Table of Contents

List of Tables

List of Figures

Abbreviations and Symbols

BLAST	Basic Local Alignment Search Tool
bp	Base pairs
C1	Papain cysteine proteases
C13	Legumain like protease
Cat-L	Cathepsin L
cDNA	complementary DNA
ddPCR	Digital droplet PCR
DM	Dry mass
DNA	Deoxyribonucleic acid
DNS	Non-stressed (Drought control)
DS	Drought stress
FM	Fresh mass
FPKM	Fragments Per Kilobase of exon model per Million mapped
GO	Gene Ontology
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KOG	EuKaryotic Orthologous Groups
LEA	Late embryogenesis abundant
MAFFT	Multi Alignment using Fast Fourier Transform
NCBI	National Centre for Biotechnology Information
NS	Non-Stressed
ns	Non significant
OCI	Oryzacystatin I

OS	Osmotic stress
PAR	Photosynthetically active radiations
PCD	Programmed Cell Death
PCR	Polymerase Chain Reaction
PI	Plastochron index
R	Recovery
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNS	Non-stressed (Recovery control)
ROS	Reactive oxygen species
RT-qPCR	Real time quantitative PCR
SNF	Symbiotic nitrogen fixation
TAIR	The Arabidopsis Information Resource
VPE	Vacuolar Processing Enzyme(s)
VWC	Vermiculite water content
WT	Wild type

CHAPTER 1

Literature review

1.1 INTRODUCTION

1.1.1 Soybean and its significance

Soybean (*Glycine max* [L.] Merr.) is a very important legume crop world-wide. It is a major source of protein and oil. Not only can it be used as feed and food but it also has an increasing importance in industrial products such as lubricants and hydraulic fluids (Choudhary and Tran; 2011; Hsien, 2015). Soybean was first domesticated in China, 3500 B.C., and from 2013-2014 an estimate of 308 million tons (Fig.1.1) of soybean were harvested throughout the whole word from 117 million hectares (Food and Agriculture Organization of The United Nations Statistics Division 2014). The main production area is concentrated in the Americas (87.7 %) whereas Africa only contributed 0.8 % of the total soybean harvested with 2.2 million tons. South Africa and Cameroon are the two counties in Africa with the highest soybean production. South Africa produced 948 000 tons according to the National Crop Estimates Liaison Committee (CELC) and 72.5 % of production was from the Free State and Mpumalanga provinces. The estimated 1.89 t/ha produced in 2013-2014 needs to be increased to 2.3 t/ha according to the BFAP Baseline, Agricultural Outlook 2014 – 2023.

Soybean seeds consists of 40 % protein, 21 % oil, 34 % carbohydrate and 5 % ash (Scott and Aldrich, 1983). The numerous benefits of consuming soybean have increased the production over the years. A few of these benefits are: prevention of cancer, lowering of cholesterol and protection against bowel and kidney diseases (Foyer *et al.*, 2016) just to mention a few. Soybean produces more protein per area of land when compared to livestock and other crops and is therefore an excellent substitution for protein. (Dovring, 1974).

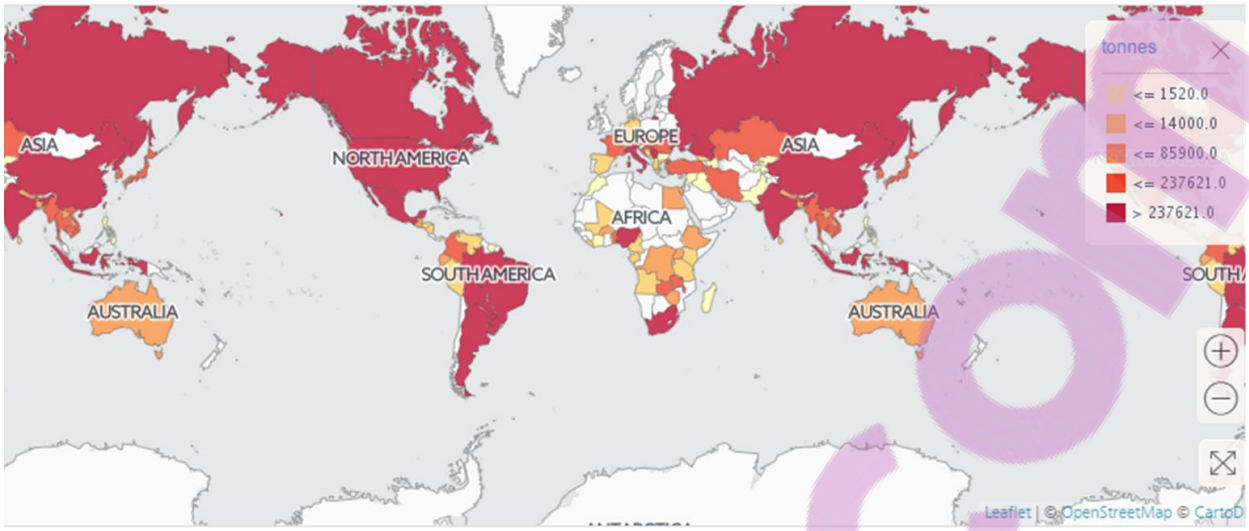


Fig. 1.1: Soybean production globally (2013-2014). Figure taken from FAO website Global crop production statistics (Food and Agriculture Organization of The United Nations Statistics Division 2014).

Another advantage of soybean is its ability to fix nitrogen due to its symbiotic relationship with the soil borne microbe *Bradyrhizobium japonicum*, which leads to the formation of root nodules (Puppo *et al*., 2005). This symbiotic relationship allows soybean to be used as a nitrogen source in intercropping and crop rotation systems (Keyser and Li, 1992), which is helpful for small subsistence farmers.

1.1.2 Root nodule development, physiology and function

For plants and the bacterial micro-symbionts to be able to establish a symbiosis, the bacteria needs to gain access to a single plant cell where they install themselves in compartments surrounded by the plants membrane called a root nodule (Stougaard, 2000). The plant supplies the carbon source to the bacteria which is then used for the reduction of di-nitrogen. Nodule formation can be described as two different processes that has to happen simultaneously,

nodule organogenesis and bacterial infection. The bacterial micro-symbiont is attracted by different plant-derived secondary phenolic compounds such as flavonoids and iso-flavonoids. Nodulation in soybean occurs after the plant lowers its endogenous defences, allowing the bacteria, *Bradyrhizobium japonicum*, to infect a root hair (Fig. 1.2). Both nodule organogenesis and bacterial infection is dependent on lipochito-oligosaccharides, also referred to as Nod factors that is released by the bacteria (Oldroyd *et al*., 2011).

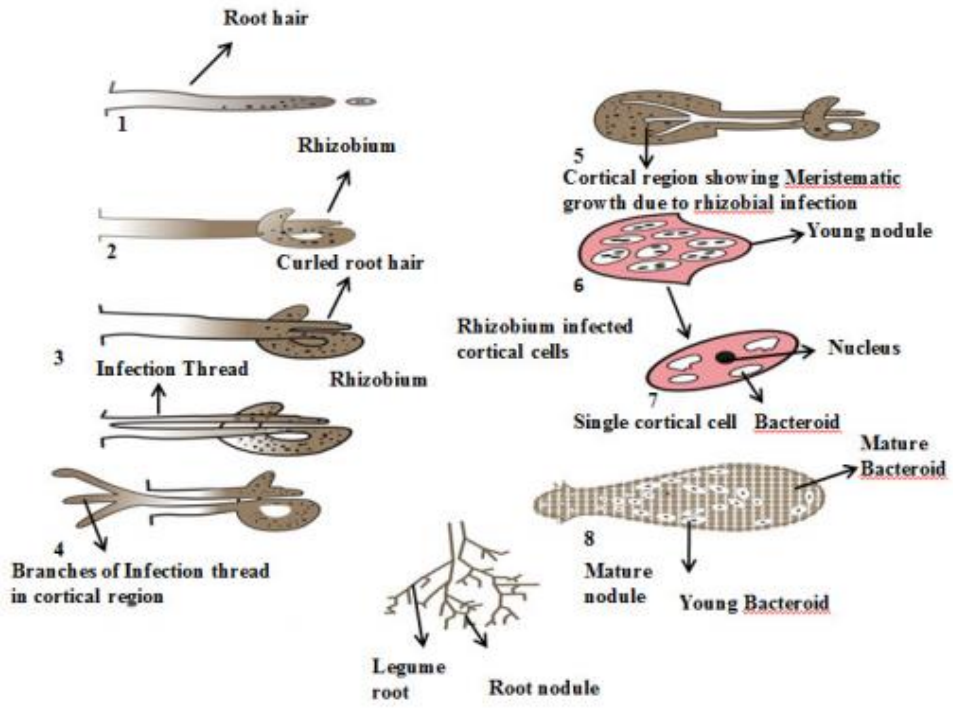


Fig. 1.2: The formation of a root nodule. 1) Rhizobium establish in the root hair 2, 3) the root hair begins to curl 4) an infection thread is formed by the bacteria 5) the infection thread branches in the cortical region 6) meristematic growth occurs 7) a young root nodule is formed containing the *Rhizobium* 8) development of the bacteriods (Muneer *et al.,* 2012).

Determinate nodules that is associated with soybean, has a transient meristem and arises from the central cortex. Nodulation factors activate the formation of the root cortex, which leads to cytokinin signalling at an early stage of nodule development (Lohar *et al*., 2006). Auxin

transportation is inhibited to allow for the initiation of the nodule meristem (Oldroyd *et al*., 2011). Meanwhile root hairs that receive the nodulation factors will bend back on itself entrapping the bacteria between the cell wall (Geurts *et al*., 2005). After the plant cell wall degrades, the bacteria will start to colonise and form an infection thread containing the bacteria in a glycoprotein matrix. This leads to the formation of the peribacteroid which allows the bacteria to be isolated from the plant cell's cytoplasm. The plant supplies the bacteria with nutrients and creates a low O_2 environment which is needed for nitrogen fixation (Colebatch *et al*., 2004; Puppo *et al*., 2005; Oldroyd *et al*., 2011).

Legumes can have two types of nodules (Fig. 1.3): determinate nodules (e.g. soybean and common bean) or indeterminate nodules (e.g. alfalfa and pea). Both these types of nodules are similar in the fact that they have a central infection zone, the inner, middle and an outer cortex. These two types of nodules can be distinguished based on their development. Determinate nodules has a globular structure due to the vascular tissue on the boundary and has no apical meristem (Van de Velde *et al.,* 2006). Indeterminate nodules have a more elongated structure due to the vascular tissue and the apical meristem is situated on the terminal ends of the nodule (Puppo *et al.*, 2005). These nodules also differ in the area where senescence is initiated. When determinate nodules undergo senescence, the process will start at the centre and spread to the outer edges whereas in indeterminate nodules, the process will start at the outer edge closest to where the nodule is attached to the plants (Puppo *et al*., 2005).

‘

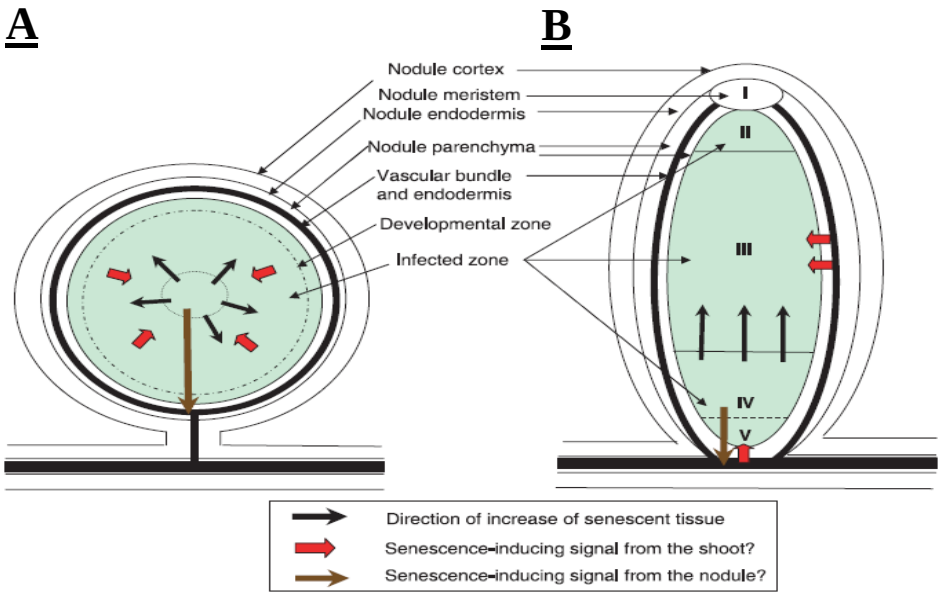


Fig. 1.3: Physiology and morphology of root nodules, A) determinate nodule with a globular structure due to no apical meristem in comparison to B) elongated indeterminate nodules (Puppo *et al*., 2005).

The nitrogenase enzyme system is required for the process of symbiotic nitrogen fixation (SNF) to occur. Nitrogenase consists of two components namely an iron (Fe) protein and a molybdenum iron (MoFe) protein. It firsts binds with ATP and with the MoFe-protein. The ATP will then be hydrolysed and an electron transfer will take place between the two proteins. The MoFe-protein will then bind to the substrate and reduce di-nitrogen to ammonia (N_2 + 8 H^+ + 16 ATP + 8 e^- → 2 NH_3 + H_2+ 16 ADP +16 Pi) which will then be converted to ammonium: NH_3 + H^+ → NH_4 (Rees *et al*., 2005). This process and production of nitrogen allows the plant to synthesise essential macromolecules such as proteins.

The ability of the symbiosis to produce nitrogen is an advantage for plant growth, development and the ultimate yield, as it lowers the demand for nitrogen fertilization. The symbiosis between

soybean and *Bradyrizobium* can fix 300 kg N ha^{-1} under optimal conditions (Keyser and Li, 1992). This symbiosis however is very short lived and only lasts 10-12 weeks, rapidly declining until the pod-filling stage (Alesandrini *et al*., 2003; Puppo *et al*., 2005). The symbiosis is also sensitive to external abiotic stress factors such as drought, cold and salt stress which leads to a shortened life cycle. This causes limited nitrogen availability for plant growth and development as well as pod filling which affects yield (Alesandrini *et al*. 2003).

1.1.3 Nodule senescence

Nodule life span depends on the plant species, the *Rhizobium* strain and on different environmental conditions (Swaraj and Bishnoi, 1996). Senescence of the nodule tissue will start in the centre of the nodule (determinate nodules) and will spread, within a few weeks, to the periphery. This process is also visible by a colour shift inside the nodules SNF zone, where the active pink nodule will change to a non-active greenish nodule due to the degradation of the heme-group of the leghemoglobin (Van der Velde *et al*., 2006). The nitrogen fixation ability of the nodule decreases when leghemoglobin degrades in nodules (Puppo *et al*., 2005). The auto-oxidation of oxygenated leghemoglobin to ferric leghemoglobin produces ROS (reactive oxygen species), such as superoxide which will cause oxidation of the bacteriod. Senescence leads to a decrease in the carbon to nitrogen ratio inside the nodule which decreases the antioxidant metabolite (ascorbate-glutathione antioxidants) availability to convert the produced ROS (Puppo *et al*., 2005). This is followed by different ultra-structural changes in the nodule and the symbiosome. The cytoplasm of the nodule will become less dense, leading to the appearance of vesicles and ghost membranes due to the disintegration of the symbiosome membrane and consequently leading to the elimination of the microbial partner (Timmers *et al*., 2000). The most prominent change during nodule senescence is the increase in protein

degradation caused by an increase in proteolytic activity within the nodules (Pladys and Vance, 1993). More than 2500 genes are activated in the nodule transcriptome during senescence where 7 % of these genes are different hydrolases and proteases (Martínez *et al*., 2007).

Nodule senescence usually coincides with pod-filling in most leguminous crops. Exogenous addition of nitrogen fertilizer during the pod-filling stage will increase plant yield and seed quality (Merbach and Schilling, 1980). Unfortunately nodule senescence can also be induced at an earlier stage in the nodules life cycle by different environmental stress conditions, such as drought (Kunert *et al*., 2016), salt stress (Du Pont *et al*., 2012) and cold stress (Van Heerden *et al*., 2008). The nitrogen fixation capability of nodules under different stress conditions is decreased prior to the degradation of leghemoglobin (Escuredo *et al*., 1996; Gogorcena *et al*., 1997; Matamoros *et al*., 1999).

1.1.4 Drought stress and drought tolerance

Biological stress can be defined as a condition or force which will ultimately inhibit plants to function normally and will also affect the well-being of the biological system of these plants (Mahajan and Tuteja, 2005). Drought stress is one of the most important threats to food security worldwide (Cutforth *et al*., 2007) and the intensity and severity of drought stress are predicted to increase (Jury and Vaux, 2007). Plants experience drought stress in two ways: firstly when the water supply for roots becomes inaccessible and secondly when transpiration rates becomes too high (Anjum *et al*., 2011). Drought impacts a lot of parameters including: membrane integrity, osmotic adjustment, water relations, photosynthesis, growth and ultimately yield (Benjamin & Nielsen, 2006). Plants will undergo molecular, biochemical and physiological changes when subjected to severe drought stress. ROS is activated during the onset of drought

stress and acts as secondary messengers to activate lipid peroxidation, protein deactivation, DNA fragmentation and cell death (Anjun *et al*., 2011). Different genes are activated or down-regulated in response to drought. Some genes are transiently expressed within a few minutes of stress perception which activates transcription factors as well as calcium sensors. These genes, in return activate delayed genes that will modulate stress tolerance effectors such as antioxidants, molecular chaperones as well as LEA-like proteins (late embryogenesis abundant) (Mahajan and Tuteja, 2005).

Soybean yield is also threatened in many parts of the world due to climatic changes and persistent drought (Foyer *et al*., 2016). Drought stress is usually combined with high light and high temperatures (Chaves *et al*., 2003). It has been observed that the soybean yield fall with 2.4 % for every 1 $^{\circ}$C rise in temperatures and that more than 30 % of yield has been suppressed in the US alone due to drought (Kunert *et al*., 2016). Eleven million tons of soybean seed was lost in Brazil due to severe drought stress in the 2008 – 2009 growing season (Franchini *et al*., 2009).

