGTGCTGCATCAA 352
Soybean         TCTTTACAGGTACTCCACTTGAAATCGAGTTCGCAGAGAAACCTTATTGTGCTGAATCCT 340
CaTPI1          TCTTGACTAATACTGATCTTGAAATTGAGTTCACTAAGAAGCCAAATTGCGTTGAATCAT 343
                * ** **   ***    ***** ** **** * *  * ** **   *** * ** ***

Medicago        CAAAATGGTTAATATTTGTTGATAATGTTATTGGAAAAGCTTGTATTGGTATTGGTGGTC 412
Soybean         CCAAATGGGTGGCGTTTGTGGACAATGAAATCCAAAAGGCATGTGTGGGTATTGGTGGTC 400
CaTPI1          CAAAATGGATTGTGTTTGTTGATGATTTTACTCCTCAAGGTTGTGTTGGTATTGGTGGTC 403
                * ****** *    ***** **  **   *      * *  *** * **************

Medicago        CTGAAAATTACCCTGGTGTGCAAACATTGAAGGGAAAATTTAATATTCAGAAACATGCAT 472
Soybean         CTGAAGGTCATCCTGGTCAACAAACATTTAGTGGCACATTTAGCATTCAGAAATATAAAT 460
CaTPI1          CTGAAAATCATCTTGGTTTAGAAATACTCAATGGCAAATTTTTAATTGTGAGACATGCTT 463
                *****  * * * ****    *** * * *  ** * ****   ***  ** * **   *

Medicago        CTGGATTTGGTTATAACTTAGGGTTTTGTGTTACTGGATCTCCTACTTGTTTGGATATTG 532
Soybean         TTGGAT------ACAAACTTGTGTTCTGTATCACTGGCTCAGGCACTTGTTTAGATATTG 514
CaTPI1          CTGGATATGTTTATAGGTTTGGATTTTGTTTGGATGTGAGTGGTGATTGTGGTTTGCTTG 523
                 *****      * *   * *  ** *** *   **          ****      ***

Medicago        GAAGATTTGAT------AATGATGAAGCTGGAAGACGTTTGAATTTGACTGAACATGAGG 586
Soybean         GAAGGTTTGATGCCAAAAATGGTGAGGGAGGAAGACGTTTGAATCTCACTGAGCATGAGG 574
CaTPI1          GATTGAATACTTTTGATTCGAGAGAAGGTGGATCACGTTTAATTTTAACTATATTTAATT 583
                **     *  *            ** *  ***  ****** * * * ***     * *

Medicago        TTTATCAAGTTGTGTTTGTTGATGCTGCTACTTATGAAGCTGAGTATATTAAATCTGT-- 644
Soybean         CCTTCGACATTGTTTTCATAGAAGCTTCTAAGGTTGA---TGGAATTATCAAGTCCGT-- 629
CaTPI1          CTTATAATGTTGTATTTGTTGATGTTGCTTCTGTTAAATCTGGACGTATTATGCCTCTTA 643
                  *   *  **** **  * ** * * **     * *   **    *** *   *  *
```

Group 2: F: GTA KTR GAC AWA MAT GGY AHC CCC
R: GAT ART TTK YSG WAC CAC CAA TAC CAA

```
IMGA|Medtr6g070540.1      ATGAAGCATCTTTTATCACTAACC---------CTTTCCTTCTTCATCTTTGTTTTCATC 51
IMGA|Medtr6g070580.1      ATGAAACATGTTTTATCACTCACC---------CTTTCCTTCCTCCTGTTTGTTTTCACT 51
IMGA|Medtr6g070470.1      ATGAATTCTGTTTGTTCACTTACCATCCTTTCCCTTTCCTTCCTCCTCTTTGTTTTCATC 60
IMGA|Medtr6g010380.1      ATGAATCCCATTTTATCCCTTAGC------CTTTCCTTCTTCCTAGTATTTGTTTTTATC 54
                          *****     ***  ** ** * *           * **** *  * ******** *

IMGA|Medtr6g070540.1      ACCAATCTTTCACTAGCTACCTCAAATGATG---TTGAGCAAGTATTAGACATAAATGGT 108
IMGA|Medtr6g070580.1      ACCAATCTTTCACTAGCTTTCTCAAATGATGCTGTTGAACAAGTATTAGACATAAATGGT 111
IMGA|Medtr6g070470.1      ACAAATCTTTCACTAGTTTTCTCAAATGACAATGTTGAGATAGTAGTGGACAAAAATGGC 120
IMGA|Medtr6g010380.1      ACCAATCTTTCAC---------CAAACAATGCTGCTAAACAAGTATTAGACATACATGGT 105
                          ** **********         ****  *      * *   **** * **** * ****

IMGA|Medtr6g070540.1      AACCCCATTTTTCCAGGTGGTCAATACTACATTTTACCAGCACTTCGTGGCCCCGGAGGA 168
IMGA|Medtr6g070580.1      AACCCCATTTTTCCAGGTGGCAAATACTACATTTTACCAGCAATTCGAGGACCCCTTGGT 171
IMGA|Medtr6g070470.1      ATCCCCCTTATCCCAGGTACCAGTTATTACATTTCGCCAGCAAAT-----ACC----GGT 171
IMGA|Medtr6g010380.1      ACCCCCCTTATCCCTGGTAGCCAATACTACATTTTTCCAGCTAGTGAAAACCCTAACAGT 165
                          * **** ** * ** ***      ** *******  *****   *      **     *