Drought does not only negatively impact plant growth, it also affects the endosymbiotic nodule bacteria and inhibits SNF (Sprent, 1972, Zahran, 1999). Within the root nodules, drought affects O_2 availability, O_2 diffusion, causes a nitrogen feedback mechanism due to an accumulation of ureides and a carbon limitation (Larrainzar *et al*., 2009). This causes a decrease in SNF which leads to less available nitrogen for the biosynthesis of proteins, resulting in lowered yields (Farooq *et al*., 2012). Nodules numbers have been seen to decrease in severe drought stress conditions (Marques-Garzia *et al.*, 2015). Certain plants and cultivars have strategies to avoid drought stress. These strategies are: drought escape, drought avoidance and drought tolerance. Drought escape can be defined as plants having a shorter life cycle and

completing the reproduction phase earlier than other cultivars, before soil water becomes limited. Avoiding tissue dehydration by reducing water loss from aerial parts by closing stomata or rolling of leaves and more efficient water uptake from a longer and deeper tap root is drought avoidance. Lastly the ability for plants to recover from drought stress is called drought tolerance. Plants should be able to maintain turgor and continue with metabolism even at a low leaf water status. This can be achieved by osmotic adjustments (Chaves *et al*., 2003). The selection of more drought tolerant soybean cultivars is required due to severe soybean yield losses after drought stress to address the threat of food and protein security (Ku *et al*., 2013).

Three ways to produce more drought tolerant crops are by natural selection, classical breeding approaches and genomic-assisted breeding (Cattivelli *et al*., 2008). Due to the growing world population it is necessary to increase crop production to a maximum. Classical breeding approaches increases the yield 1 % per year (Kucharik and Ramankutty, 2005). Unfortunately classical breeding has focused on improving yield which is not always sustainable under stressful environments (Kunert *et al*., 2015). To breed for drought tolerance the stress has to be introduced year after year which makes this method difficult in a natural environment. Yield increases also have to be associated with improved stress tolerance to ensure improved crop production in changing climatic conditions. Genomic-assisted breeding sounds like a good alternative but no single-gene target modifications have been reported in literature (Kunert *et al*., 2015) due to the unpredictability of the duration and severity of drought. Delaying the onset of stress symptoms and senescence can possibly enhance yield in soybean cultivars but this requires the identification of possible molecular markers for drought tolerance. One adaptive mechanism of plants that are activated during drought conditions is the expression of proteases

to facilitate active nitrogen reserves via proteolysis (Simova-Stoilova *et al.,* 2010; Kidrič *et al*., 2014), which could be a possible starting point in identifying molecular markers.

1.1.5 Cysteine proteases and inhibitors

Proteolytic enzymes can be found in all organisms (Beers *et al*., 2004) and play a key role in plant development. These enzymes' mode of action can be divided into two groups. The first is limited proteolysis where only a limited number of peptide bonds are cleaved to activate another protein. The second is unlimited proteolysis, where proteins are degraded completely and provide free amino acids to synthesize new proteins (Fan and Wu, 2005). Enzymes that hydrolyse peptide bonds either internally (endopeptidase), or externally (exopeptidase), cleaving at the amino and carboxy terminal ends are referred to as proteases according to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB, 1992). Proteases are involved in the regulation of plant growth, development, remobilization of storage proteins, defence mechanisms, senescence and cell death (Solomon *et al*., 1999). Proteases can be divided into different functional classes according to their optimal pH range and according to the amino acid residue in the functional centre of the enzyme (nucleophile). These functional classes are: aspartic, serine, cysteine, metallo, threonine, glutamate and asparagine proteases (Grudkowska and Zagdanska, 2004). Threonine, cysteine and serine proteases has a Thr, Cys and Ser residue as nucleophile respectively, whereas metallo, aspartic and glutamate proteases uses water activated by the metal ion Me^{2+}, or aspartate (Van der Hoorn *et al.,* 2008). The asparagine peptide lyase uses an asparagine residue as nucleophile (Deu *et al.,* 2012). Proteases' mode of action is illustrated in Fig. 1.4.

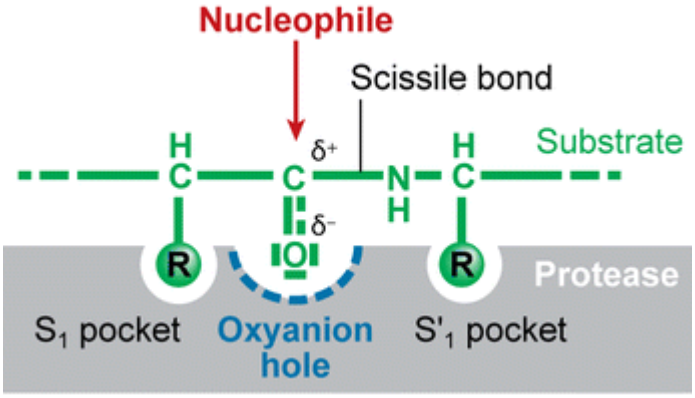


Fig. 1.4: Protease mode of action: the substrate protein (green) binds with an amino acid residue (R) to the substrate binding site of the proteases (grey) by interacting with the substrate pockets (S) of the enzyme. The peptide bond is next to the carbonyl group which get stabilised by the oxyanion (blue) which makes the carbonyl group available for a nucleophilic attack (Van der Hoorn *et al*., 2008).

1.1.5.1 Cysteine proteases

The most prominent group of proteases is cysteine proteases. They account for 90 % of the total degradation of storage proteins in wheat (Botarri *et al*., 1996) and 27 out of 42 proteases were found to be involved with seed germination (Jones and Zhang, 1995). Cysteine proteases are involved in a variety of processes in plants such as growth and development, hormone signalling, embryogenesis and morphogenesis (Salas *et al*., 2008). Cysteine proteases are also involved in: senescence to mobilize nutrients to growing plant organs (Beers *et al*., 2000), during programmed cell death (PCD), with tissue differentiation and during a wide variety of processes during abiotic and biotic stress (Salas *et al*., 2008).

Cysteine proteases can further be grouped into six different families: papain, calpain, clostripains, streptococcal cysteine proteases, viral cysteine proteases and caspases (Dubey *et*

al., 2007). Most of the plant proteases can be grouped into papain (C1, C2 and C10) and caspase families, C13 and C14 (Grudkowska and Zagdanska, 2004). The papain family has been subdivided into C1A that contains disulphide bridges and accumulates in vesicles, and C1B that lacks disulphide bridges and accumulates in the cytoplasm but is not present in plants (Van der Hoorn *et al*., 2008). C1A proteases will be referred to only as C1 cysteine proteases from here on.

1.1.5.2 Papain-like (C1) cysteine proteases

C1 cysteine proteases, the model cysteine proteases, were first isolated in the latex of papaya fruit (Schaller, 2004). C1 cysteine proteases were one of the first proteases to have their three-dimensional structure determined (Grudkowska and Zagdanska, 2004). The structure of C1 cysteine proteases (Fig.1.5) consists of two domains, an α-helix and a β-sheet that is linked to each other in such a way that a cleft is formed containing the amino acid sequence Cys-His-Asn within the substrate binding region (Turk *et al*., 2001). Soybean C1 cysteine proteases consists of ± 496 amino acids. C1 cysteine proteases are stable although they are found in harsh environments such as vacuoles and lysosomes. This is due to the fact that they are present in the cell as pre-proproteins (Fig. 1.6 A) and have an auto-inhibitory pro-domain to avoid unnecessary proteolysis. C1 cysteine proteases could carry a signal for retention in the endoplasmic reticulum at the C-terminus called the KDEL motive (Than *et al*., 2004) or at the N-terminus (NPIR) for vacuole targeting (Ahmed *et al*., 2000). C1 cysteine proteases are grouped according to their closest animal counterparts, Cathepsin (Cat). After a recent phylogenetic analysis Cathepsin can be subdivided into nine different Cathepsin-like families. The first six families are closely related to Cat-L, the seventh to Cat-F, eight to Cat-H and the ninth to Cat-B. A more detailed description can be seen in Fig. 1.6 B. More than 600 C1A cysteine proteases have been included into the MEROPS database (Martínez, *et al*., 2012).

These enzymes are involved in catalysing protein remobilization and are involved in the senescence of different organs, abscission, seed germination, seed ripening as well as PCD (Beers *et al*., 2000; Grudkowska and Zagdanska 2004, Van der Hoorn *et al.*, 2008). They are also important genes in pathogen resistance and systemic defence (Solomon *et al*., 1999; Van der Hoorn *et al.*, 2008). C1 cysteine proteases are also enhanced during different abiotic stress conditions where they trigger the reorganization of the metabolism, degrade and remobilize unnecessary proteins and nutrients as well as remodelling of cell protein components (Martínez *et al*., 2012).

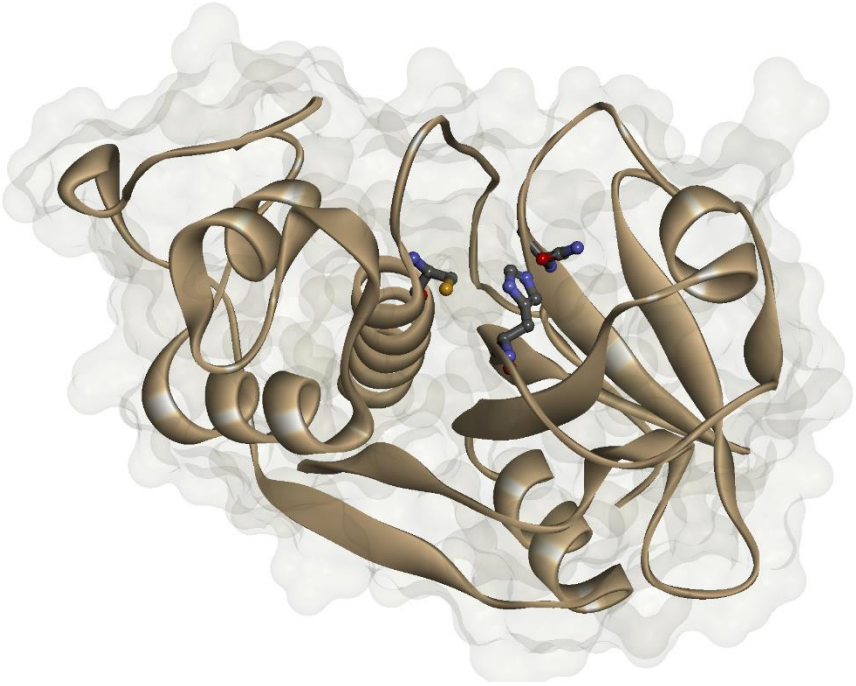


Fig. 1.5: Three dimensional structure of a C1 cysteine proteases. (Image provided by Dr Juan Vorster, UP).

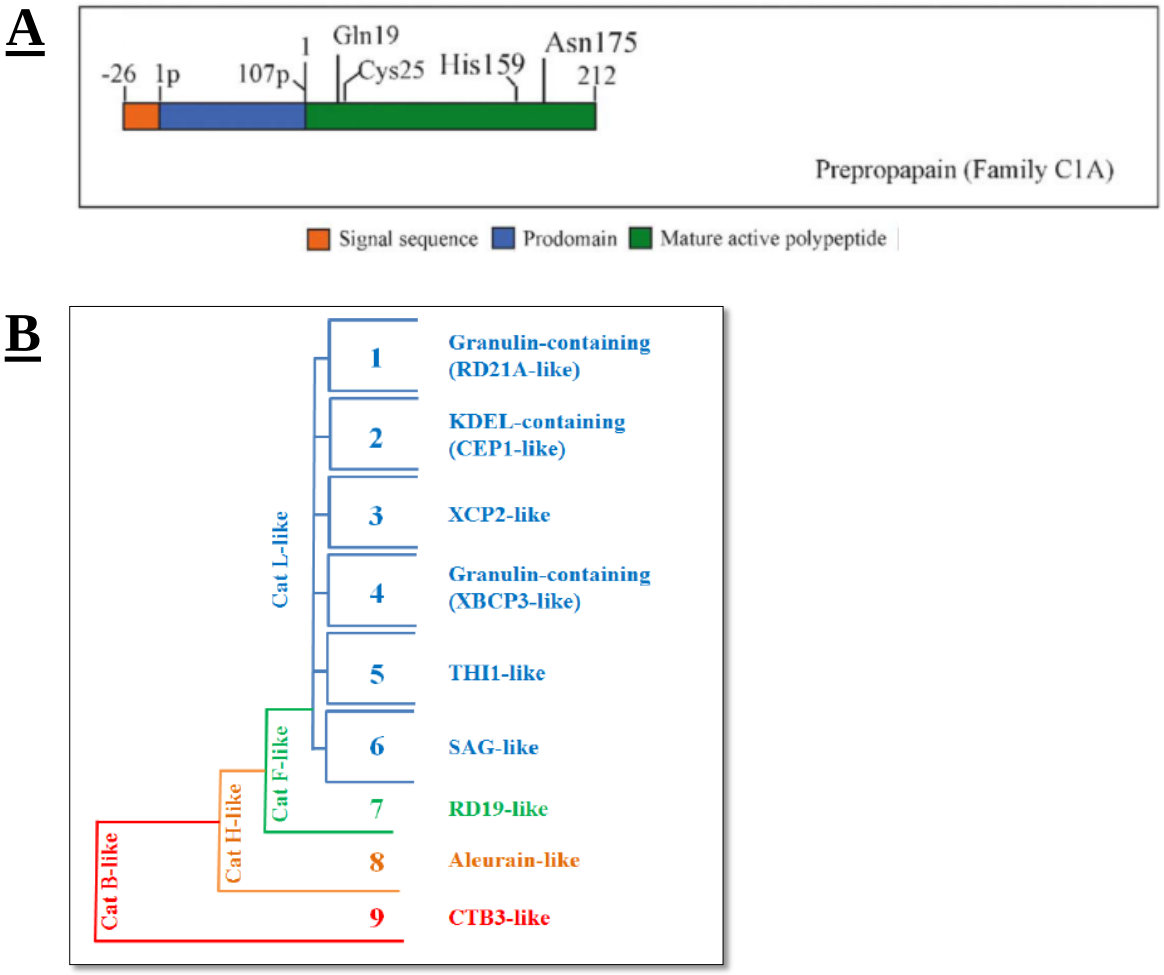


Fig. 1.6: A) Schematic representation of pre-propapain's structure indicating its active site residues (Cys25-His159-Asn175) along with its conserved residue (Gln19) (Beers *et al*., 2004). B) Phylogenetic tree of the different C1 cysteine protease classes divided into 9 subfamilies (text on the right) in four different Cathepsin classes (text on the left) according to *Arabidopsis* C1 cysteine proteases (modified from Richau *et al*., 2012).

1.1.5.3 Legumain-like (C13) cysteine protease

Legumain like (C13) cysteine proteases, also known as vacuolar processing enzymes (VPE), have shown caspase-like activity (Hara-Nishimura *et al*., 2005). Caspases are a type of cysteine proteases that contain an aspartate specific centre that can regulate PCD. The structure (Fig.1.7) consists of a central six-stranded β-sheet, and five major α-helices (Dall and Brandstetter,

2013). C13 cysteine proteases are asparaginyl endopeptidases and cleaves Asn- as well as Asp residues. C13 cysteine proteases has ± 494 amino acids. These proteases proteases are synthesised as propeptides that are co-transcriptionally segregated in the rough endoplastic reticulum as a prolegumain. The prolegumain will then get transported to the cell wall or vacuole where it will self catalytically activate in an acidic environment by the removal of the C and N-terminal propeptides (Christoff *et al*., 2014).

C13 cysteine proteases are important enzymes as they are involved in the process of activating (post translational processing) proteins into their mature forms (Hara-Nishimura *et al*., 2005). It has been observed that they are active in plant tissue such as seeds, cotyledons, roots and leaves (Christoff *et al*., 2014). During germination in seedlings it was also observed that legumains activate C1 cysteine proteases which is involved in the degradation of storage proteins (Christoff *et al*., 2014).

C13 cysteine proteases can be divided according to their expression in different tissues and sequence homology into two different groups. The first group is the seed-type and the second the vegetative-type (Muntz *et al*., 2002). Four C13 cysteine proteases have been classified in *Arabidopsis: a-VPE* and *γ-VPE* are specific to vegetative organs, *β-VPE* are specific to seeds and lastly *δ-VPE* which does not belong to the subfamilies proposed (Yamada *et al*., 2005).

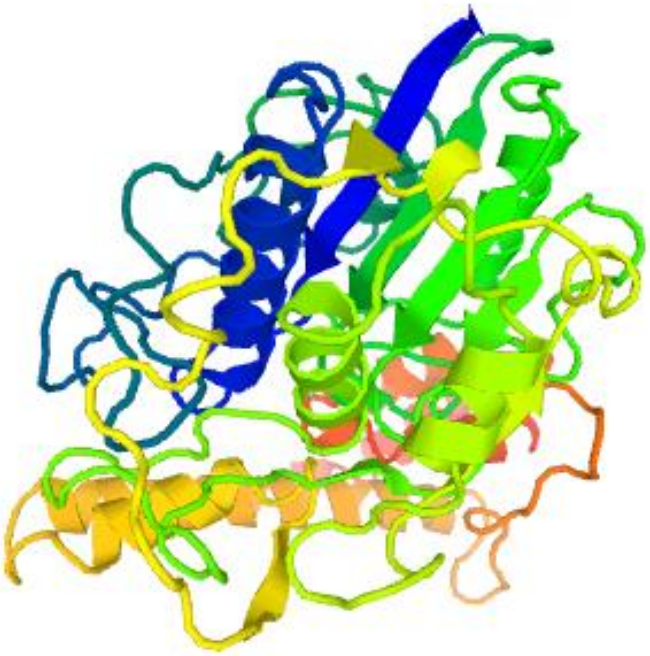


Fig. 1.7: Three dimensional structure of a C13 cysteine protease, γ-VPE. (http://www.uniprot.org/uniprot/P49047).

C13 cysteine proteases play an important role in PCD (Fedorova and Brown, 2007) and have been isolated in senescing leaves and wounded leaves under a stress response (Fig. 1.8) where they act as processing enzymes in events like senescence and PCD (Müntz *et al*., 2002). During a loss of function mutation in VPE genes, cell death was prevented during the hypersensitive response to a fungal toxin (Kuroyangani *et al*., 2005) which leads to the degradation of cytosolic compounds or autophagy during senescence (Yamada *et al*., 2005).

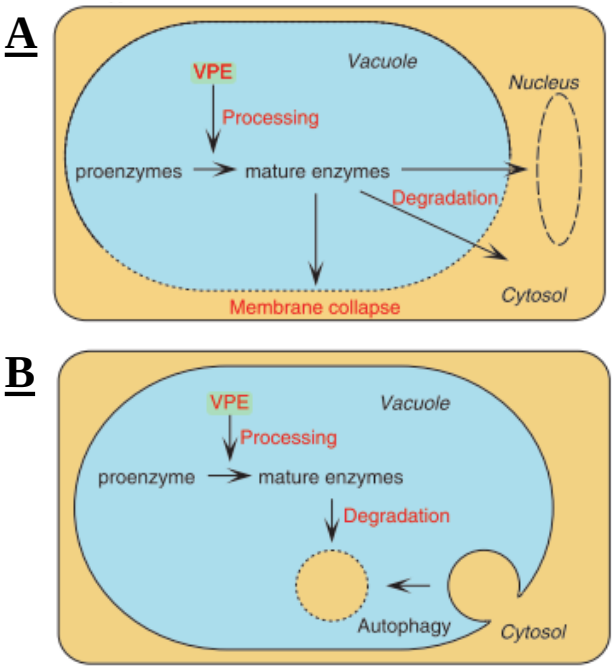


Fig. 1.8: Two models of the role of C13 cysteine proteases during A) hypersensitive response and PCD as well as B) senescence. VPE's mature active enzymes such as proteases and hydrolases (Yamada *et al*., 2005).

1.1.5.4 Nodule cysteine proteases

The expression of cysteine proteases in root nodules have been reported on previously by different research groups (Pfeiffer *et al*. 1983, Kardailsky and Brewin 1996; Espinosa-Victoria *et al*., 2000; Lee *et al*., 2004; Oh *et al*., 2004). Increased proteolytic activity in nodules affect the SNF and the amount of ammonia produced. These proteases directly target the bacteriod and leghemoglobin. (Pladys and Vance, 1993; Pfeiffer *et al*., 1983). Lee *et al*. (2004) divided nodule proteases into two groups: nodule specific and nodule enhanced proteases after investigating cysteine proteases involved in late nodule senescence.

Cysteine proteases' involvement in nodule developmental and induced senescence has been confirmed and the inhibition of certain cysteine proteases has resulted in increased nodule lifespan. A cysteine protease, *CYP15A*, found in *Medicago trunctula* were inhibited and

showed to induce delayed nodule senescence (Sheokand *et al*., 2005). In Chinese milk fetch, a cysteine protease called *ASNODF33*'s, transcription levels increased 5-fold from 14 day old nodules to 30 day old senescing nodules. The involvement of *ASNODF*32 in nodule senescence was confirmed when nodule life span was extended after silencing this proteases (Li *et al*., 2008). A C1 cysteine protease (*MtCP6*) and a C13 cysteine protease (*MtVPE*) were down regulated in *Medicago trunctula* and showed delayed nodule senescence and increased nitrogen fixation whereas over expression of these genes promoted senescence (Pierre *et al*., 2014). Both above mentioned proteases, *MtCP6* and *MtVPE* were found in the vacuole. They are suggested to be involved in amino acid recycling, correction of miss folded proteins and plays a role in autophagic bodies. They were also seen to be involved in vacuole mediated cell death (Pierre *et al*., 2014). Another study showed that three C13 cysteine proteases and one C1 cysteine proteases decreased in transcript level after 21 days of drought and four C1 cysteine proteases transcripts increased slightly after drought stress conditions (Márquez-Garcia *et al*., 2014). Eighteen C1 cysteine proteases were identified to be transcribed in developing soybean nodules as well as in early senescent nodules. Five of these identified C1 proteases were highly expressed in young developing nodules (4-8 weeks). Thirteen C1 cysteine proteases showed to be associated with natural senescence due to an elevated transcript abundance in 14 week old nodules (Van Wyk *et a*l., 2014). They also identified six C13 cysteine proteases that had high transcript levels in older senescing nodules (Van Wyk *et al*., 2014). Esteban-García *et al.* (2010) suggested that C1 cysteine proteases are similar to RD19 and RD21 subfamilies as identified in *Arabidopsis* and are involved in stress responses such as drought stress. Albertini *et al.* (2014) showed that *a-VPE* is involved in drought tolerance in *Arabidopsis*. However, the exact involvement of C1 and C13 cysteine proteases in drought is still unclear. The exact function of C1 and C13 cysteine proteases in root nodules is still unclear as well as their function during nodule drought stress.

1.1.5.5 Cysteine protease inhibitors

Cysteine proteases' activity is regulated *in situ* by proteinaceous inhibitors. C1 cysteine proteases are regulated and inhibited by cystatins in a tight and reversible manner (Fig. 1.9) (Martinez *et al*., 2012). These inhibitors form part of the cystatin superfamily. Cystatins are small protein molecules of 11 to 16 kDa, but some can have a mass of 23 kDa due to a carboxy-terminal extension. Cystatins will regulate protein turnover during plant growth and development. The rice cystatin, oryzacystain 1 (*OCI*), is the best characterised cystatin. *OCI* contains 102 amino acids and in its tertiary structure it has an α-helix and a five stranded anti-parallel β-sheet that contains no disulphide bonds with a conserved QXVXG motif that is needed for cysteine protease inhibition (Benchabane *et al.*, 2010). Some cystatins are also able to inhibit C13 cysteine proteases activity if it has a c-terminal extension with a SNSL amino acid motif (Martinez *et al*., 2007).

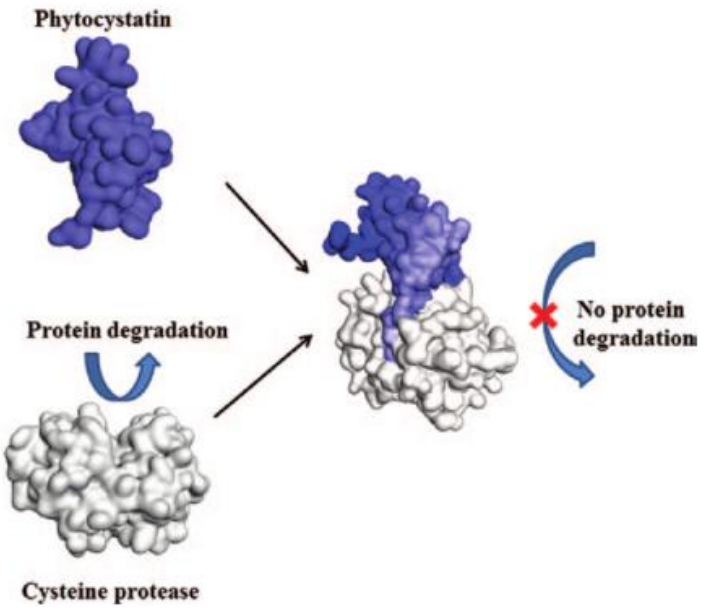


Fig. 1.9: Formation of a cysteine protease-inhibitor complex to inhibit protein degradation (Kunert *et al*., 2015).

Cystatins play an important role in protein regulation during stress-induced senescence as their expression decreases so that cysteine proteases activity are able to increase allowing protein remobilization to occur (Benchabane *et al*., 2010). Cystatins can be used for housekeeping purposes and physiological regulation. Housekeeping cystatins will then have a broad expression pattern range, whereas cystatins involved in development of stress responses have a more restricted expression pattern (Massonneau *et al*., 2005). Although several functions of cystatins have been proposed all of these highlight a balanced interplay between cysteine proteases and proteolytic activity (Grudkowska and Zagdanska 2004). One such example is during seed development where cystatins will protect deposited protein by accumulating in expression to inhibit cysteine proteases and preventing protein remobilization (Benchabane *et al.*, 2010). Tajima *et al*. (2011) illustrated a detailed analysis of C1 cysteine protease and a cystatin complex in senescent spinach leaves. Another key function of cystatins is protecting the plant against Coleopteran insects and nematodes by inhibiting cysteine proteases activity which is needed for digestion (Benchabane *et al.*, 2010). The *AtCYS1* cystatin in *Arabidopsis* blocked cell death that was activated by wounding due to pathogens as well as nitric oxide (Belenghi *et al*., 2003).

Using plant cystatins to regulate cysteine protease expression and activity during biotic and abiotic stress, developmental senescence and PCD have been proposed to improve agronomical important traits (Kunert *et al*., 2015). When ectopically expressing *OCI* in tobacco, the growth and development of stems and leaves were slowed down (Van der Vyver *et al*., 2003). This has also been observed in soybean and *Arabidopsis* (Quain *et al*., 2014). Plants also proved to be more tolerant to abiotic stress such as drought stress, heat and chilling stress and even high and low salt stresses (Prins *et al*., 2008; Demirevska *et al*., 2010; Quain *et al*., 2014). In cowpeas, cystatins accumulated in both drought sensitive and drought tolerant cultivars. However, this

accumulation happened more rapidly in the tolerant cultivars suggesting toward cystatins function in drought tolerance (Diop *et al*., 2004). In lupins cystatins were seen to be also drought responsive but their expression increases even more after rehydration suggesting their involvement in recovery after drought stress (Pinheiro *et al*., 2005).

Various research groups found that cystatins are also involved in nodulation (O'Rourke *et al*., 2014; Van Wyk *et al*., 2014). Van Wyk *et al*. (2014) identified 19 cystatins that were expressed in the soybean root nodules. Three of these showed to be significantly up-regulated during natural senescence. Quain *et al*. (2015) showed more nodules are present although they are smaller when *OCI* is overexpressed in soybean. A recent study identified seven non nodule specific cystatin genes involved in nodulation, nodule development and nodule senescence (Yuan *et al*., 2016). The cystatin, *Glyma.15G227500* were found to be involved in nodulation (Yuan *et al*., 2016). Keyster *et al*. (2013) identified four soybean cystatins which were also present in the nodules that are sensitive to applied nitric oxide and a nitric oxide synthase inhibitor, which is associated with drought stress. However, how cystatins expression is regulated in drought stressed root nodules is still unclear and their ability to be used as a molecular marker for drought tolerance should be investigated.

1.2 PROBLEM STATEMENT

Recent advances in understanding how drought effects soybean production has primarily been focussed on above ground traits such as flowering and seed production (Kunert *et al*., 2016). However, drought stress does not only affect soybean growth but also affects symbiotic nitrogen fixation. A decrease in SNF leads to reduced yields and an increasing need for applications of extra nitrogen fertilizer. The production and selection of drought-tolerant cultivars is needed to improve food security (Ku *et al*., 2013). Due to the difficulty of natural selection and classical breeding approaches under drought stress conditions, marker-assisted breeding could lead to cultivars with improved yield. Marker-assisted breeding can assist in enhancing difficult to measure below ground characteristics, such as prolonged nitrogen fixation. Biotechnology strategies, such as investigating the soybean transcriptome under drought stress conditions, has been used to identify possible drought molecular markers (Manavalan *et al*., 2009; Ku *et al*., 2013). Unfortunately this has not been done on root nodules.

The involvement of proteases, specifically cysteine proteases, and their inhibitors in protein remobilization in nodules, have been studied by various groups. C1 cysteine proteases have known functions in root nodules, bacterial symbiosomes and in the process of nitrogen fixation (Van de Velde *et al*., 2006). Quain *et al*. (2014) also showed that the inhibition of C1 cysteine protease activity by increasing nodulation improved drought tolerance in soybean. However, even though the nodule transcriptome has been investigated during nodule development and early senescence (Van Wyk *et al*., 2014), proteases and their inhibitors' expression under drought stress conditions have not been studied. Whether or not drought stress induces specific cysteine proteases as well as the method how these proteases are activated, prior to the activation of developmental senescence is unknown. This knowledge is needed to possibly

identify drought molecular markers in the nodules protease transcriptome to establish gene expression profiles.

This leads to the establishment of the working hypothesis of this PhD: that genes of the C1 and C13 cysteine proteases and cystatins are involved in premature soybean root nodule senescence caused by drought stress. Further, that an increase in C1 and C13 proteases and cystatins transcript levels leads to a decrease in symbiotic nitrogen fixation. Also, that these induced proteases are possible candidates for drought molecular markers and drought selection. Lastly, that C13 cysteine proteases are involved in the activation of C1 cysteine proteases which could possibly which leads to the premature senescence of root nodules.

1.3 AIMS AND OBJECTIVE

The aim of this study was to advance our knowledge of the expression and function of drought-induced proteases and cystatins found in soybean root nodules. Ultimately the knowledge gained by this study will identify possible drought molecular markers which can be used to improve root nodule lifespan and functionality in drought stressed periods to withstand induced senescence. To achieve this aim, this PhD study had the objectives of i) designing a drought trial which will induce premature senescence in soybean and its root nodules; ii) thereafter characterizing gene expression profile changes in the drought stressed root nodules by using RNA-sequencing (RNA-Seq) as a gene discovery technique to identify possible candidate genes for drought tolerance; iii) to identify all the members of the cysteine and cystatins families through homology searches; iv) analyse the expression of the identified drought-induced proteases and cystatins over different degrees of drought stress as well as to investigate

if they are able to recover after the onset of premature senescence; v) and lastly to investigate if C13 cysteine proteases could be responsible for the activation and increase in activity of C1 cysteine proteases during drought stress.

CHAPTER 2

Growth and development of soybean and crown root nodules experiencing drought stress.

ABSTRACT

Drought stress causes a severe decline in soybeans' (*Glycine max*) symbiotic nitrogen fixation and yield. To possibly identify changes in gene regulation involved in the drought-induced senescence process through a gene expression profile analysis, a drought experiment, representative of drought in a natural environment, were designed to induce different levels of stress in a controlled environment. A potted drought trial with three levels of drought stress measured in vermiculite water content (VWC) at 60 %, 40 % and 30 % were conducted. Other than a decline in growth that could be visually observed, drought stress also led to reduced moisture content of leaves and shoots at 30 % VWC and roots at all the different levels of stress. The vegetative development of plants were also inhibited at 40 % and 30 % VWC. Crown root nodule numbers and water potential were only affected by severe drought stress at 30 % VWC. The function of nodules were negatively affected by drought visible as the colour change in nodules from an active fixating pink/red to an inactive greenish colour. Ureides accumulated in the nodules during early drought but the extent of the severe drought stress was efficient to inhibit ureide formation and accumulation in the nodules. It was concluded that the drought stress levels of 40 % and 30 % VWC were sufficient to initiate induced senescence as it showed physiological differences. The 60 % drought stress level is still a valuable level to add for gene induction studies.

2.1 INTRODUCTION

Climate predictions indicate that in the near future more frequent and severe weather conditions, with possible significant higher temperature and extended periods of drought, will be experienced (Zinta *et al*., 2014). World-wide more than 40 % of crop losses can be attributed to drought stress conditions (Manavalan *et al*., 2009). Changes in climatic conditions could also cause severe soybean yield losses (Ku *et al*., 2013; Daryanto *et al*., 2015).

Different growth parameters can be used to measure the effect of different levels of drought stress for example nodule number and water use efficiency of different organs (Fenta *et al*., 2011) as well as the vegetative growth development (Van Heerden *et al*., 2004). Water potential is a dependable performance indicator during drought stress studies which can give an indication of the plants' drought tolerance and should be used as a selection indicator (Jongdee *et al*., 2002; Siddique *et al*., 2000). Extended periods of drought not only affects natural plant growth and development but also lower nodule water potential, decreases nitrogenase activity and ultimately affects nodule formation and life span (Fernandez-Luquen *et al*., 2008; Gil-Quintana *et al*,. 2015).

Drought has various effects on symbiotic nitrogen fixation (SNF) such as decreasing *Rhizobia* bacteria in the soil leading to an inhibition of nodule formation and leghemoglobin which is needed for oxygen supply to the simbiosome (Fenta *et al*., 2011). One product of SNF in temperate legumes like soybean plants is the nitrogenous compound known as ureides (allantoin and allantoate) (Márquez-García e*t al*., 2015). These compounds are important signals in a plants' stress response (Marino *et al*., 2007). Studies have documented both the accumulation of ureides in nodules (Sinclair and Serraj, 1995; Serraj *et al*., 1999) as well as a

decrease in nodule ureide content (King and Purcell, 2005; Todd *et al*., 2006; Ladrera *et al*., 2007). It has been speculated that a decrease of nodule activity due to ureides inhibition can occur as a feedback mechanism from nodules, or directly from the shoots (Serraj *et al*. 2001).

The objective of this part of the study was to investigate the effect that different levels of drought stress had on different growth parameters of soybean plants and particularly on root nodules. Whether or not the levels of drought were sufficient was important before nodule transcriptome sequencing could continue to analyse gene expression profiles.

2.2 MATERIALS AND METHODS

2.2.1 Plant material and quantification of vegetative development

Soybean (*Glycine max* L.) seeds, cultivar: Prima 2000 (Pannar Seed, South Africa), were planted in pots (17.5 cm x 20 cm diameter) containing fine-grade vermiculite (Mandoval PC, South Africa). Individual seeds were each treated with 0.5 g of commercial *Bradyrhizobium japonicum* inoculum of the WB 74-1 strain (SoyGro Bio-Fertiliser Limited). Plants were grown under natural light conditions, extended with artificial metal-halide lights for up to a 13 hours (h) photoperiod at 600 µmol $m^{-2}s^{-1}$ photosynthetically active radiations (PAR). The greenhouse temperature was set to 27 °C/ 25 °C day/night temperature and humidity was maintained at 60 %. Plants were watered twice a week with de-ionised water and once a week with a nitrogen-free Hoagland nutrient solution (Fenta *et al*., 2012). This watering regime stimulated optimum growth for root nodules responsible for symbiotic fixation (Van Heerden *et al*., 2007).

All plants were grown until they reached the same vegetative growth stage (plastochron index of 3.6) as described by (Erickson and Michelini, 1957), using 25 mm as the reference lamina length. The plastochron index was calculated as follows: Plastochron index = n + ($\log L_n$ - logR)/($\log L_n - \log L_{n+1}$), where n is the youngest trifoliate leaf which is longer than the reference value of R=25 mm counting acropetally from the cotyledonary node. L_n and L_n+1 are the lengths of the trifoliate leaves in mm of n and n+1. To produce less error, only the central pinna was measured from the base to the tip (Handa and Yong Son, 1974).

Thereafter drought stress was initiated by refraining from watering the plants. Pots, together with the vermiculite surface, were covered with a plastic bag to ensure that water loss only occurred through transpiration (Fig. 2.1). The vermiculite water content (VWC) was lowered to 60 %, 40 % and 30 % by withholding both water and Hoagland solution. The respective VWC was calculated as follows: VWC = (fresh mass-dry mass)/fresh mass) X 100. The dry mass used in each pot was 300 g of dry vermiculite. When the desired VWC (60 %, 40 % and 30 %) was reached, plants were kept at the desired VWC for five days before plants and nodules were harvested on the fifth day for further analysis. Each drought stress (DS) treatment had its own control, which will be called DNon-stressed (DNS) hereafter. Although this experimental design added an age-factor to the three treatments, it was considered the most realistic replication of a natural drought progression in the field. The ages of the plants at harvest were 56 days (60 % VWC), 67 days (40 % VWC) and 75 days (30 % VWC) after sowing.