IMGA|Medtr6g070540.1      GGAGGAGTAAGATTAGGAAGAACCGGTGATTTAAAGTGTCCGGTTACCGTCCTACAAGAT 228
IMGA|Medtr6g070580.1      GGAGGACTAAGACTGGGAAAATCAAGTAATTCAGATTGTGAAGTTACTGTTGTACAAGAT 231
IMGA|Medtr6g070470.1      GGACGAATTACACTAGGTAAAACTGTTGATTCAGATTGTTCTTTTCTTGTGTTACAGGAT 231
IMGA|Medtr6g010380.1      GGAGGACTAACCCTAAACAAAGTTGGTAATTTAGAGTGTCCAGTAACTGTCCTACAAAAT 225
                          *** ** * *   *    * *     * *** * * ***    *    **  ****  **

IMGA|Medtr6g070540.1      CGTAGAGAAGTCAAAAATGGTCTACCAGTGAAATTCACCATACCAG---GAATAAGTCC- 284
IMGA|Medtr6g070580.1      TACAATGAAGTTATCAATGGTGTACCAGTGAAATTCAGTATACCAG---AAATAAGCCC- 287
IMGA|Medtr6g070470.1      GATGAAAAAATGATTTACGGTCGACAGGTGAAATTCAGCTTATCCGTCGGAATAATCCCG 291
IMGA|Medtr6g010380.1      AAT------GCTATGATAGGTTTACCAGTTAAATTCACCATCCCAG---AAAACAGCAC- 275
                                *     ***  **  ** *******   *  * *    **  *   *

IMGA|Medtr6g070540.1      --TGGTATAATTTTCACTGGTACACCACTTGAGATCGAGTACACGAAAAAACCTAGTTGT 342
IMGA|Medtr6g070580.1      --TGGTATAATCTTCACTGGTACACCAATTGATATCGAGTTCACAAAGAAGCCTAATTGC 345
IMGA|Medtr6g070470.1      GCTAGTTTAATCTTCACGAATACCGCGCTAGACATTGAGTTCGTATATAAGGATAGTTGT 351
IMGA|Medtr6g010380.1      --TGGTAATATCTTGACCGGTACCGATCTTGAGATCGAGTTCACTAAAAAGGCTGATTGT 333
                            * **   ** ** **   ***     * ** ** **** *    * **   *  ***

IMGA|Medtr6g070540.1      GCTGCATCAACAAAATGGTTAATATTTGTTGATAATGTTATTGGAAA---AGCTTGTATT 399
IMGA|Medtr6g070580.1      GTTGAATCATCAAAATGGTTGATATTTGTTGATAGTGTTATTCAAAA---AGCTTGTGTT 402
IMGA|Medtr6g070470.1      GTTGAATCATCAAAATGGTTGATATTTGTTGATAATGTTAATAACAACAAATCATTTGTT 411
IMGA|Medtr6g010380.1      GCTGAAGCATCAAAATGGTTGATGTTTGTTGATCATAATACTCAACT---AAGTTGTGTT 390
                          * ** * ** ********** ** *********  *  ** *        *   * * **

IMGA|Medtr6g070540.1      GGTATTGGTGGTCCTGAAAATTATCCT---GGTGTGCAAACATTGAAGGGAAAATTTAAT 456
IMGA|Medtr6g070580.1      GGTATTGGTGGTCCTGAAAATTATCCT---GGTTTTAGAACATTGAGTGGCACATTTAAT 459
IMGA|Medtr6g070470.1      GGTATTGGTGGTCCTGAAAACTATCCTCAAGGTACACAAATATTGAATGGAAAATTTAAT 471
IMGA|Medtr6g010380.1      GGTATTGGTGGTGCAACAAACTATCAT---GGTATAGAAACAATAAGTGGCAAATTCTTA 447
                          ************ *   *** **** *   ***     ** * * *  ** * ***

IMGA|Medtr6g070540.1      ATTCAGAAACATGCATCTGGATTTGGTTATAACTTAGGGTTTTGTGTTACTGGATCTCCT 516
IMGA|Medtr6g070580.1      ATTGAGAAGCATGAATCTGGATTTGGTTATAGGCTTGGATATTGTGTGAAAGATTCTCCT 519
IMGA|Medtr6g070470.1      ATTAAGAAATCTGGATCTGAAAATGCTTATAAGTTTGGATTTTGTGTTAAGGAAACTCCT 531
IMGA|Medtr6g010380.1      ATAGTGAAACATGGATCTGGTCATGTTTATAGGCTTGGATTTTGTTTGGATGTAACTGGA 507
                          **   ***   ** *****    ** *****   * ** * **** *    *   **