Five plants' organs (young leaves, old leaves, root tips, shoots crown and lateral nodules) (Fig. 2.2) were flash frozen with liquid nitrogen and six plants' samples from all five organs from both harvests were used in the subsequent growth analysis. The two youngest trifoliates were used to represent young leaves and the remaining trifoliates were used to represent old leaves.

Fig. 2.1: A plastic bag was used at initiation of drought stress, to cover the vermiculite surface, to ensure that water loss only occurred through transpiration.

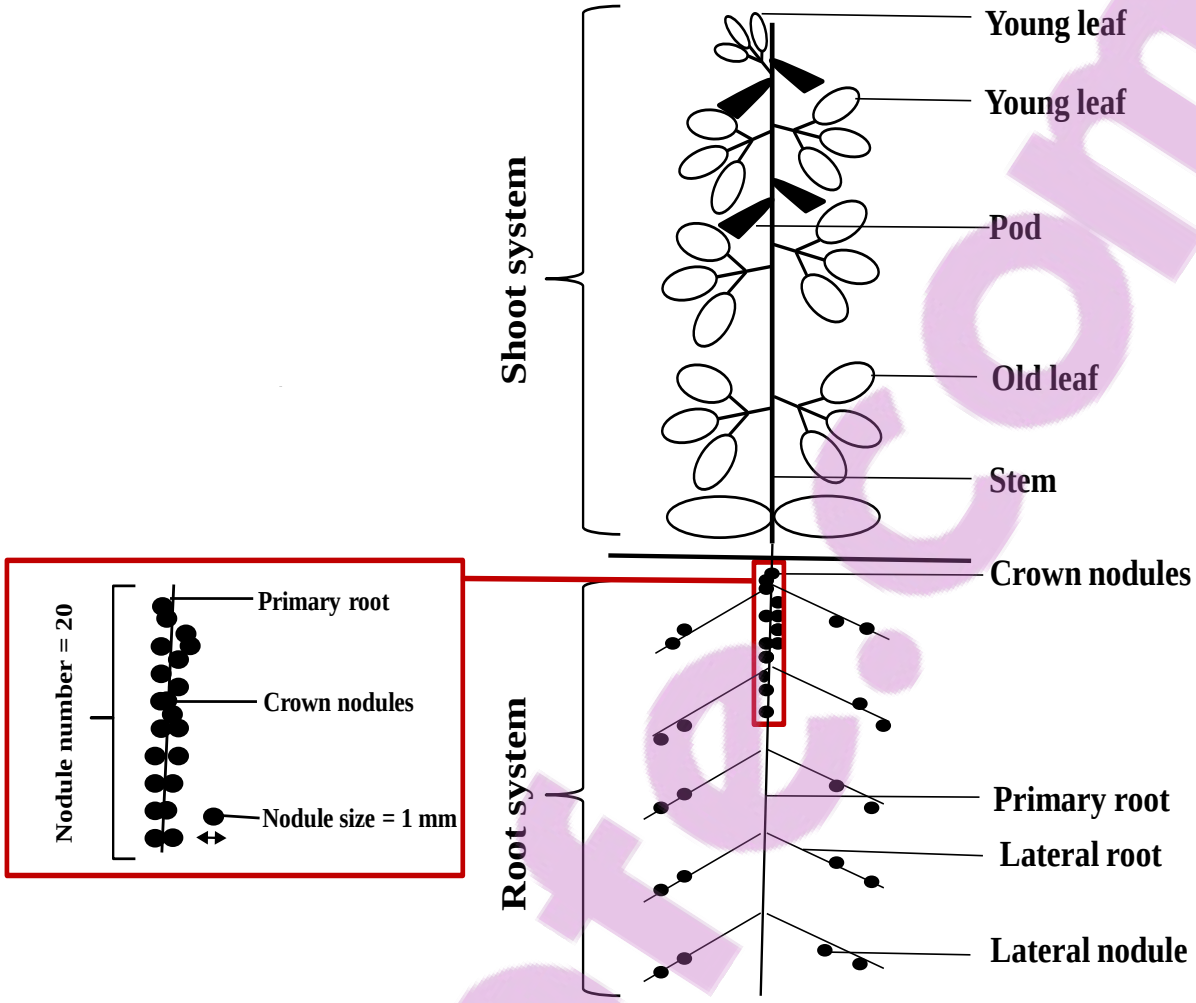


Fig. 2.2: Morphology of a soybean plant showing the above ground shoots system and the below ground root system with different types of nodules. An average of 20 nodules (1 mm in diameter) can be seen DNon-stressed plants.

2.2.1.1 Moisture content

The fresh mass (FM) of the two youngest trifoliate leaves (not fully expanded), remaining trifoliate leaves, roots, shoots, crown and lateral nodules were determined by using a AB104-5 Mettler-Tolendo balance. The dry mass (DM) was then determined by drying all samples in an oven at 60 ºC for approximately 48 h until a constant mass were obtained. Fresh and dry mass can be seen in Appendix A, Tables A.1-2. The moisture content was then determined on a wet basis as follows: Moisture content = (Fresh mass-Dry mass)/Fresh mass X 100.

2.2.1.2 Nodule number, leaf and nodule water potential

Leaf water potential was determined using a pressure bomb (Model 3005 ITC International Australia), while crown nodules (nodules found on the tap root) water potential were determined using a WP4 Dew Point Potential meter (Decagon, USA).

Leaf water potential (Ψ_{Leaf}) was determined pre-dawn on the day of harvest using DS and DNS samples. The central pinna of the third trifoliate leaf counting acropetally from the cotyledonary node was used by placing the cut end of the stem protruding through the specimen holder. Nitrogen gas was then applied to the chamber until a droplet of sap could be observed on the stem (Fig. 2.3). The pressure that is required to produce a sap droplet is equivalent to the force of absorption and capillary with which water is held to plant tissue (Valenzuela-Vazquez *et al*., 1997).

Crown nodules were collected by hand and counted. Nodule water potential (Ψ_{Nod}) was determined immediately after the harvest commenced at 09:00, using 0.1 g of crown nodules together with a WP4 Dew Point Potential meter (Decagon, USA) as described by Guerin *et al.*

(1990). A pre-drought stress measurement was taken on both nodule number and nodule water potential.

2.2.2 Nodule ureide content

To measure the effect that drought stress had on biological nitrogen fixation, the ureide (allantoin and allantoic acid) content of nodules was determined. After determining the weight of nodule tissue, ureides were extracted with 100 µl of 0.2 M NaOH. Samples were then boiled for 20 min to convert allantion to allantoic acid. Samples were cooled and centrifuged at 10 000 g for 10 min where after 5 µl of the supernatant together with 35 µl of H_2O were used for further analysis according to Young and Conway (1942). The diluted plant extract (40 µl) was boiled together with 8 µl of 0.5 M NaOH for another 10 min where after16 µl of a mixture of a 1:1 ratio of 0.33 % Phenylhydrazine (Sigma, USA) and 0.65 M HCl was added and boiled for another two min. A 40 µl solution of 1.67 % pottasium ferricyanide (Sigma, USA) and HCL (36.5-38.0%, used for molecular biology) were incubated together with the plant mixture for 10 min before the absorbance was read at 525 nm. A standard curve was set up by using 1µg, 2 µg, 4 µg, 6 µg and 8 µg of allantion (Sigma, USA) (Appendix A, Fig. A.1).

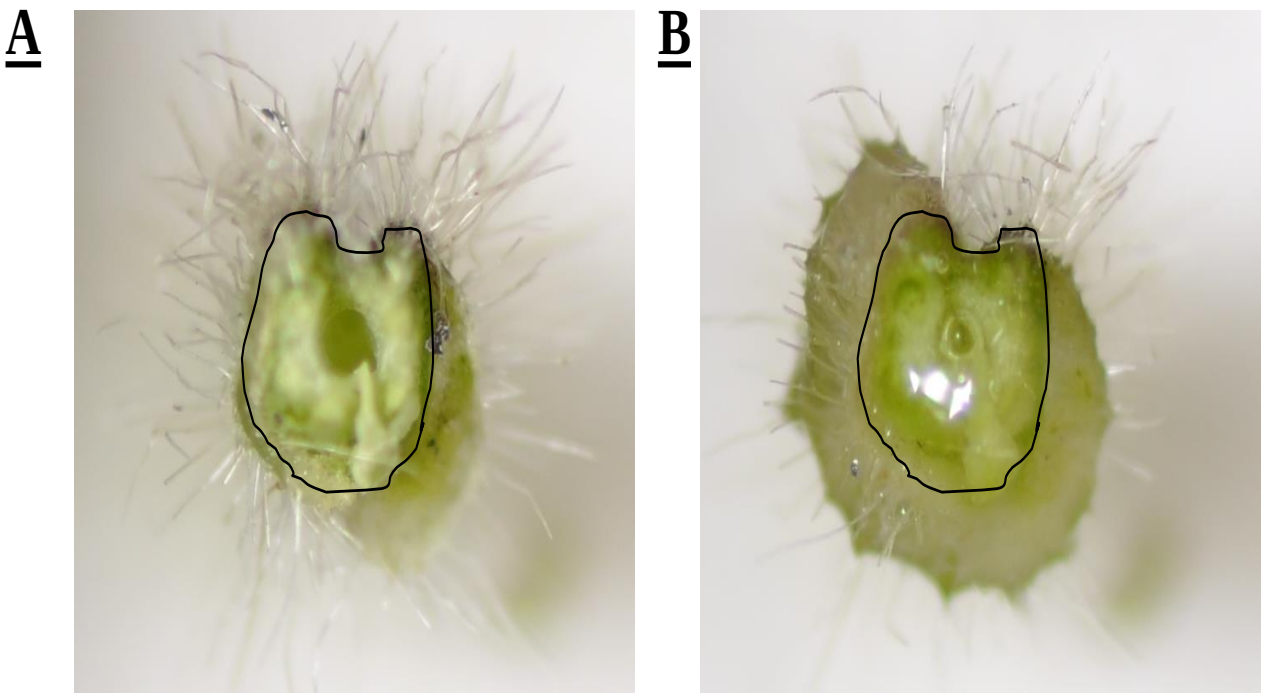


Fig. 2.3: Cut stem protruding from pressure bomb chamber. A) Before nitrogen gas was applied and B) after nitrogen gas was applied producing a water droplet. The line drawn indicates the outline of the stem.

2.2.3 Statistical analysis

To determine statistical significant changes during soybean root nodule development and during soybean gene expression studies, a one-way analysis of variance (ANOVA) was conducted across treatments (DNS (60 %) *vs* DS 60 %) and within a treatment (DS 60 % *vs* DS 40 %, *vs* DS 30 %), if results showed a normal distribution over residual. An example of measurements that has a normal distribution can been seen in Fig. 2.4. This was followed by a Duncan post-test. The SPSS© Version 23 and IBM © Software was used for the statistical analysis. A P-value of $P \leq 0.05$ were seen as significantly different.

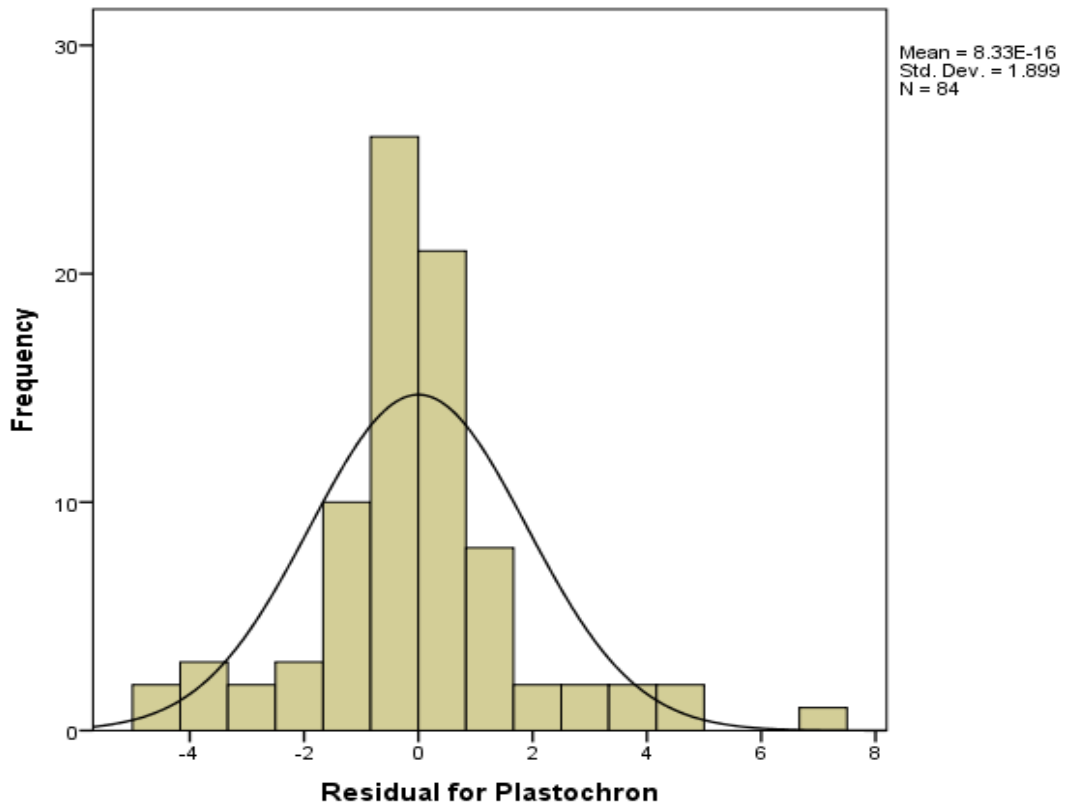


Fig. 2.4: A Histogram of the residuals of the plastochron index was found to be consistent with the normal distributions.

2.3 RESULTS

2.3.1 Plant growth and quantification of vegetative development

A first set of experiments was conducted to quantify the level of stress according to growth and physiological characteristics of soybean plants and crown nodules subjected to different levels of drought stress (60 %, 40 % and 30 % VWC). The greatest effect on plant growth occurred when plants were grown for five days at 30 % VWC (Fig. 2.5 A). This effect on growth could be visually observed, as plants that were drought stressed were noticeably smaller. This visible stunting of growth in DS plants could be quantitatively confirmed by measuring the vegetative development. A significant difference ($P \leq 0.05$) in DS plants compared to it corresponding DNS control (Fig. 2.6) was seen in the vegetative development of drought stress treatments of 40 % and 30 % VWC. No difference in vegetative development was seen in DS plants at 40 % and 30 % VWC with a plastochron index of 7.64 ± 0.19 and 7.80 ± 0.20 respectively. The DNS control showed a significant increases ($P \leq 0.05$) over time in vegetative development with a plastochron index of 6.46 ± 0.19, 11.52 ± 0.97 and 16.36 ± 0.77 at 60 %, 40 % and 30 % VWC respectively (Fig. 2.6). No significant difference was observed in vegetative growth in DS samples at 40 % VWC and 30 % VWC. A significant difference ($P \leq 0.05$) was observed between 60 % VWC compared to 40 % VWC and 30 % VWC after water deficit stress was induced (Plastochron index = 3.6). DS samples at 60 %, 40 % and 30 % were respectively 6.14 ± 0.09, 7.64 ± 0.19 and 7.80 ± 0.20.

Further, crown nodules were collected by hand and dissected to investigate the colour of the nodules (Fig. 2.5 B). It was observed that the internal nodule colour changed from red/pink (indicating active nitrogen fixation) to brown/greenish colour under drought conditions. This visual colour change under drought indicated an inactivation of nodules and degradation of

leghemoglobin and a loss of nitrogen fixation ability. The colour change in DS nodules was more severe than in DNS samples caused by natural senescence.

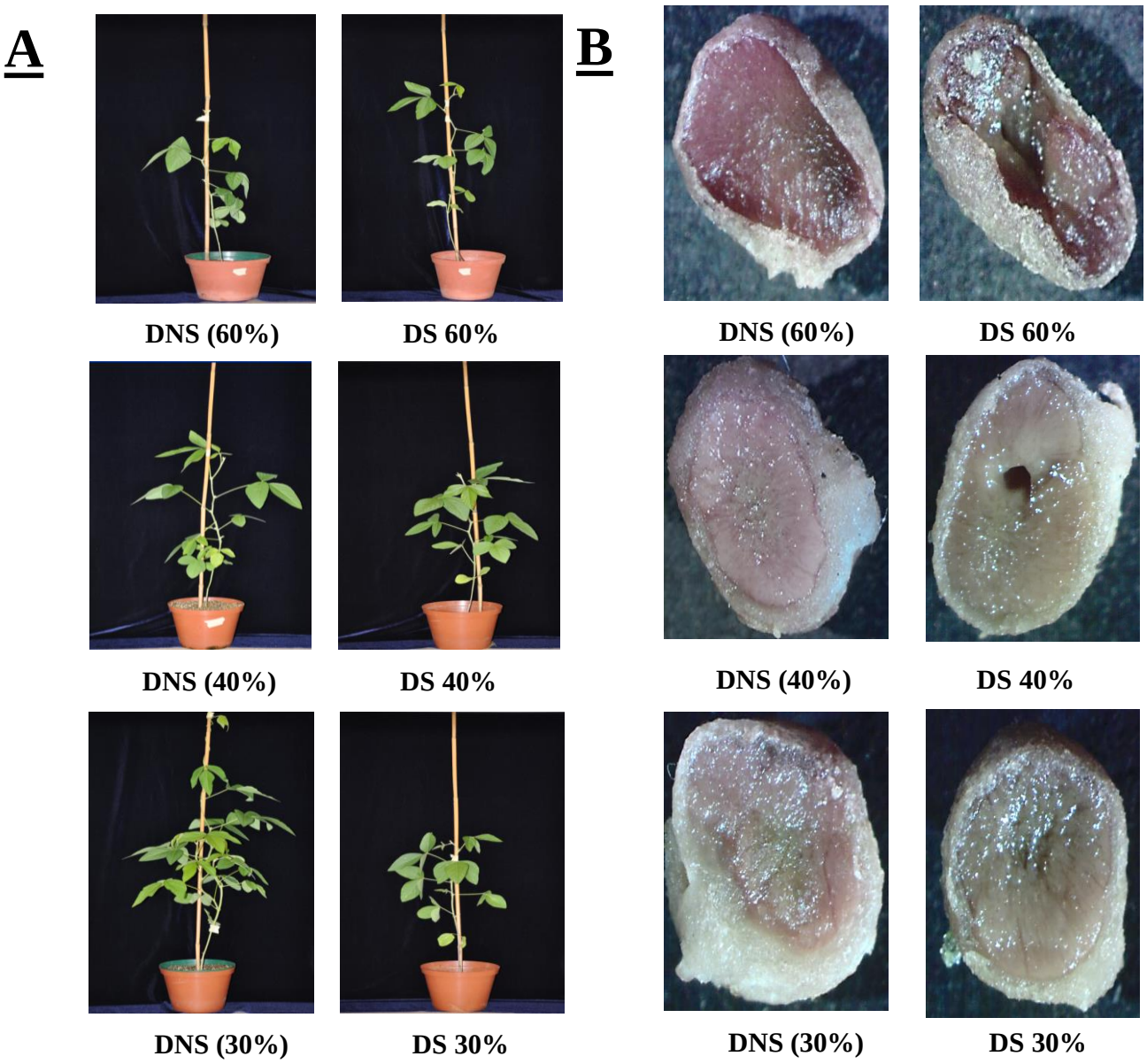


Fig. 2.5: (A) Soybean growth under DNon-stressed and drought stressed conditions applied as different percentage of VWC. (B) Crown nodule cross sections, of DNS and DS plants exposed to different levels of drought (60 %, 40 % and 30 % VWC). The plants age after sowing was 56 days (60 % VWC), 67 days (40 % VWC) and 75 days (30 %) at harvest.