IMGA|Medtr6g070540.1      ACTTGTTTGGATATTGGAAGATTTGATAATGA---------TGAAGCTGGAAGACGTTTG 567
IMGA|Medtr6g070580.1      ACTTGTTTGGATATTGGGAGGGCTCATGAAGAAGTTGAAGATGAAGGTGGATCACGTTTA 579
IMGA|Medtr6g070470.1      AGTTGTTGGGATATTGGGAGGTAT-ATGAGTA--TAGGTGAAGAAGGTGGAAGACGTTTG 588
IMGA|Medtr6g010380.1      AATTGTGGCTATATTGG-ACTTCAAATGTTCAATTC--TGAAGAAGGTGGATCACGTTTG 564
                          * ****    ******* *      **    *          **** ****  ******

IMGA|Medtr6g070540.1      AATTTGACTGAACATGAGGTTTATCAAGTTGTGTTTGTTGATGCTGCTACTTATGAAGCT 627
IMGA|Medtr6g070580.1      CATTTGACTCATCAAGTGGCTTTTGCAGTTGTGTTTGTTGATGCTGCTTCTTATGAAGCT 639
IMGA|Medtr6g070470.1      AGTTTTAATGCTACTGAAGATTTTGAGGCTGTGTTTGCTGCTATTGCTACTTGA------ 642
IMGA|Medtr6g010380.1      TTTTTGACTGCAATTGATGTTTATTCTGTTCTATTTGTTGAAGCTAATGGAAATAGTGCT 624
                            *** * *      *  * ** *   * * * **** **    *  *

IMGA|Medtr6g070540.1      GAGTATATTAAATCTGTTGTTTGA 651
IMGA|Medtr6g070580.1      GGAATTATTAAGTCTGTTGCTTGA 663
IMGA|Medtr6g070470.1      ------------------------
IMGA|Medtr6g010380.1      TTGTAA------------------ 630
```

GROUP 3: F: GAY AWA WWT GGR AAC CCC RT
R: GAY CAA ACG CMD TCC ATT

```
IMGA|Medtr6g070770.1      ATGAAATCAACATTATTT---ACCTTTTCTCTCCTC---TTTTCCTT---TACCTACTTT 51
IMGA|Medtr6g070630.1      ATGAAATCAACATTATTT---ACCTTTTCTCTCCTCCTCTTTTCCTT---TACCTACTTT 54
IMGA|Medtr6g070510.1      ATGAAACCTACTTTACTCGCCACCCTCCCTTTCCTCCTCTTTGCCCTAACCACCTACTTT 60
IMGA|Medtr6g043660.1      ATGAAGCCTACATTAGTTACTACCCTTTGTTTCCTCCTCTTTTCCTTCACCATCTACTTT 60
IMGA|Medtr6g043640.1      ATGAAGCCTACATTAGTTACTACCCTTTGTTTCCTCCTCTTTTCCTTCACCATCTACTTT 60
                          *****  * ** *** *    *** *   * *****   *** ** *    * *******

IMGA|Medtr6g070770.1      CCATTAGCTTTTACTGAAACAGTTGAA------------GACATAAATGGAAACCCCGTA 99
IMGA|Medtr6g070630.1      CCATTAGCTTTTACTGAAACTGTTGAA------------GACATAAATGGAAACCCCGTA 102
IMGA|Medtr6g070510.1      CCATTACCCTTTACTCAAGGTGTTGAACAA---GTGAAGGATAAAAATGGAAACCCCATT 117
IMGA|Medtr6g043660.1      CCATTACCGTTTACACATGCAAATGATTTTATCGTCAAAGATATATTTGGGAACCCCGTT 120
IMGA|Medtr6g043640.1      CCATTACCCTTTACACATGCAAATAAAATTATCGTCAAAGATATATTTGGGAACCCCGTT 120
                          ****** * *****  *      * *            ** * *  *** ****** *

IMGA|Medtr6g070770.1      TTTCCTGGTGGCAAATACTATATTGCACCACTTATATCTAAAGGAGGAGGTGGTGGATTG 159
IMGA|Medtr6g070630.1      TTTCCTGGTGGCAAATACTATATTGCACCACTTATATCTAAAGGAGGAGGTGGTGGATTG 162
IMGA|Medtr6g070510.1      CTTGTGAGTAAGAAATACTTTATTTGGCCAG------CTGATGGAAGTGGTGGTGGACTT 171
IMGA|Medtr6g043660.1      GTTCCTAGTGGCAGCTACTATATTTGGCCTGAT-TACTTAGT--AAGTGGTGGTGAATTG 177
IMGA|Medtr6g043640.1      GTTCCTAGTGGCAGCTACTATATTTGGCCTGAT-TACTTAAT--AAATGGTGGTGAATTG 177
                           **    **   *  **** ****   **         *     *   ******* * *