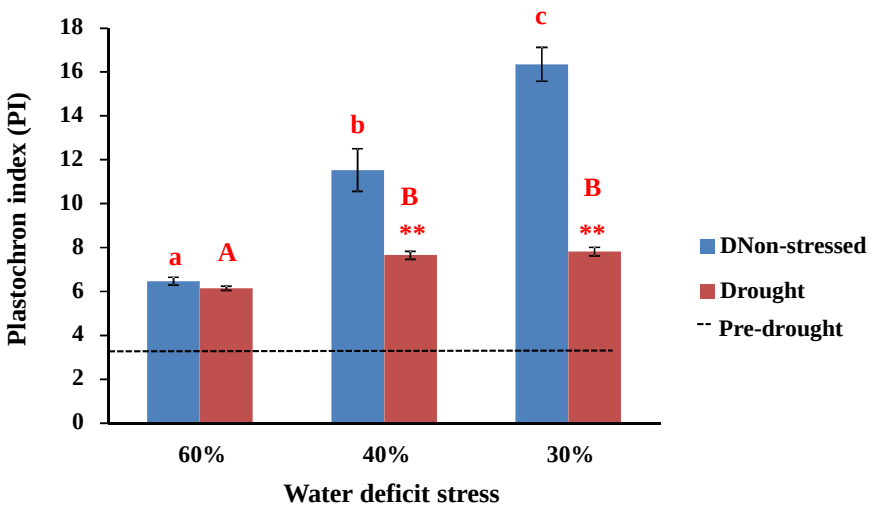


Fig. 2.6: Plastocron index (PI) of plants under DNon-stressed conditions and after exposure of different levels of drought stress (60%, 40% and 30% VWC). Pre-drought plants had a Plastochron index of 3.60. Data represents the mean ± SE of the PI from 14 individual plants.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40 % *vs* DS 30 %) were compared to each other.

2.3.2 Nodule number and moisture content of soybean plant organs

The development and senescence of root nodules was of particular interest to this study. The effect of drought on crown nodules development and growth were measured. Drought conditions prevented any further nodule formation (Fig. 2.7 A) and an increase in nodule numbers at different levels of drought stress (11.00 ± 1.36 60 % VWC, 9.66 ± 1.05 40 % VWC and 9.00 ± 1.26 30 % VWC) were similar to the number of nodules before drought stress was initiated (10.2 ± 2.1). In contrast, in DNS plants (16.00 ± 2.94 40% VWC and 17.33 ± 2.43 30 % VWC) the number of crown nodules increased significantly ($P \leq 0.05$) over time when compared to the number of nodules measured at pre-drought (Fig. 2.7 A). Plants subjected to 30 % VWC DS showed a significantly lower number of nodules ($P \leq 0.05$) compared to its DNS same age control.

Nodules of plants grown at 30 % VWC proved to be the most affected by drought stress as seen by a significant ($P \leq 0.05$) decline in moisture content (56.84 % ± 1.36 %) when compared to the other levels of DS (73.94 % ± 6.6 % at 60 % VWC and 74.17 % ± 2.05 % at 40 % VWC, Fig. 2.7 B). A significant decline ($P \leq 0.05$) was also observed in moisture content of nodules in both 40 % and 30 % VWC when each treatment was compared to its DNS control. The moisture content of DS nodules at 30 % also showed a significant decline ($P \leq 0.05$) compared to nodules before drought stress was initiated (80.61 % ± 3.53 %). Fresh and dry mass measurements are provided in Appendix A, Tables A.1-2.

The moisture content of young leaves, old leaves and shoots (Fig. 2.8 and Fig. 2.9B) decreased significantly in 30 % VWC when compared to its respective DNS control and over the different DS treatments (60 %, 40 % and 30 % VWC). The moisture content of young leaves and shoots

drought stressed at 30 % VWC were also significantly lower from pre-drought young leaves and shoots (83.12 % ± 2.08 % and 79.65 % ± 1.19 %). The moisture content of the roots (Fig. 2.9 A) on the other hand showed significant changes ($P \leq 0.05$) in all three DS treatments compared to the respective DNS controls. Only the 30 % VWC treatment showed a significant difference ($P \leq 0.05$) when the different DS treatments (60 %, 40 % and 30 % VWC) were compared to each other. Both 40 % and 30 % VWC stressed roots were significantly lower ($P \leq 0.05$) to the pre-drought moisture content of roots (90.25 % ± 2.45 %).

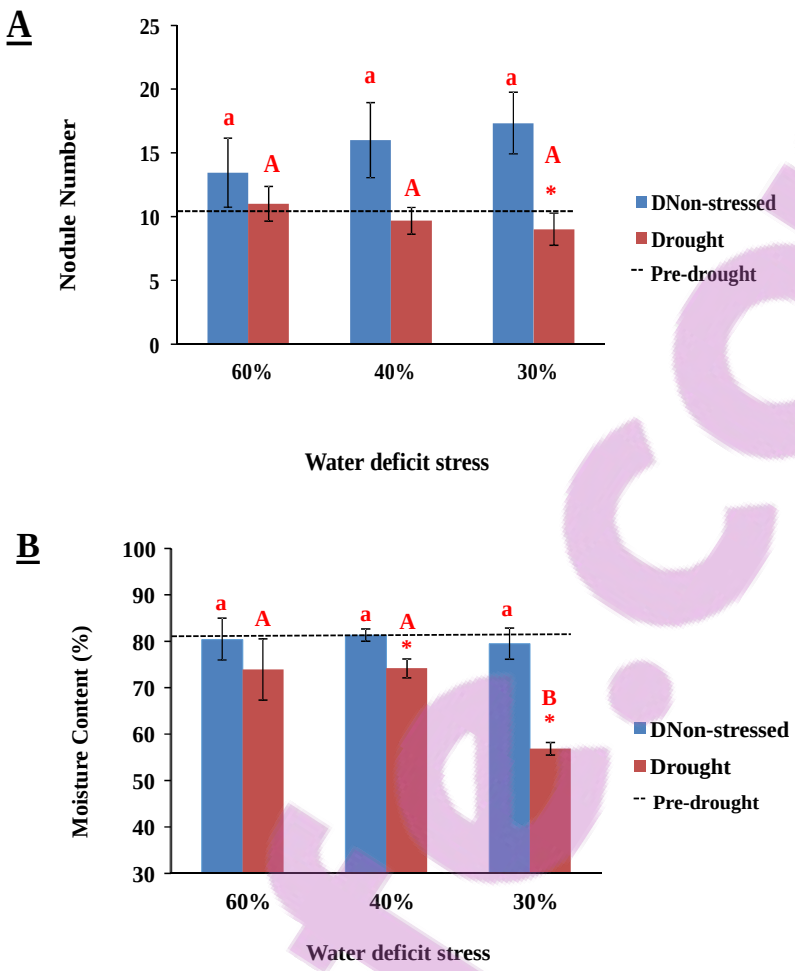


Fig. 2.7: A) Crown nodule number under DNon-stressed conditions and after exposure of different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stress nodules had a nodule number of 10.20 ± 2.12. B) Moisture content of DNS crown nodules and after exposure of nodules with different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stress nodules had a moisture content of 80.61 % ± 3.53 %. Data represent the mean ± SE of nodules derived from nine individual plants for nodule number and derived from eight individual plants for nodule moisture content.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40 % *vs* DS 30 %) were compared to each other.

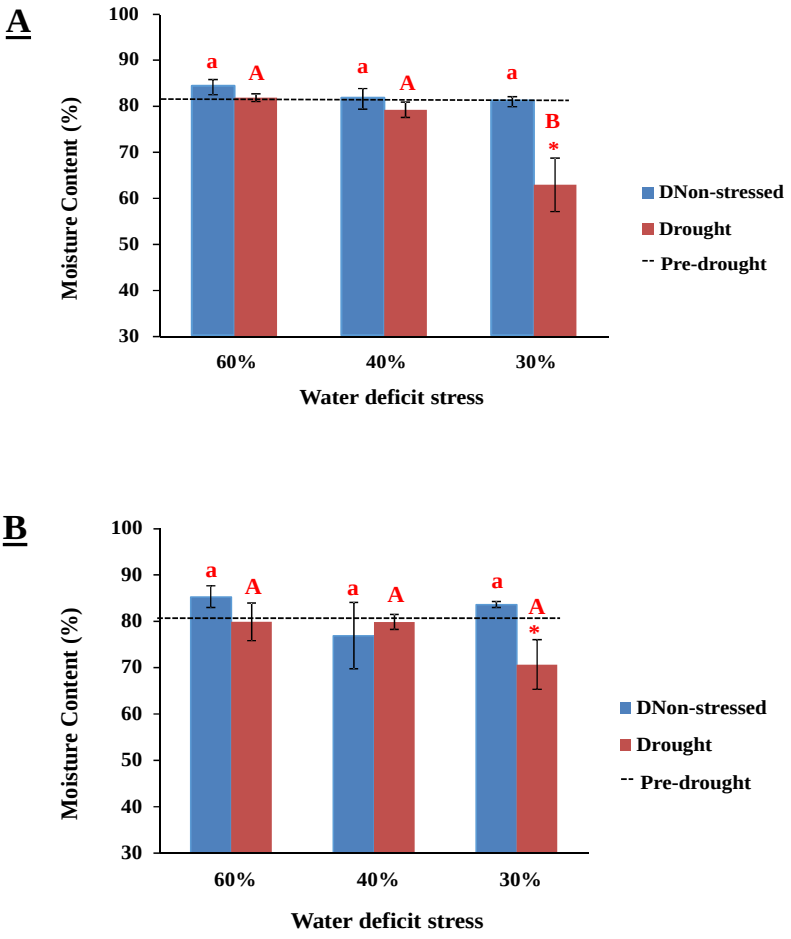


Fig. 2.8: A) Moisture content of young leaves under DNon-stressed conditions and after exposure of different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stressed young leaves had a nodule moisture content of 83.12 % ± 2.08 %. B) Moisture content of old leaves under DNon-stressed conditions and after exposure of different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stress nodules had a moisture content of 80.4 % ± 1.89 %. Data represent the mean ± SE of the moisture content of eight individual plants.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40% *vs* DS 30%) were compared to each other.

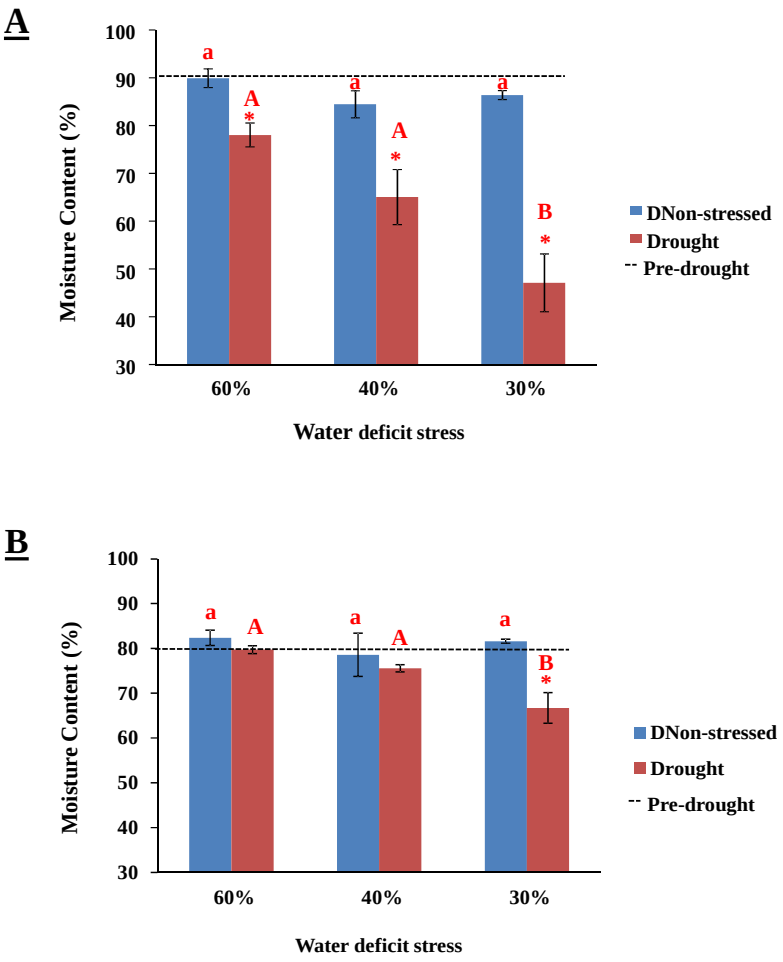


Fig. 2.9: A) Moisture content of roots under DNon-stressed conditions and after exposure different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stressed roots had a moisture content of 90.25 % ± 2.45 %. B) Moisture content of shoots under DNon-stressed conditions and after exposure of different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stressed shoots had a moisture content of 79.65 % ± 1.2 %. Pre-drought stress shoots had a moisture content of 79.65 % ± 1.19 %. Data represent the mean ± SE of the moisture content of eight individual plants.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40 % *vs* DS 30 %) were compared to each other.

2.3.2 Nodule and leaf water potential

Leaf and nodule water potential was measured to examine whether the levels of drought were sufficient. The 30 % VWC stress treatment affected leaf water potential the most ($P \leq 0.05$) with a leaf water potential of -0.63 MPa ± 0.08 MPa compared to its respective DNS control with a Ψ_{Leaf} of -0.41 MPa ± 0.04 MPa (Fig. 2.10 A). When water potential of leaves at 30 % VWC was compared to other drought stress treatments (60 % and 40 % VWC), it also showed to be significantly lower (Fig. 2.10 B). However, only the 30 % VWC DS treatment were significantly ($P \leq 0.05$) lower to its pre-drought samples at (-0.31 MPa ± 0.02 MPa).

The nodule water potential was lower in DS nodules when compared to their respective DNS controls (Fig. 2.10 B). Water potential of DS nodules at 30 % VWC (-1.03 MPa ± 0.20 MPa) was thereby significantly lower ($P \leq 0.05$) than in the both the respective DNS same age control nodules (-0.42 MPa ± 0.08 MPa) and when comparing them to the initial levels before stress were initiated. Both drought stress treatments of 40 % and 30 % VWC were significantly lower from their pre-drought control at -0.41 MPa ± 0.74 MPa.

2.3.3 Nodule ureide content

Ureides are used as a major source of nitrogen in plants and is the main product of symbiotic nitrogen fixation. Initially there was a significant increase ($P \leq 0.05$) in ureides in DS nodules (8.45 μg g^{-1}FW ± 1.32) at 60 % VWC compared to its same age DNS control (5.08 μg g^{-1}FW ± 1.46) (Fig. 2.11). The ureide content then showed a significant decrease ($P \leq 0.05$) in both 40 % and 30 % VWC stressed nodules (3.81 μg g^{-1}FW ± 0.39, 40 % VWC and 4.75 μg g^{-1}FW ± 0.56, 30 % VWC), compared to 60 % (Fig. 2.11). This was in contrast to the DNS control

which showed no significant difference over time than in the absence of stress, as seen in the DNS controls (Fig. 2.11).

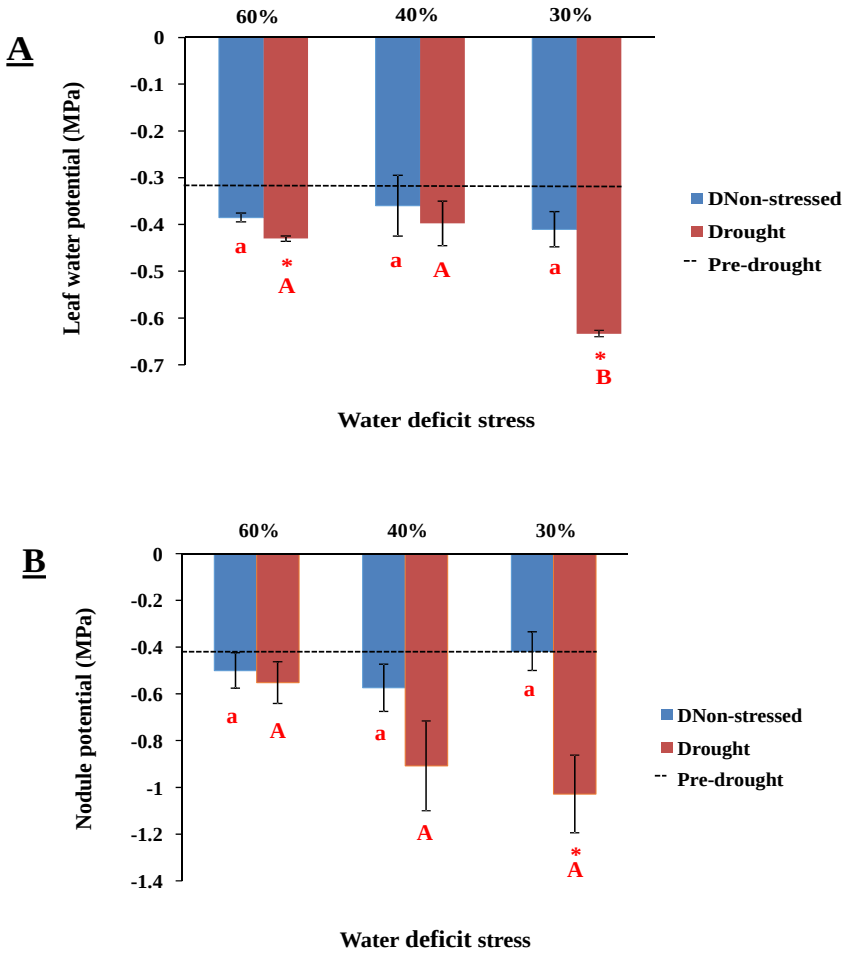


Fig. 2.10: A) Leaf water potential of DNon-stressed plants and after exposure of plants to different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought stressed leaf water potential was -0.31 MPa ± 0.02 MPa B) Nodule water potential of DNon-stressed nodules and after treatment of nodules with different levels of drought stress (60 %, 40 % and 30 % VWC). Pre-drought nodule water potential was -0.42 MPa ± 0.07 MPa. Data represent the mean ± SE of leaves derived from four individual plants for leaf water potential and derived from eight individual plants for nodule water potential.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40 % *vs* DS 30 %) were compared to each other.