IMGA|Medtr6g070770.1      AAGCTTGGTAAAACAGGAGACTCAGAATGTCCAGTTACTGTCATACAAGATTTTTCTGAG 219
IMGA|Medtr6g070630.1      AAGCTTGGTAAGACAGGAGACTCAGAATGTCCAGTTACTGTCCTACAAGATTTTTCTGAG 222
IMGA|Medtr6g070510.1      AGGCTTAATGAGACAGAACA------ATGTCCACTTGTTGTACAACAAGCTTTTTCTGAG 225
IMGA|Medtr6g043660.1      AGGCTTGGTGAAACAGAAAATTCAACATGTCCATTTACTGTACTTCAAGATTATTCTAAC 237
IMGA|Medtr6g043640.1      AGGCTTGGTGAAACAGAAAATTCAACATGTCCATTTACTGTACTTCAAGATTATTCTAAC 237
                          * ****  * * **** * *      ******* **  ***    **** ** **** *

IMGA|Medtr6g070770.1      GTTGTTAGAGGCTTTCCAGTTAGGTTTATCATTCGAGTAAGAC---GTGGTGTCATCTTT 276
IMGA|Medtr6g070630.1      GTTGTTAGAGGCTTACCAGTTAGATTTACCATAATAGTAAAAC---GTGGTGTCATCTTT 279
IMGA|Medtr6g070510.1      GATGTTAAAAGCTTGCCACTAAAATTTATACCAACAGAAAACATCAATGATTTCATCTTT 285
IMGA|Medtr6g043660.1      CTTGGTCCTGGCCTGCCAGTAAAATTTACCCCACAAAATCAAACAAGTGGTGATGATCCC 297
IMGA|Medtr6g043640.1      CTTGGTCCTGGTCTGCCAGTAAAATTTACCCCACAAAATCAAACAAGTAGTGATGATCCC 297
                            ** *    *  * *** * *  ****       *           *  *

IMGA|Medtr6g070770.1      ACAACTGACGAA---C---TAGATATTGAGTTTGTTAAGAAACCAAAGTGTGCTGAATCT 330
IMGA|Medtr6g070630.1      ACAACTGATGAA---G---TAGATATTGAGTTTGTTAAGAAACCAAAGTGTGCTGAATCT 333
IMGA|Medtr6g070510.1      ACTGGCTATACATCAC---TAGATATTGTGTTTGAAAAGAAGACAAAATGTGCTGAATCC 342
IMGA|Medtr6g043660.1      ATCACTTTAAGCTTGCATATAGACATTGCATTTGAAAACAAGCCAGATTGTGCAGAATCC 357
IMGA|Medtr6g043640.1      ATCACTTTAATGTTGCCTATTGAGATTACATTTGAAAACAAACCAGATTGTGCAGAATCC 357
                          *                 * ** ***   ****  ** **  ** * ***** *****

IMGA|Medtr6g070770.1      GCAAAATGGGTTCTAGCTCATG------ATGACTTCCCTACATCTTGGGTAGGTATTGGT 384
IMGA|Medtr6g070630.1      GCCAAATGGGTGCTTGCTCATG------ATGATTTCCCTACATCTTGGGTAGGTATTGGT 387
IMGA|Medtr6g070510.1      TCCAAGTGGGTGGTAGTTAAAG------GTGGTTTCATGGAACCATGGATAGGTATTGGT 396
IMGA|Medtr6g043660.1      TCCAAGTGGTTGGTGGTTGAAGCAGAAAATGAGTACCCTACACCATGGCTGGCTATTGAT 417
IMGA|Medtr6g043640.1      TCCAAGTGGTTGGTGGTTGAAGCAGAAAATGAGTACCCTACACCATGGGTGACTATTGAT 417
                           * ** *** *  * * * * *       **  * *     * * *** *   ***** *

IMGA|Medtr6g070770.1      GATAATATTGATG-------CATTTCAA--GGTAAATTTAAAATTGAGACACTTGGTTCA 435
IMGA|Medtr6g070630.1      GATAATATTGATG-------CTTTTCAA--GGTAAATTTAAAATTGAGACACTTGGTTCA 438
IMGA|Medtr6g070510.1      GGTGGTGTTAATGGTAAGAGTGTTATAGATGGTTTGTTTAAGATTGAGACAATTAG---G 453
IMGA|Medtr6g043660.1      GGTACTGGAAAGA--AGGTTTATGATGA-TGGCTGGTTTGAAATTATTGGATACAA---G 471
IMGA|Medtr6g043640.1      GGTACAAACAAGA--ATGTTTATGATG----GTTATTTTATGATAGTTGGATTCAA---G 468
                          * *       *           *        *    ***   **      *

IMGA|Medtr6g070770.1      GGATCAGGAGCATACAAGCTTGTGTATTGTCCATTGTTCAGTGCTCCACCTGGTGCTTGT 495
IMGA|Medtr6g070630.1      GGATCAGGAGCATACAAGCTTGTGTATTGTCCATTGTTCAGTGCTCCACCTGGTGCATGT 498
IMGA|Medtr6g070510.1      AGTTTTCGTGGATACAAGCTTGTGTTTTGTCCCACCATCAGTGATCCAACTGGTCAGTGT 513
IMGA|Medtr6g043660.1      AAAACGGGATACCTTATCTATTTCTGTCACAAGTTAT---CTCCTACACTAGGTGAATGT 528
IMGA|Medtr6g043640.1      AAGACGGGATACCTTATCTTTTTCTGTCACAAGTTATTATCTCCTACACCAGGTGTATGT 528
                                 *       *    * * * *             *  * **   ***   ***