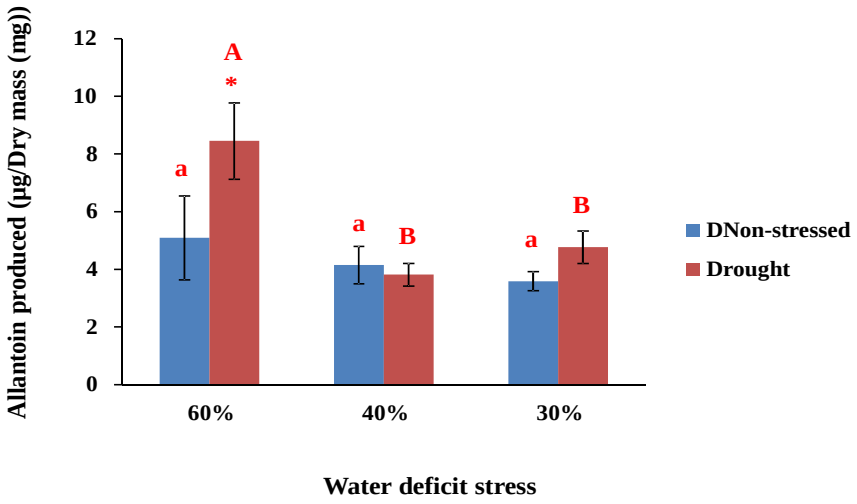


Fig. 2.11: Ureide content of nodules under DNon-stressed conditions and after exposure of different levels of drought stress (60 %, 40 % and 30 % VWC). Data represents the mean ± SE of the ureides content if plants were derived from three individual plants.

* Significant differences between DS treatments and its corresponding DNS controls.

[a-c] Significant differences where the three DNS controls (DNS (60 %) *vs* DNS (40 %) *vs* DNS (30%)) were compared to each other.

[A-C] Significant differences where the three DS treatments (DS 60 % *vs* DS 40 % *vs* DS 30 %) were compared to each other.

2.4 DISCUSSION

Soybean plants were planted in a controlled environment to eliminate the variability which can be experienced in field trials such as light intensity, soil and air temperature, humidity and various biotic stresses (Passioura, 2006). This permitted drought stress to be the only stress experienced by the plants and ensured the repeatability of the trial. Care was taken to ensure that the drought experiment represented a natural environment. Using vermiculite as a growth medium, which allows for adequate aeration, protects the roots from hypoxia (Passioura, 2006), as hypoxia could cause changes in protein levels usually associated with abiotic stress (Seki *et al*., 2002). Distilled water and hoaglands nutrient solution that was used to water plants, were always kept at the same temperature as the green house to prevent an inhibition of growth due to watering with cold water (15 ºC) (Brockwell and Gault, 1976).

Drought stress was initiated at the same time and age on all plants to facilitate the realistic drying of vermiculite over time. This also allowed all plants to be at the same growth stage and nodules to be the same age, eliminating differences in nitrogenase activity due to nodules short life cycle (Alesandrini *et al*., 2003; Puppo *et al*., 2005). This, however, induced a difference in the age of plants at harvest. If drought stress was initiated at different times to allow for a harvest where all the plants were the same age, other factors such as plant drought tolerance along with a difference in nitrogenase activity of nodules would have caused additional variables.

By using plant and nodule growth data as well as water potential measurements, it was verified to what extent drought stress was experienced by the nodules. It was confirmed that the drought treatment affected soybean plant growth, crown root nodule formation and nodule activity.

Plant growth and vegetative development were inhibited and the formation of new nodules repressed due to drought stress. Márquez-García e*t al*. (2015) illustrated that nodule function is not immediately impaired upon exposure to mild drought conditions, but that nodule numbers were decreased in soybean plants subjected to severe drought stress, this was also seen in other studies (Fernandez-Luquen *et al*., 2008). Fenta *et al.* (2012) also saw that three different soybean cultivars had reduced nodule numbers after drought stress was introduced at the third trifoliate leaf stage of development. Visually nodule tissue indicated inactivation, as were seen by a tissue colour change from red to greenish (Puppo *et al.*, 2005; Fenta *et al*., 2011; Fenta *et al*., 2014). Ureides, the product of SNF, could have possibly accumulated in mild drought stressed nodules. Ureides has been seen to accumulate in nodules due to reduced catabolism of foliar ureides during drought stress (Vadez and Sinclair 2000). Nitrogenase activity is suspected to be decreased by a negative feedback due to the accumulation of ureides in nodules (Serraj *et al*., 1999; Van Heerden *et al*., 2008). Nitrogenase activity could have been measured with gas chromatography but this was not possible at the time of harvest. A low nodule water potential and moisture content decreases respiratory capacity of nodules resulting in a decline in nodule permeability (Purcell and Sinclair, 1995). When the cell turgor is lost in the nodule cortex, the microbial partner is eliminated due to limited O_2 diffusion (Guerin *et al*., 1990). The decline of ureides in more severe drought conditions could have been a result of induced accumulation of ureides in the shoots due to protein translocation in senescing organs (De Silva *et al*., 1996) but have to be investigated in the future. The extent of the severe drought stress was sufficient to inhibit ureide formation and accumulation in the nodules. Overall, results from the growth analysis proved that the drought stress levels of 40 % and 30 % VWC were sufficient for RNA isolation from nodule samples to investigate changes in the crown

nodule transcriptome and gene expression profiles. Although 60 % VWC doesn't affect soybean growth it is still a valuable addition to the study to investigate gene expression profiles.

CHAPTER 3

Soybean crown nodule gene expression profile analysis under drought conditions.

ABSTRACT

A transcriptome analysis of soybean crown nodules were conducted to investigate gene expression profiles under varying drought-stress conditions using next generation RNA sequencing (RNA-Seq) technology to identify genes which could possibly be candidate genes for enhanced drought tolerance. A total of 324 genes were significantly up-regulated from their non-stressed same age controls (DNS) and 20 genes were significantly up-regulated across drought stress treatments without an increase in DNS controls. A Defensin-like protein *(Glyma.13G27800*) and a *LEA*-D11 group 2 protein (*Glyma.05G112000*) were the most up-regulated genes as a result of drought stress and with further investigation might be possible candidate genes for drought tolerance. Expression of nodule cysteine proteases and cystatins, involved in the regulation of the bacterial symbiosis and leghemoglobin degradation, increased during nodule senescence. Eight C1 cysteine proteases, three C13 cysteine proteases and four cystatins were induced by drought stress. Expression of one C1 cysteine protease, *Glyma.10G207100,* was highly induced during drought but not under natural nodule senescence. A C1 cysteine protease, *Glyma.14G085800* and a cystatin, *Glyma.05G149800,* will be useful indicators for drought-induced premature senescence of soybean due to their high expression after drought exposure. The C13 cysteine protease *Glyma.05G055700,* increased in expression after drought stress more severely than during natural senescence. Overall, the results identified genes that are associated with premature senescence of soybean root nodules as a consequence of drought.

3.1 INTRODUCTION

Advances in next generation RNA-Sequencing (RNA-Seq) together with the release of the soybean genome (Schumtz *et al*., 2010), makes it possible to study gene expression changes in soybean root nodules after the onset of drought stress. RNA-Seq is a powerful technique to investigate differential gene expression. The soybean genome size is 1, 15 MB (Schumtz *et al*., 2010), and more than 66 000 genes (Findley *et al*., 2010) have been sequenced. Recent transcriptome data revealed that 55 616 transcripts have been annotated in the soybean genome (Libault *et al*., 2010). Using this information will allow us to significantly improve our understanding of the genes involved in drought-induced senescence, resulting in a decline in the symbiotic nitrogen fixation (SNF) ability of the root nodules.

RNA-Seq has been used in recent years to study the transcriptome of different organs of soybean during development and under different abiotic and biotic stress conditions. Severin *et al*. (2010) looked at 14 different tissue types at an age of 20-25 days after seeds were planted in soil inoculated with *Bradyrhizobium japonicum,* to establish a transcriptome atlas of soybean genes. Jones and Vodkin (2013) investigated the soybean seed from fertilization until maturity focussing on genes involved in protein storage. A recent transcriptome study on drought-exposed soybean roots identified expression changes of genes associated with osmo-protectant biosynthesis as well as genes coding for kinases, transcription factors controlling root growth and phosphatase 2C proteins (Ha *et al*., 2015; Song *et al*., 2016). RNA-Seq has also been used to investigate the defence mechanisms of different diseases associated with biotic stress in soybean (Kim *et al*., 2011). Although the nodule transcriptome has been investigated during natural developmental senescence (Van Wyk *et al*., 2014), limited studies have focused on nodule development under stress-induced senescence conditions.

Nodule cysteine proteases, the focus of this study, have shown to play a role in SNF (Van de Velde *et al.*, 2006; Li *et al*., 2008). They also modulate nodule development and senescence with increased protease activities in senescent nodules (Kardailsky and Brewin, 1996, Vorster *et al*., 2013, Van Wyk *et al*., 2014). Márquez-Garcia *et al*. (2015) further reported that drought can enhance the expression of cysteine proteases in soybean crown nodules belonging to the C1 (papain-like) cysteine protease family. These C1 cysteine proteases are synthesised as pre-proproteins and undergo proteolytic processing (Simova-Stoilova *et al.,* 2010) with the help of C13 (legumain-like) cysteine proteases (also referred to vacuolar processing enzymes or VPEs) (Okamoto *et al*., 1995). VPEs are involved in developmental senescence, programmed cell death (PCD) and activation of pre-proteases (Hara-Nishimura *et al*., 2005; Roberts *et al*., 2012). VPEs are further bound to an endogenous cysteine protease inhibitor that is released upon perception of PCD triggers (Ge *et al*., 2016). However, whether VPEs are actively expressed in drought-stressed nodules are still unclear.

Cystatins (cysteine protease inhibitors) form a tight reversible interaction with C1 cysteine proteases (Chu *et al*., 2011). This leads to the regulation of protein turnover during different developmental processes including senescence. The main function of cystatins is to protect the plant against pests (Benchabane *et al*., 2010), but they were also seen to be induced during natural senescence (Van Wyk *et al*., 2014).

Very few studies have been done on the transcriptome of soybean nodules, in particular with a focus on proteases and their inhibitors, and specifically cysteine proteases and cystatins expression. According to the Phytozome database (www.phytozome.net), seven hundred and six C1 cysteine protease and 75 C13 cysteine protease sequences were found in the soybean genome and 300 cystatin-like sequences. Some have been previously identified to vary in

expression in developing soybean nodules (Van Wyk *et al*., 2014). However, a study has not been carried out how drought stress changes the soybean nodule transcriptome.

The objective of this study was to use RNA-Seq to identify gene expression profiles in soybean nodules following drought exposure. Secondly, to investigate which genes in the entire nodule transcriptome's expression was the most effected by drought and not by natural senescence. Thirdly, to identify all the C1 and C13 cysteine proteases expressed in nodules during drought stress and to establish a protease expression profile for cysteine proteases. As a last objective, to examine to what extent these C1 and C13 cysteine proteases and cystatins are drought inducible.

3.2 MATERIALS AND METHODS

3.2.1 RNA extraction and quantification

Crown nodules were harvested from drought stressed (DS) and DNon-stressed plants (DNS) from each of the 60 %, 40 % and 30 % VWC treatments as described in chapter 2 and were immediately flash frozen. For RNA extraction, 100 mg of nodules from three biological replicates were grinded into a fine powder. RNA was extracted applying a Direct-zol™ RNA extraction kit (Zymo Research, USA) together with Tri-Reagent® (Sigma-Aldrich, Germany) following the manufacturer's instructions. RNA quantity was assessed with a Thermo Scientific NanoDrop 2000 applying 1 µl of RNA following the instructions provided by the manufacturer. RNA quality was further assessed on a 1 % (w/v) agarose gel in 1X Tris-acetate-EDTA (TAE) buffer. A 4 µl reaction volume was prepared by adding the RNA sample (3 µl) and 6X RNA Loading dye (1 µl) (Thermo Scientific, USA) containing Gelred™ (Biotium, USA) to stain nucleic acids. The reaction mixture was loaded in the pre-set agarose gel and electrophoresis was carried out at 100 V for 20 min which was then visualised using a ultra-violet light. For RNA-Seq, three biological crown nodule replicates were pooled for each of the three treatments (60 %, 40 % and 30% VWC). These crown nodule samples were then sent to BGI (China) for transcriptome sequencing on an Illumina HiSeq TM 2000 Sequencer. Sample libraries were constructed with the TruSeq RNA library preparation kit.

3.2.2 Data processing, normalization and data mining

RNA-Seq results were analysed using the Galaxy platform [http://galaxy.bi.up.ac.za/] (Bioinformatics Unit, Forestry and Agricultural Biotechnology Institute, University of

Pretoria) employing the Tuxedo pipeline. The reference soybean genome Gmax 275 Wm 82A2.V1 for annotation and mapping of reads was applied directly from the Phytozome database (version 11.0) [https://phytozome.jgi.doe.gov/pz/portal.html] (Schmutz *et al*., 2010). Sequencing results were converted to a Sanger FASTQ format with a FASTQ Groomer (version 1.0.04). RNA-Seq reads quality scores were assessed applying FASTQ Quality Trimmer (Version 1.0.0) (Blankenberg *et al*., 2010, Cock *et al.,* 2010). Trimmed paired reads, each 74 base pairs (bp), were mapped to the reference genome with Tophat2 (version 0.7) (Kim *et al*., 2013). Thereafter, Cufflinks (version 0.0.7) was applied to assemble all the aligned reads to the transcript/exon isoforms (Trapnell *et al*., 2010). Cuffdiff (version 0.0.5) was finally used to find fold changes in transcription time points over the different drought stress treatments (Trapnell *et al*., 2010). The various bioinformatic tools used in the Galaxy pipeline to analyse generated RNA-Seq data can be seen in Fig. 3.1. An alignment summary was made with a Collect Alignment Summary tool (version 1.126.0) to view the percentage of aligned bases over the different treatments. Specific parameters used for each of these tools are listed in Appendix B, Tables B.1- 3. All other programmes used were with the default set of parameters.

FPKM data (Fragments per Kilobase of exon model per Million mapped fragments) generated with an expression value of higher than 1.5 FPKM and a 2X $\log_2$-fold increase or decrease in expression, were analysed further. Differential patterns of gene expression and proteases in root nodules, due to different drought stress conditions (60 %, 40 % and 30 % VWC), were illustrated in a Venn-diagram (Table 3.1) constructed with Venny 2.1 (Oliveros, 2007). Only significant changes in gene expression will be discussed further. The Phytozome, the Merops databases (https://merops.sanger.ac.uk/) and the online resources of NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.,* 1990) were applied to identify genes of interest and cysteine proteases found in both the nodule and soybean genome.

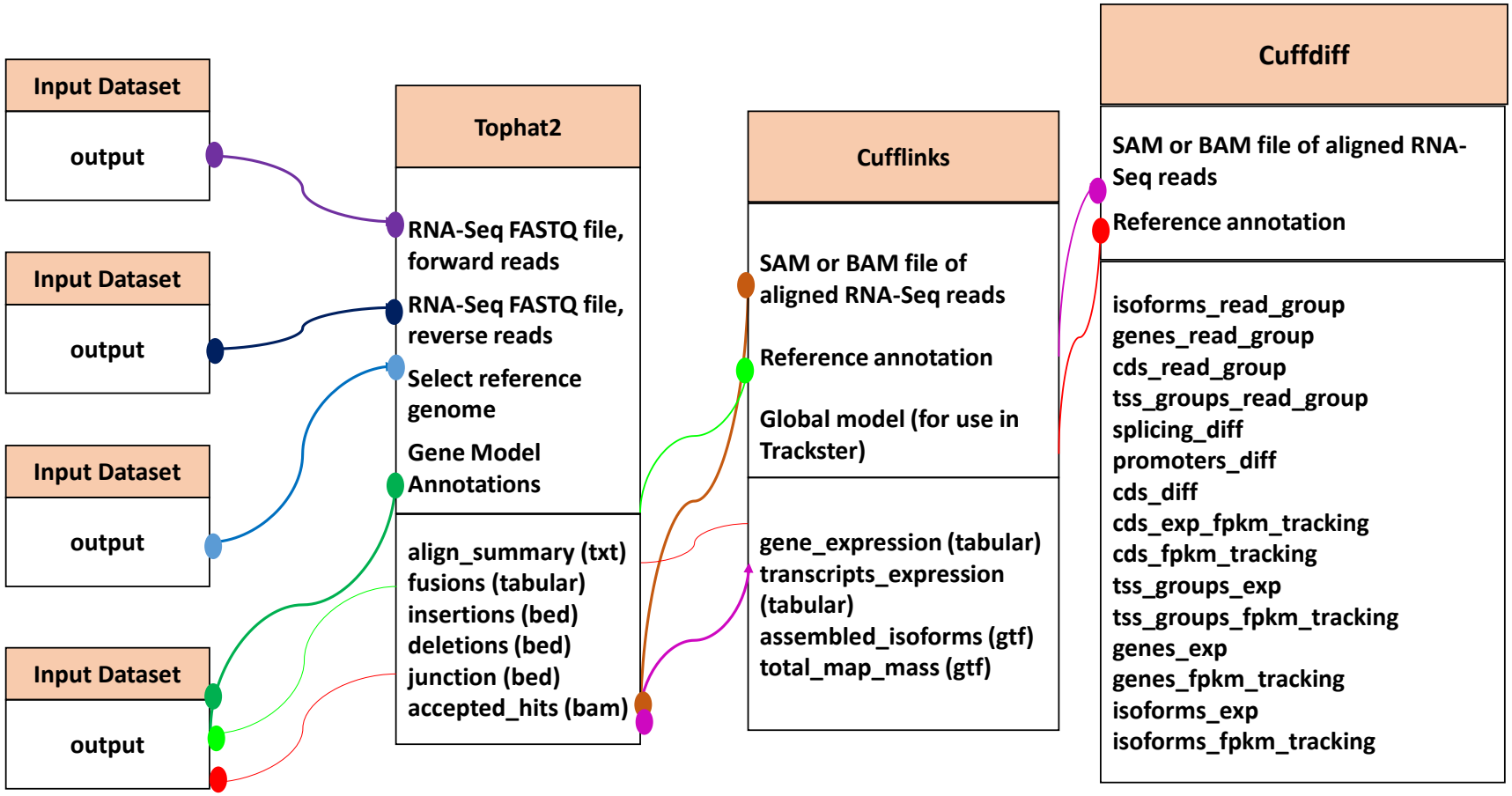


Fig. 3.1: The PostQC Galaxy pipeline applied to map reads to the genome. This pipeline was employed for each of the respective treatments DS 60 % *vs* D 60 %, DS 40 % *vs* D 40 %, DS 30 % *vs* D 30 % and also for D 60 % *vs* D 40 % *vs* D 30 % and DS 60 % *vs* DS 30 % *vs* DS 30 %.