IMGA|Medtr6g070770.1      AGTGATATTGGGAGGTAT---AGAGATGAGAATGGATGGCGTTTGGTCCCAACTGAAAAT 552
IMGA|Medtr6g070630.1      AGTGATATTGGGAGGTAT---AGAGATGAGAATGGATGGCGTTTGGTCCCAACTGAAAAT 555
IMGA|Medtr6g070510.1      AATAATATTGGAAGATTTTTTGATAATGAGAATGGACTGCGTTTGATCATGAGCGAAAAT 573
IMGA|Medtr6g043660.1      ATTTACTTAAGTAGGAAG---AATGACAAAAATGGAATGCGTTTGGTCTATGAAATGGAT 585
IMGA|Medtr6g043640.1      ATTTATTTAAGTAGGAGG---AATGACGAAAATGGAATGCGTTTGGTCTATGAAATGGAT 585
                          * * *  *  * **           *  * ******  ******* **          **

IMGA|Medtr6g070770.1      ---GATCCCTTTAGGGTTGTATTTGTTGATG---------CTACTGAATCTGAAAAAGCT 600
IMGA|Medtr6g070630.1      ---GATCCCTTTAGGGTTGTATTTATTGATG---------CTACTGAATCTGAAAAAGCT 603
IMGA|Medtr6g070510.1      TTTAAACCTTTTGAAGTTGTATTCGTTGATGTTGAAGATACTGCTGGATTCGGAAGATCT 633
IMGA|Medtr6g043660.1      GGTGATGCCTTAGCAGCGGTATTCGTTAACATTAATGATGCTGCTAGAGCTAGAAGATCA 645
IMGA|Medtr6g043640.1      GGTGATGCCTTAGGAGCGGTATTCGTTAACGTTAATGATGCTGCTAGAGCTCGAAGATCA 645
                              *  * **    *  *****  ** *           ** **  *     ** * *

IMGA|Medtr6g070770.1      ---GTAGTTTGA------------------------------- 609
```

```
IMGA|Medtr6g070630.1      ---GTAGTTTGA-------------------------------- 612
IMGA|Medtr6g070510.1      ---GTAGTTTGA-------------------------------- 642
IMGA|Medtr6g043660.1      TCGGCGATTTGA-------------------------------- 657
IMGA|Medtr6g043640.1      TCGGTACTTAAGAAGGACCACGCATTTAGGCTACCTATGATCTGA 690
                             *   **
```

Group 4: F: CTC TTT GCM TTA RCC AYC TRC TTT YCA T
R: TCA TCA TRC CTC CCA ATA TCA TRA CAA

```
IMGA|Medtr6g071280.1      ATGAAGCCTACATTCCTTACT---CTTTCTTTGATCCTCTTTGCATTAACCATCTGCTTT 57
IMGA|Medtr6g070610.1      ATGAAGCCTACATTCCTTACCACCCTTTCTCTACTACTCTTTGCCTTAGCCACCTACTTT 60
                          ********************    ****** *  * ******** *** *** ** ****

IMGA|Medtr6g071280.1      TCATTAGCTTTTGCACAAGTTTCTGAACAAGTTTTTGACACAAATGGAAACCCCATTTTT 117
IMGA|Medtr6g070610.1      CCATTAGCTTTTAC---------TGAACAAGTGCGTGACTCAAATGGAAACCCTATTTTC 111
                           *********** *         *********   **** ************* *****

IMGA|Medtr6g071280.1      CCTGGTGGCACATTCTATATTATGCCATCAATCTTTGGAGCAGCTGGTGGTGGATTGAGG 177
IMGA|Medtr6g070610.1      TTTTCTAGTAGATTCTATGTTAAGCCATCTATATTTGGAGCTGCTGGTGGTGGAGTAAAG 171
                            *  * * * ******* *** ****** ** ******** ************ * * *

IMGA|Medtr6g071280.1      CTTGGTAAGACAAAAAACTCAAAATGTCCACTTACTGTACTTCAAGATTATTCTGAGGTT 237
IMGA|Medtr6g070610.1      CTTGGTGAAACAGGAAACTCATCATGTCCACTTACTGTACTTCAAGATTATTCTGAGGTC 231
                          ****** * ***  *******  ***********************************

IMGA|Medtr6g071280.1      GTCAATGGTCTGCCTGTAAAGTTTACCA--GGC-TAGAAGCAGGCCATGATATAATCTCA 294
IMGA|Medtr6g070610.1      GTCAACGGTCTCCCGGTAAAATTTAGCACTGACGCTGAAATTTTTATTGATTTGATATCT 291
                          ***** ***** ** ***** **** **  * *   ***        **** * ** **

IMGA|Medtr6g071280.1      ACAAACACAGCAC---TTGATATCGCATTTACAACAAAGCCCGACTGTGCTGAATCCTCC 351
IMGA|Medtr6g070610.1      ACAGACACGTCAAGAGTGGATATTGTATTTCCAGAAAAGCCAGAGTGTGCCGAATCCTCG 351
                          *** ****  **    * ***** * **** **  ****** ** ***** ********

IMGA|Medtr6g071280.1      AAGTGGGTGTTGGTTGATGATTTTAACAAATTAACTGGACCATGGGTTGGTATTGGTGGT 411
IMGA|Medtr6g070610.1      AAGTGGTTGCTGATTGAAGAT------GATTTCCCTAGACCATGGGTTGGTATTGGTGGT 405
                          ****** ** ** **** ***       * **  ** ***********************

IMGA|Medtr6g071280.1      ACTGAAGATAATGAAGATA---------------------------------------- 430
IMGA|Medtr6g070610.1      ATTGAAGACTATATAGGTAAGCATATCATAGATGGTAAGTTTAAAATCGTGAAACATGGT 465
                          * ******  **  ** **

IMGA|Medtr6g071280.1      ------------------------------TCACTGCCCCACCTGGTGCTTGTTATGAT 459
IMGA|Medtr6g070610.1      TTTGGATACAAACTTGTGTTTTGTCCTACATTCACTGCTCCACCAGGTCTTTGTCATGAT 525
                                                  ******* ***** ***  **** *****

IMGA|Medtr6g071280.1      ATTGGGAGGCATGATGACTTTACTGGAAGACTTCTGGTCCTGGCTAATAATGATCCCTAT 519
IMGA|Medtr6g070610.1      ATTGGGAGGTATGATGATAAGAATGGACGGCGTCTCATTCTTACAGAAGATGATCCATAT 585
                          ********* *******    * **** * * ***  * **  *  *  ******* ***

IMGA|Medtr6g071280.1      GAAGTTGTATTTGT---TGATGCCATGGGAAATTGA-------------- 552
IMGA|Medtr6g070610.1      GAGGTTGTCTTTGAGCATGTTGCCATTGGAACTGAGAGATCTGTTGTTTGA 636
                          ** ***** ****    ** ****** **** *
```

Appendix 4: Sequence alignments for designing primers to obtain the 3'UTR regions

```
TrKPI2  ------------------------------------------------------------
TrKPI3  ------------------------------------------------------------
TrKPI5  ------------------------------------------------------------
TrKPI4  ----------------------------------------CTCTTTGCATTAACCATC 18

TrKPI2  ----------------------------------------GATATAAGTGGCAACCCCCTT 22
TrKPI3  ----------------------------------------GACTTACATGGTACCCCCATT 21
TrKPI5  --------------------------------TGTAGTGGACAAAAATGGCAACCCCGTT 28
TrKPI4  TGCTTTCCATTAGCTTTTAGTTCTATTGAACAACTAGAAGACTTAAATGGAAACCCCATC 78
                                                 **   *  *** * * ** *

TrKPI2  ATTCC-GGTGGAAAATACTACATTTTTCCGGTATCTCATGATGAGACATATGGTGGAGGA 81
TrKPI3  TTCCCTGGTGGAAAATACTACATTTTTCCAGTATCTCATGATGATACATATGGTGGAGGA 81
TrKPI5  GTCTCTGGTAAACAATACTTCATTTTTCCAGCAACTGATAAC---CCTAAAAAGGGAGGA 85
TrKPI4  TTTTATTCTACTCGTTTCTATATTATGCCATCTATCTTTGGA---GCTGCAGGTGGTGGA 135
         *      *      * **  *** * **                *       ** ***

TrKPI2  TTAAGACTAGCAAAAACCGG------CGATTCAAAGTGTGATGTTACTGCCTTACAAGAT 135
TrKPI3  TTAAGATTAGCAAAAACTGG------TGATTCAAAGTGTGAAGTTACTGCCTTACAAGAT 135
TrKPI5  CTAACCTTAAACAATGTCGGCGACGATGATTCAAAATGTCCAGTAACTGTCCTACAAAAC 145
TrKPI4  CTCAAGCTTGGTGAAACTGG------AAAATTGACATGTCCACTTACTGTACTTCAAGAT 189
         * *   *     *    **        * *  *  ***    * ***    * *** *

TrKPI2  GACAATATAGTTATCGACAATATACCAGTGAAGTTCAGTATACCAG---GAATAAGTCCT 192
TrKPI3  GACAATATAGTTATCGACAATATACCAGTGAAGTTCAGTATACCAG---GAATAAGTCCT 192
TrKPI5  AAT------GCCATAACAGGTTTACCAGTTAAATTCACCATCCCAC---AAACCACCACC 196
TrKPI4  TATTCTGAAGTTATCAATGGTCTGCAACTAAAATTTACCCCTCCAGGTGAAATTTTCGTT 249
              *  *       * * * * * ** ** *     ***      *

TrKPI2  GGTATAATTTTCACTGGTACGCCGATTGAG---ATTGAGTTTACAAAGAAGCCTAGTTGT 249
TrKPI3  GGTATAATTTTCACTGGTACGCCGATTGAG---ATTGAGTTTACAAAGAAGCCTAGTTGT 249
TrKPI5  GATAATATCGTAACAGGTACCGATCTTGAC---ATCGAGTTCACCGAGAAACCTGATTGT 253
TrKPI4  GATTTGATAAGTACAGACCAACCACTGAAAGGTATTGAATTTGTAGAGAAGCCAGAGTGT 309
        * *   **    ** *          *  *    ** ** **     **** **    ***