Table 3.1 Comparisons between significant data sets for RNA-Seq analysis

Data set	Gene expression	Comparison	Requirements
1	Increase	DS 60 % *vs* DS 40 % VWC DS 40 % *vs* DS 30 % VWC DS 60 % *vs* DS 30 % VWC	2X log_2-fold change in expression
2	Increase	DS 60 % *vs* DS 40 % VWC DS 40 % *vs* DS 30 % VWC DS 60 % *vs* DS 30 % VWC	2X log_2-fold change in expression No 2X log_2-fold increases in DNS
3	Decrease	DS 60 % *vs* DS 40 % VWC DS 40 % *vs* DS 30 % VWC DS 60 % *vs* DS 30 % VWC	2X log_2-fold change in expression
4	Decrease	DS 60 % *vs* DS 40 % VWC DS 40 % *vs* DS 30 % VWC DS 60 % *vs* DS 30 % VWC	2X log_2-fold change in expression No 2X log_2-fold decreases in DNS
5	Increase	DNS (60 %) *vs* DS 60 % VWC DNS (40 %) *vs* DS 40 % VWC DNS (30 %) *vs* DS 30 % VWC	2X log_2-fold change in expression
6	Decrease	DNS (60 %) *vs* DS 60 % VWC DNS (40 %) *vs* DS 40 % VWC DNS (30 %) *vs* DS 30 % VWC	2X log_2-fold change in expression

3.2.3 cDNA Synthesis

Complimentary deoxyribonucleic acid (cDNA) was synthesized, from four nodules for all three levels of DS and their DNS controls in a 38 µl reaction using the ImProm-II™ Reverse transcription system (Promega, USA). Manufacturer instructions were followed by applying 1 µg of RNA and 100 µM nM Random Hexamer Primers (Thermo Scientific, USA). cDNA synthesis was carried out at 42 °C for 60 min in a thermo cycler prior to inactivation at 70 °C for 5 min. The same method was applied to all samples. Plants used for cDNA synthesis was grown independently in a 2nd replicated experiment, than those used for RNA-Seq and thus served as a second control.

3.2.4 RNA-Seq validation

RNA-Seq validation was done with Real Time RT-qPCR. Primers for the different cysteine proteases as well as cystatins were designed with the IDT's PrimerQuest Design Tool [http://eu.idtdna.com/PrimerQuest/Home/Index]. Primer sequences and amplicon product sizes can be found in Appendix B, Tables B.10. All amplicons were sequenced at the Bioinformatics and Computational Biology Unit at the University of Pretoria (www.bi.up.ac.za/seqlab/). Thermo cycling was carried out with the Bio-Rad CX5 Thermo cycler (Bio-Rad, USA) and a KAPA SYBR® FAST qPCR Kit (Kapa Biosystems, USA). Reactions were set up at 95 °C for 10 min followed by cycling at 95 °C for 15 sec, 60 °C for 30 sec and 72 °C for 30 sec over 39 cycles. Melting curves (75 °C - 95 °C) confirmed the specificity of PCR amplicons. Amplicons were then sequenced and BLAST searched on Phytozome to confirm the sequence of amplified products. All reactions were setup in triplicate and three biological replications were used for each treatment. The Bio-Rad CFX Manager

v2.1 software was applied for data analysis. Normalization of expression values was carried out with housekeeping genes ribosomal 40S protein subunit S8 (NCBI - XM_003532110) and elongation factor 1-beta (ELF1) (NCBI - XM_003545405). The ΔΔCq method was applied for relative quantification and normalization and standard curves were used to validate primer pairs. All standard curves had a PCR efficiency of between 90 % - 110 % with a R^2 value higher than 0.9.

3.2.4 Statistical analysis

Significant transcription changes in the RNA-Seq data was determined by applying a False Discovery Rate (FDR) of 0.05. Significant changes was determined after correction with the Benjamini-Hochberg correction for multiple-testing.

3.3 RESULTS

3.3.1 RNA-Seq quality scores

RNA-Seq was used as a gene discovery technique to identify either up- or down-regulation of nodule genes due to drought stress treatment. Three biological replications of the three drought stress treatments (60 %, 40 % and 30 % VWC) and their DNS controls were pooled and sequenced. This was done to reduce costs. Gene expression levels were subsequently confirmed on single samples by RT-qPCR. Before further analysis took place, FASTQ was used to determine the Kmer quality of sequences and appropriate action was taken for each parameter to improve the quality. Per base sequence quality and content (Fig. 3.2), GC content and sequence duplication was assessed (Fig. 3.3). All parameters were at an expectable standard after sequences were trimmed, except sequence duplication, due to the paleopolyploidy nature of the soybean genome (Severin *et al*., 2010). After QC-filtering of RNA-Seq results (Table 3.2), reads that were 74 bp in length, were aligned to the soybean genome (Gmax 275 Wm 82A2.V1). More than 20 million paired-end reads were generated for each of the six treatments. Approximately 95 % of bases were further successfully aligned with Tophat 2 (version 0.7).

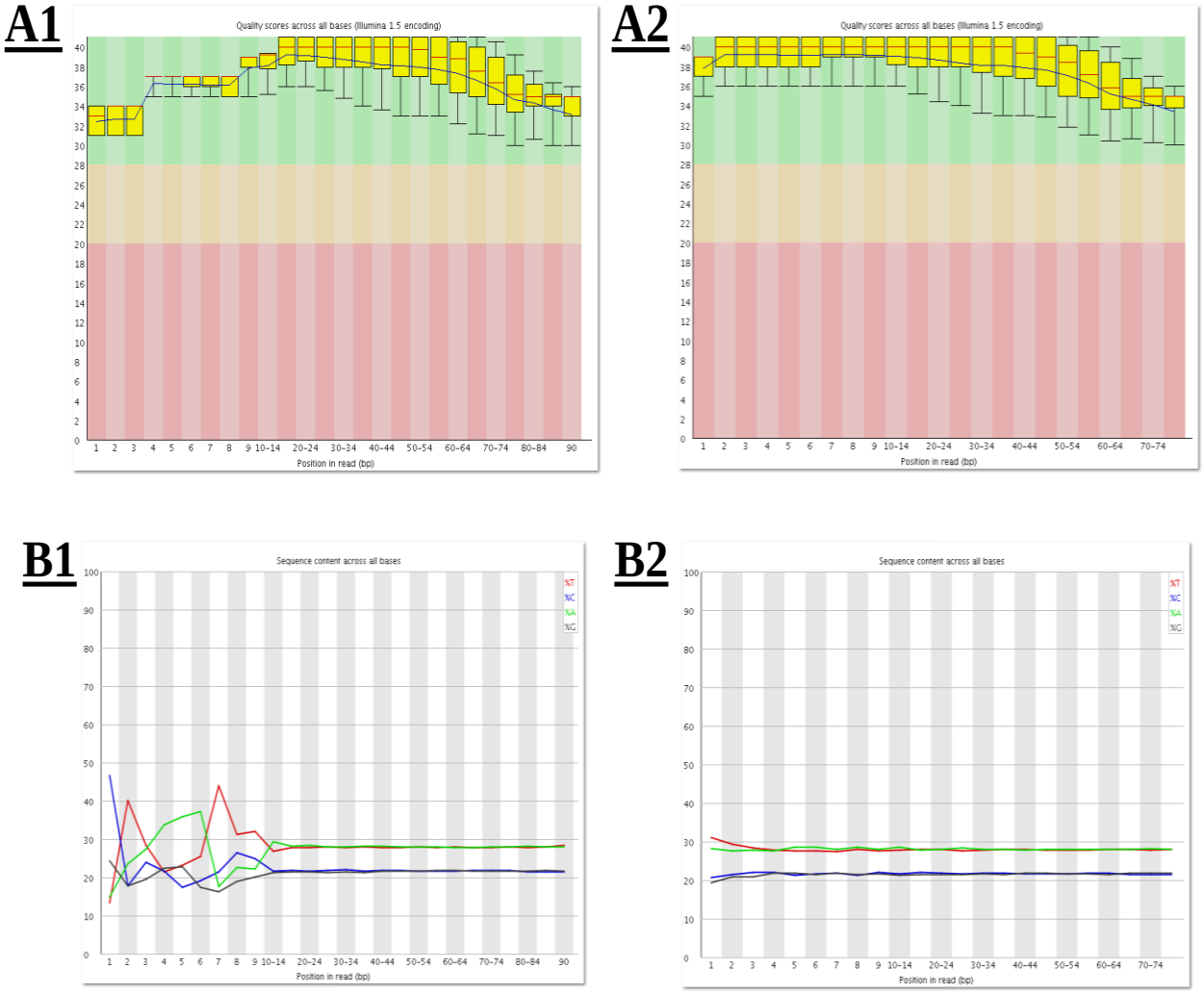


Fig. 3.2: A) Per base sequence quality scores obtained from FASTQC where each base is on the x-axis and the quality score is indicated on the y-axis, A1) prior to trimming A2) after read trimming. The median score base is indicated by the red line. B) Per base sequence content quality, B1) prior to trimming and B2) after read trimming. Each base's position is indicated on the x-axis and the percentage per base is indicated of the y-axis.

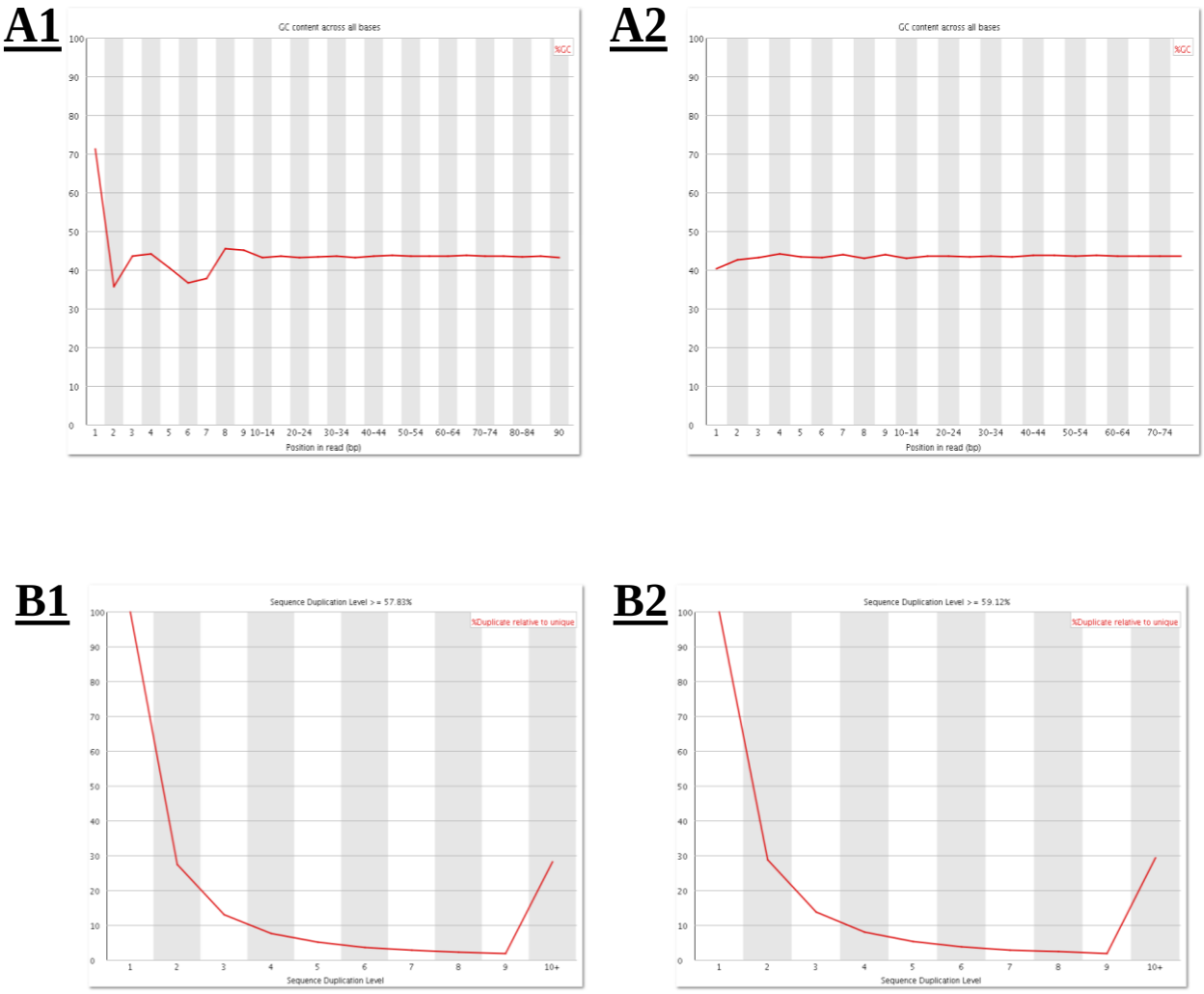


Fig. 3.3: A) Per base GC quantity scores obtained from FASTQC where each base is on the x-axis and the GC content is indicated on the y-axis, A1) prior to trimming A2) after trimming. B) Sequence duplication levels, B1) prior to trimming and B2) after read trimming. The level of sequence duplication is indicated on the x-axis and the sequence that showed to be given in duplication is indicated on the y-axis.

Table 3.2: Quality statistics of the Illumina HiSeq data in comparison to the *Glycine max* reference genome (Gmax 275 Wm 82A2.V1) of each drought stress treatment after processing data using the Collect Alignment Summary tool (version 1.126.0).

Treatment	Post QC read length	Total reads post QC	Total reads aligned	Total bases aligned (%)
DNS (60 %)	74 bp x 74 bp	22 629 610	20 862 078	1 614 327 404 (92 %)
DNS (40 %)	74 bp x 74 bp	22 857 985	21 426 257	1 657 714 931 (94 %)
DNS (30 %)	74 bp x 74 bp	22 442 149	21 544 982	1 669 447 083 (96 %)
DS 60 %	74 bp x 74 bp	23 063 885	22 235 911	1 721 314 332 (96 %)
DS 40 %	74 bp x 74 bp	22 227 744	21 380 393	1 655 721 536 (96 %)
DS 30 %	74 bp x 74 bp	22 112 072	21 039 287	1 630 216 616 (95 %)

3.3.2 Gene expression profile analysis

3.3.2.1 Gene expression overview

The transcriptome of soybean crown nodules was analysed to identify and annotate drought stressed induced gene transcripts over the different levels of drought and investigate gene expression profiles. All genes that had an expression value lower than 1.5 FPKM across all three treatments, were eliminated. Genes that were either 2X log_2-fold up- or down-regulated and showed significant changes were used for further analysis. This ensured that the changes in the genes were statistical relevant and that the changes were changes involved in the gene expression profiles of drought stressed nodules. Drought stress treatments compared to different levels of drought stress and drought stress compared to its DNon-stressed control is shown with Venn diagrams (Fig .3.4-Fig .3.6) that was constructed using the Cuffdiff analysis data.

Overall, using the specified criteria (Table 3.1), only 48 gene transcripts were significantly induced during drought stress (Fig. 3.4 A). More genes (12 transcripts) were induced in the 60 % VWC treatment *vs* the 40 % VWC treatment (data set 1) and in the 60 % VWC *vs* 30% VWC (11 transcripts) when compared to the 40 % *vs* 30 % VWC (four transcripts) were DS samples were compared to other levels of DS (60 % DS *vs* 40 % DS *vs* 30% DS). For all three drought stress treatments combined, only one transcript (*Glyma.20G201800*), a non-annotated hypothetical protein also found in *Phaseolus vulgaris* and *Medicago sativa*, had at least a 2X log_2-fold or higher increase in expression. The amount of significant genes that was only induced by drought stress (data set 2, Fig. 3.4 B) and not affected by developmental senescence as seen with the DNS controls of 60 %, 40 % and 30 % VWC, decreased to 20 gene transcripts. No genes transcripts were found with the above mentioned requirements in all three treatments.

When compared to genes that showed an increase in transcript level as explained above, a different trend was seen in genes that decrease in expression (Fig. 3.5 A). Fifty one gene transcripts (data set 3) were significantly down-regulated over the drought stress treatments. More transcripts showed a decrease in expression in the 40 % VWC *vs* 30 % VWC comparison (14 transcripts) than in the 60 % *vs* 40 % VWC (1 transcripts). One transcript, *Glyma.20G14200* a perioxidase, were seen to decrease over all three treatments. When genes were excluded that also showed a decrease in expression during DNon-stressed conditions, 41 gene transcripts (data set 4) showed to be down-regulated and no genes (Fig. 3.5 B) showed to be down-regulated over all three treatments.

When a comparison was made between the drought stress treatments and their same age DNS controls, it was apparent that more transcripts were up-regulated from DNS nodules to DS nodules (data set 5, Fig. 3.6 A) at 40 % VWC (136 transcripts) and 30 % VWC (98 transcripts) than at 60 % VWC (27 transcripts). The opposite was true for down regulated genes (data set 6, Fig. 3.6 B) at 40 % VWC (16 transcripts) and 30 % VWC (24 transcripts) less genes were down regulated compared to 60 % VWC (30 transcripts).