TrKPI2  GTTGAATCATCGAAATGGTTGATATTTGTTGACGATGTTATTCAAAAAGCTTGTGTTGGT 309
TrKPI3  GTTGAATCATCGAAATGGTTGATATTTGTTGACGATGTTATTCAAAAAGCTTGTGTTGGT 309
TrKPI5  GCTGAATCATCAAAATGGTTACTAGTTACTGATGATAATACTCAACAAAGCTATGTTGGT 313
TrKPI4  GCTGAATCCTCCAAGTGGGTGGTGGTCGAAGACGATGATTTCCCTCGACCATATGTCGGA 369
        *   **** ** ** *** *  *  *    **  **  *        *   * *** **

TrKPI2  ATAGGTGGTCCTGAAAATTATCCTCAT---------------------------------- 336
TrKPI3  ------------------------------------------------------------
TrKPI5  ATTGGTGGTACGAAAAACTATC------------------------------------- 335
TrKPI4  ATTGGTGGTATTGAAGACAATAAAGGTAAGAGGATCATAAA---TGGTAGCTTTAAAATT 426

TrKPI2  ------------------------------------------------------------
TrKPI3  ------------------------------------------------------------
TrKPI5  ------------------------------------------------------------
TrKPI4  GTGAAACATGGTTTTGGATACAAGATTGTGTTTTGTCCCCGATTCACTGCACCACCTGGT 486

TrKPI2  ------------------------------------------------------------
TrKPI3  ------------------------------------------------------------
TrKPI5  ------------------------------------------------------------
TrKPI4  CTTTGTTATGATATTGGGAGGCATGATGAA------------------------------ 516
```

Appendix 5: Sequence of primers used

Primer Name	Forward/Reverse	Sequence (5'-----3')
GroupIDegF	F	GGAYAYAARTGGYAACCCCMTTWTYCC
GroupIDegR	R	CCARCRTRAYYTTCAGGACCACC
GroupIIDegF	F	GTAKTRGACAWAMATGGYAHCCCC
GroupIIDgR	R	GATARTTTKYSGWACCACCAATACCAA
GroupIIIDegF	F	GAYAWAWWTGGRAACCCCRT
GroupIIIDegR	R	GAYCAAACGCMDTCCATT
GroupIVDegF	F	CTCTTTGCMTTARCCAYCTRCTTTYCAT
GroupIVDegR	R	TCATCATRCCTCCCAATATCATRACAA
3'RACE OligoT Adapter	R	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTV
3'RACE Adapter	R	GACTCGAGTCGACATCG
3' RACE Tr-KPI1	F	CGTAGAGAAGTCAAAAATG
3' RACE Tr-KPI2	F	GATGACAATATAGTTATCGACA
3' RACE Nested Tr-KPI2	F	CCGATTGAGATTGAGTTTACA
3' RACE Tr-KPI4	F	GAAGTTATCAATGGTCTGCAAC
3' RACE Nested Tr-KPI4	F	CCACTGAAAGGTATTGAATTTGTAG
3' RACE Tr-KPI5	F	CTGTCCTACAAAACAATGCC
Tr-KPI1TFF	F	ATGAAGCATGTTTCATCACTCA
Tr-KPI1TFR	R	TTAAACTTCAGACTTAATATACT
Tr-KPI2TFF	F	ATGAAGCCTATGTTATCACTCATC
Tr-KPI2TFR	R	TCAAACAACAGACTTAATTCTAGCTTC
Tr-KPI4TFF	F	ATGAAACCTACAATGCTTACCACCC
Tr-KPI4TFR	R	TCAAGCAACAGATCTTCTAGGAATC
Tr-KPI5TFF	F	ATGAATCCTACTTTATCCATTACCC
Tr-KPI5TFR	R	TTAAATAGGCAATGCAATATTTG

Primer Sequences for developing RNAi knockdown constructs, as indicated

		Tr-KPI1
Sense	F	gacgac**gaattc**TTACAACATTATCAGCGAATACC
	R	gacgac**ctcgag**AGAAGTCAAAAATGGTTTACCAG
Antisense	F	gacgac**tctaga**AGAAGTCAAAAATGGTTTACCAG
	R	gacgac**ggatcc**TTACAACATTATCAGCGAATACC
		Tr-KPI2
Sense	F	gacgac**gaattc**GAATAACATCGTCAACAAATATC
	R	gacgac**ctcgag**TATAGTTACTGAAAGTATACCAGTG
Antisense	F	gacgac**tctaga**TATAGTTACTGAAAGTATACCAGTG
	R	gacgac**ggatcc**GAATAACATCGTCAACAAATATC
		Tr-KPI4
Sense	F	gacgac**gaattc**CACCAAGCTTGAGTCCAC
	R	gacgac**ctcgag**ATGAAACCTACAATGCTTACC
Antisense	F	gacgac**tctaga**ATGAAACCTACAATGCTTACC
	R	gacgac**ggatcc**CACCAAGCTTGAGTCCAC
		Tr-KPI5
Sense	F	gacgac**gaattc**TTGTTTAAGGTTAGTCCTCCC
	R	gacgac**ctcgag**ATGAATCCTACTTTATCCATTACC
Antisense	F	gacgac**tctaga**ATGAATCCTACTTTATCCATTACC
	R	gacgac**ggatcc**TTGTTTAAGGTTAGTCCTCCC
Primers used to screen for RNAi positive construct		
iYAB5-1		GATATGCGAGACGCCTATGA
OCS3'-1		GAGTTCCCTTCAGTGAACGT

Appendix 6: Primers for genome walking (Section 2.3.5)

Primer	GSP11	GSP12
AP1/NA46		GTAATTCGCATCACTATAGCTC
AP2/NA47		ACTATAGCTCACCGCTGGT
TrK-PI1	CTACGATCTTGTAGGACAGTAACCGGAC AC	CAGGGCCGCGAAGTGCTGGCAAAATG
Tr-KPI2	AAACCCACCTTCCTCTGTCACATTACCAG ATC	GGCAGTAACATCACACTTTGAATCGCCGG
Tr-KPI4	CACCTGGAGGGGTAAATTTTAGTTGCAG ACC	ACCACCTGCAGCTCCAAAGATAGATGGCAT
Tr-KPI5	GGTGAACTCGATGTCAAGATCGGTACCT GT	GTCGCC GACATTGTTTAAGGTTAGTCCTCC

Primer	GSP21	GSP22
Tr-KPI1	ACAGAAAGTTGCTTTGAGAGGACCTTTT CGCG	GACAGGCGAGGTAAATGATACAAGGAGTTCAG
Tr-KPI2	CTACAACAAATTGCAAGCGTTCCACGAA GGG	GCTGACGACAGTCTTATGTTCTGTAGGGCA
Tr-KPI4	GAGGAGTAAAGAAAGGGTGGTAAGCAT TGTAGG	CCAAGTTCTTACCTAGCCGGCCACATATGA
Tr-KPI5	GAGATGCACACCATTGACGTGAATAAT TCAAATAGG	CAATGCAGTATTTAAGGCGTGAGACTC TCTCAC

Appendix 7: Sequences of primers used for qPCR to determine expression of the genes indicated

Gene	Forward primer	Reverse Primer
β-actin	CGTATGAGCAAGGAGATCACTG	CATCTGCTGGAAGGTGCT
GAPDH	TCCAGTATTGAACGGTAAATTGAC	TCTGATTCCTCCTTGATAGCAG
Tr-KPI1	GGTAACGCCATCTTCCCAG	CTACGATCTTGTAGGACAGTAACC
Tr-KPI2	GGGAGATCTGGTAATGTGACAG	TCAAGGTATCAAACAACAGACTTAAT
Tr-KPI4	GATATTGGAAGGCATGATGATGAG	GTCTAACAAGTGATCAGCTAACCT
Tr-KPI5	CCATCCCACAAACCACCAC	GGTCCACCAATACCAACATAGC

Appendix 8: Hoagland's Medium (Gibeaut *et al.,* 1997)

Macronutrients

Macronutrients	Working concentration (mM)	Weight (g) per L of 10X stock
KNO_3	1.25	1.264
$Ca(NO_3)_2$	1.50	3.543
$MgSO_4$	0.75	1.849
KH_2PO_4	1.00	1.360

Micronutrients

Micronutrients	Working concentration (µM)	Weight (mg) per L of 100X stock
KCl	50.0	372.8
H_3BO_3	50.0	3.543309.2
$MnSO_4$	10.0	1.849151.0
$ZnSO_4$	2.0	1.36057.5
$CuSO_4$	1.5	37.5
$(NH_4)_6Mo_7O_{24}$	0.075	9.30
Na_2O_3Si	0.10	2.10
FeNaEDTA	72.0	2642.80

-P Macronutrients

Macronutrients	Working concentration (mM)	Weight (g) per L of 10X stock
KNO_3	1.25	1.264
$Ca(NO_3)_2$	1.50	3.543
$MgSO_4$	0.75	1.849
KH_2PO_4	0.01	0.0136

Appendix 9: Tissue culture media used for RNAi white clover lines

CR0	
• MS salts (4.4 g /L)	
• B5 vitamins	
• sucrose	30 g/L
• pH5.8	(using KOH)
• agar	8.0 g/L
CR5	
• MS salts	
• B5 vitamins	
• sucrose	30 g/L
• BA	0.1mg/L
• NAA	0.05mg/L
• pH 5.8	(using KOH)
• agar	8.0g/L
CR7	
• MS salts	
• B5 vitamins	
• sucrose	30 g/L
• BA	l.0mg/L
• NAA	0.05mg/L
• pH 5.8	(using KOH)
• agar	8.0g/L
MGL	
• Mannitol	5.0g/L
• L glutamic acid	1.0g/L
• KH2PO4	250mg/L
• MgSO4	100mg/L
• NaCl	100mg/L
• Biotin	100mg/L
• Bactotryptone	5.0g/L
• Yeast extract	2.5g/L
• pH7.0	(using NaOH)