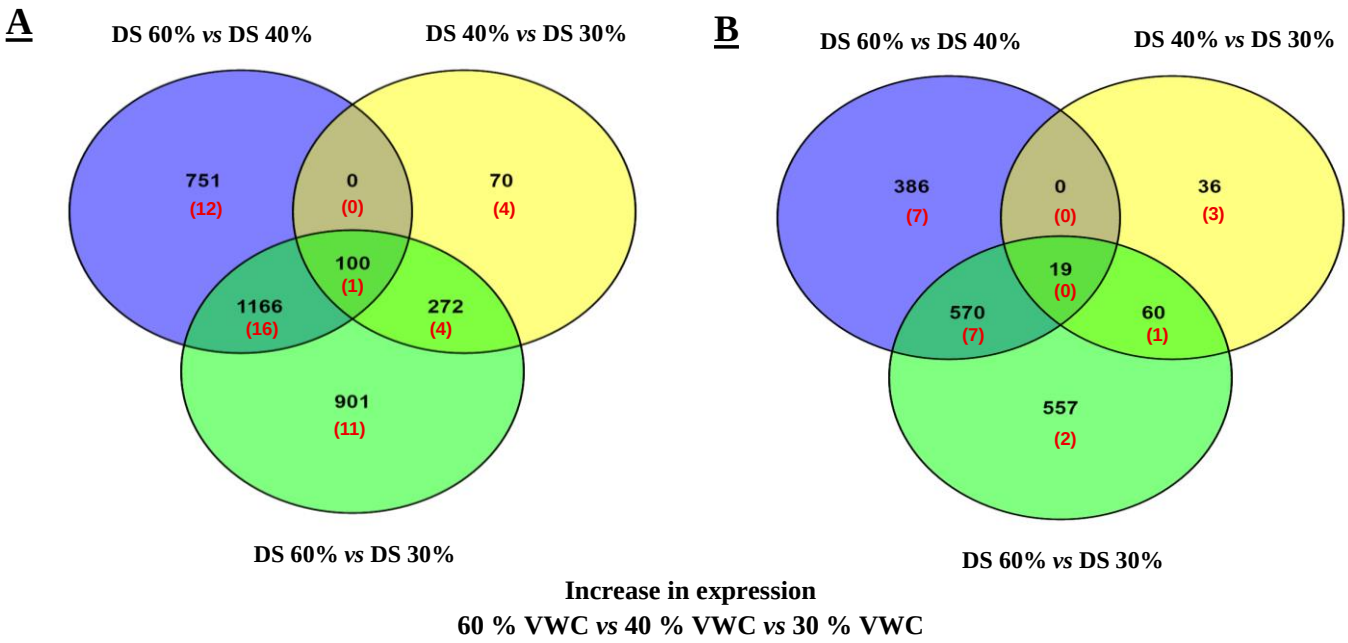


Fig. 3.4: Venn diagram of genes unique to each drought stress treatment, as well as genes with overlapping activity with a FPKM ≥ 1.5 and a change in expression of ≥ 2X log2-fold. Genes from DS samples at 60 %, 40 % and 30 % VWC were compared to each other. A). All genes that showed to have an increase (data set 1) in expression from one drought stress treatment to the other. B) All genes that showed to have an increase in expression (data set 2) from one drought stress treatment to the other which did not show to have a 2X log_2-fold increase in expression over all three of the DNS controls. Significant changes are indicated in red.

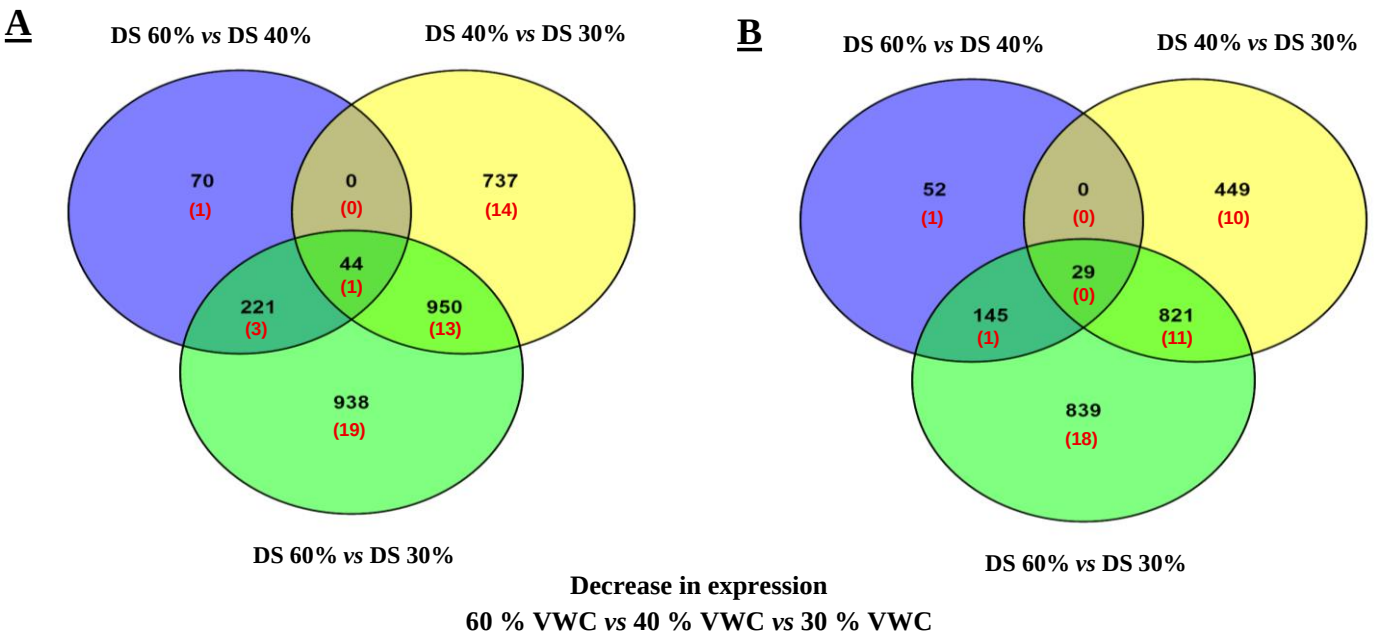


Fig. 3.5: Venn diagram of significant genes unique to each drought stress treatment, as well as genes with overlapping activity with a FPKM ≥ 1.5 and a change in expression of ≥ 2X log_2-fold. Genes from DS samples at 60 %, 40 % and 30 % VWC were compared to each other. A) All genes that showed to have a decrease (data set 3) in expression and B) genes that showed to have a decrease in expression from one drought stress treatment to the other (data set 4) which did not show to have a 2X log_2-fold decrease in expression over all three of the DNS controls. Significant changes are indicated in red.

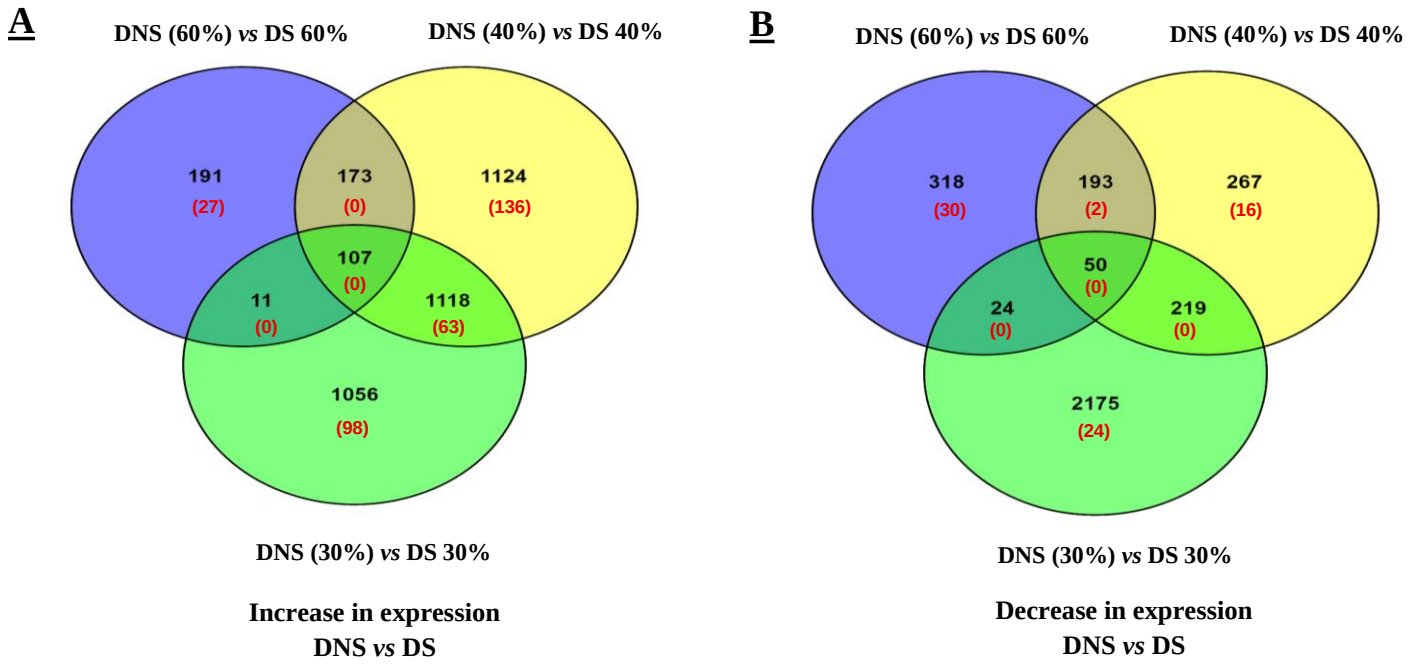


Fig. 3.6: Venn diagram of genes unique to each drought treatment, as well as genes with overlapping activity with a FPKM ≥ 1.5 and a change in expression of ≥ 2X $\log_2$-fold. Genes from DS samples at 60 %, 40 % and 30 % VWC were compared to each DNS controls A) All genes that showed to have an increase (data set 5) in expression and B) decrease in expression (data set 6) which did not show to have a 2X $\log_2$-fold decrease in expression over all three of the DNS controls. Significant changes are indicated in red.

3.3.2.2 Gene Ontology

The next step was to determine the gene ontology (GO) terms of significantly over expressed genes. The most important biological processes, cellular compartments and molecular function of the transcripts at each drought stress treatments were determined using AgriGO (http://bioinfo.cau.edu.cn/agriGO/) (Du *et al*., 2010). The subset of genes that was unique to each drought stress treatments as well as genes that overlapped over drought stress, were used in the analysis. Network maps of the various biological processes, cellular components or molecular functions were constructed. If no gene ontology was found, different ontology databases such as the PFAM > Panther > KOG > KEGG ontology was used to elucidate a possible gene function.

AgriGO provide illustrations of over-represented genes terms in bar charts and hierarchical tree graphs. The reference list used was compiled by AgriGO from data collected from the soybean genome found on the Phytozome database (Gmax 275 Wm 82A2.V1). This data was generated from leaf tissue of two week old plants and after etiolation for five days prior to harvest. Therefore the GO term percentages cannot be considered as a comparison due to the difference in tissue type (leaf *vs* nodule), the age of the tissue (2 weeks *vs* ± 8 weeks, ± 9 weeks or ± 11 weeks) and the tissue was not stressed.

Bar charts were constructed using only four data sets. The first two being genes that significantly increased and decreased in expression during drought stress but did not show to have a 2X $\log_2$-fold increase/decrease in expression over all three of the DNS controls (data set 2 and 4). The last two sets (data set 5 and 6) were nodules that showed a significant increase/decrease of DS samples compared to their DNS control. The x-axis indicates the

specific GO term and the y-axis indicates the percentage of genes mapped to each specific GO term divided by the total number of genes mapped in the input list.

No GO terms was available for data set 2. This is due to the genes but being present in the soybean leaf transcriptome used as reference. Bar Charts (Fig. 3.7 A and B) were constructed with significant over-represented GO terms of the data set of nodules that decreased at 40 % *vs* 30 % VWC and 60 % *vs* 30 % VWC (data set 4). Both the data sets were grouped only to biological processes such as cellular and metabolic processes and molecular function such as catalytic activity and binding. More genes were significantly up-regulated and down-regulated when comparing each DS treatments to it DNS control (Fig. 3.8, 3.10, 3.12, 3.14 A and B).

The over expressed GO terms for each category (biological process, cellular compartment and molecular function) were also illustrated as a hierarchical tree graphs (Fig. 3.9, 3.11, 3.13, 3.15). Each box contains the GO term labelled with a GO ID and the term definition. All non-significant terms are shown in white boxes whereas significant terms are indicated according to the level of enrichments which can be seen in the degree of saturation. Solid, dashed, and dotted lines indicate two, one and zero enriched GO terms at both ends connected by the line, respectively. Red arrows indicate positive regulation and green arrows indicate negative regulation.

In the gene list where nodules stressed at 60 % VWC was compared to its DNS control and expression increases significantly (Fig. 3.9, data set 5), six genes (*Glyma.04G076900, Glyma.12G149100, Glyma.16G043200, Glyma.13G279900, Glyma.19G108800* and *Glyma.06G248900*) were over-represented according to GO terms in biological processes. These six genes are involved in processes such as DNA binding and transcriptional regulation

of metabolic, cellular biosynthesis, genes expression and nitrogen compound metabolic processes.

GO molecular function terms that were over-represented and significant in the gene list (data set 5) where nodules were stressed at 40 % VWC compared to its DNS control (Fig. 3.11), were enzyme inhibitor activity (GO:0004857), endopeptidase inhibitor activity (GO:0004866), peptidase inhibitor activity (GO:0030414) and enzyme regulator activity (GO:0030234). Seven genes (*Glyma.08G342100, Glyma.09G163600, Glyma.18G003700, Glyma.12G234800, Glyma.16G212200 Glyma.16G212100* and *Glyma.20G205700*) were involved in enzyme inhibitor activity, endopeptidase inhibitor activity, peptidase inhibitor activity and three were involved in enzyme regulator activity (*Glyma.03G029100, Glyma.07G237300* and *Glyma.03G216000*). These genes were over-represented in enzyme inhibitor activity and enzyme regulator activity functions.

Twelve genes were seen to be functional in the membrane of the cellular component (GO: 0044425) using data set 5 where nodules samples at 30 % VWC were compared to its DNS control (Fig 3.13). The genes were *Glyma.19G159000, Glyma.19G009900, Glyma.14G188000, Glyma.17G185600, Glyma.17G186400, Glyma.04G198400, Glyma.16G141000, Glyma.06G078700, Glyma.04G198600, Glyma.08G204500, Glyma.06G154200, Glyma.06G154200* and *Glyma.13G186400*. Where genes decrease significantly in expression in the data set where nodules samples at 60 % VWC was compared to its same age control (Fig. 3. 15), five genes (*Glyma.11G058100, Glyma.10G200800, Glyma.03G007700, Glyma.06G269700* and *Glyma.06G275900*) were over-represented in GO terms. These genes are involved in the biological process of oxidation and the molecular process of cat-ion and metal ion binding.

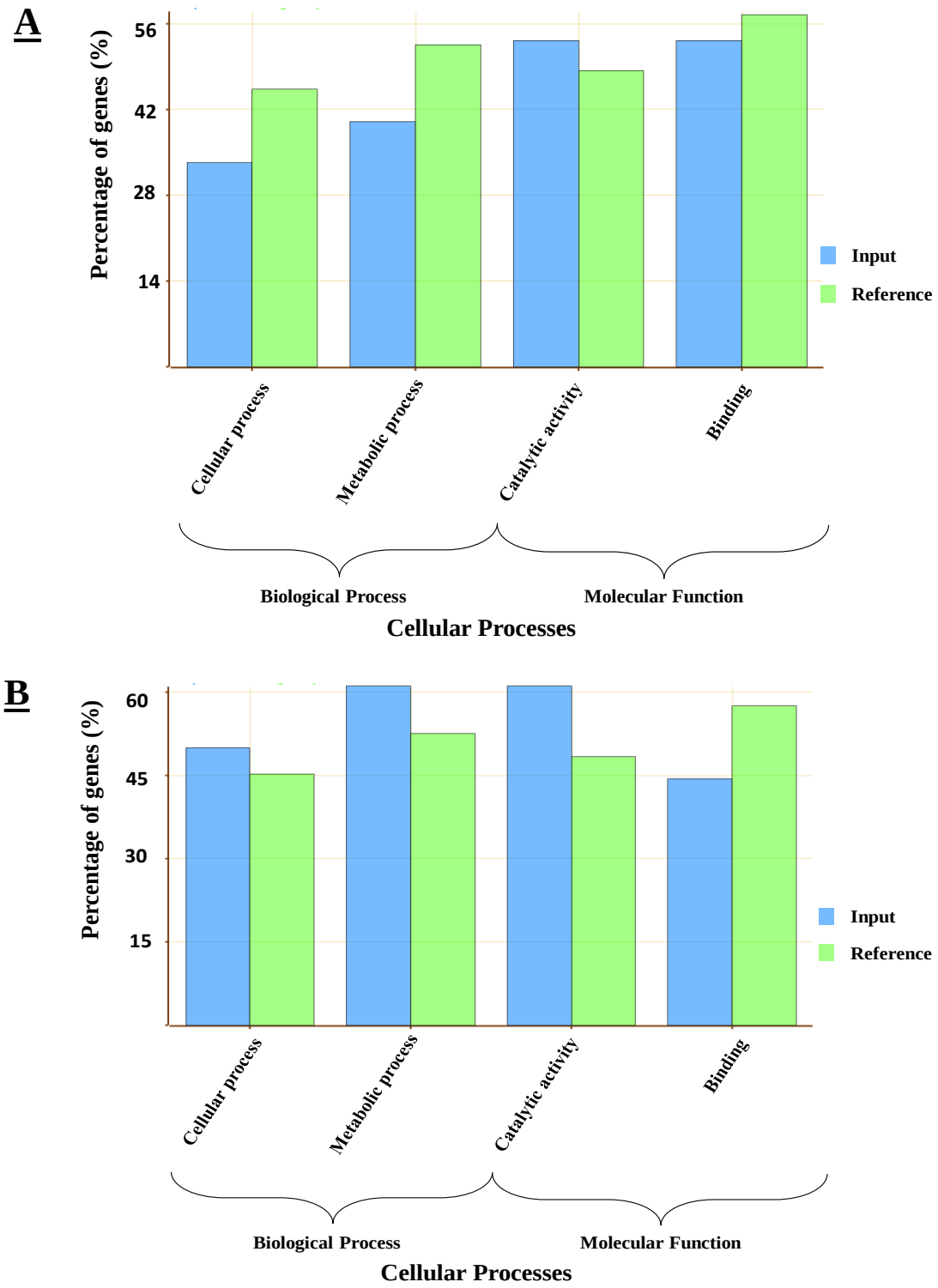


Fig. 3.7: GO terms of over-represented genes with a significant decrease (data set 4). A) DS 40 % VWC *vs* 30 % VWC B) DS 60 % VWC *vs* 30 % VWC which did not show to have a 2X $\log_2$-fold decrease in expression over all three of the DNS controls. No GO terms for the 60 % *vs* 40 % VWC treatment were over-represented.

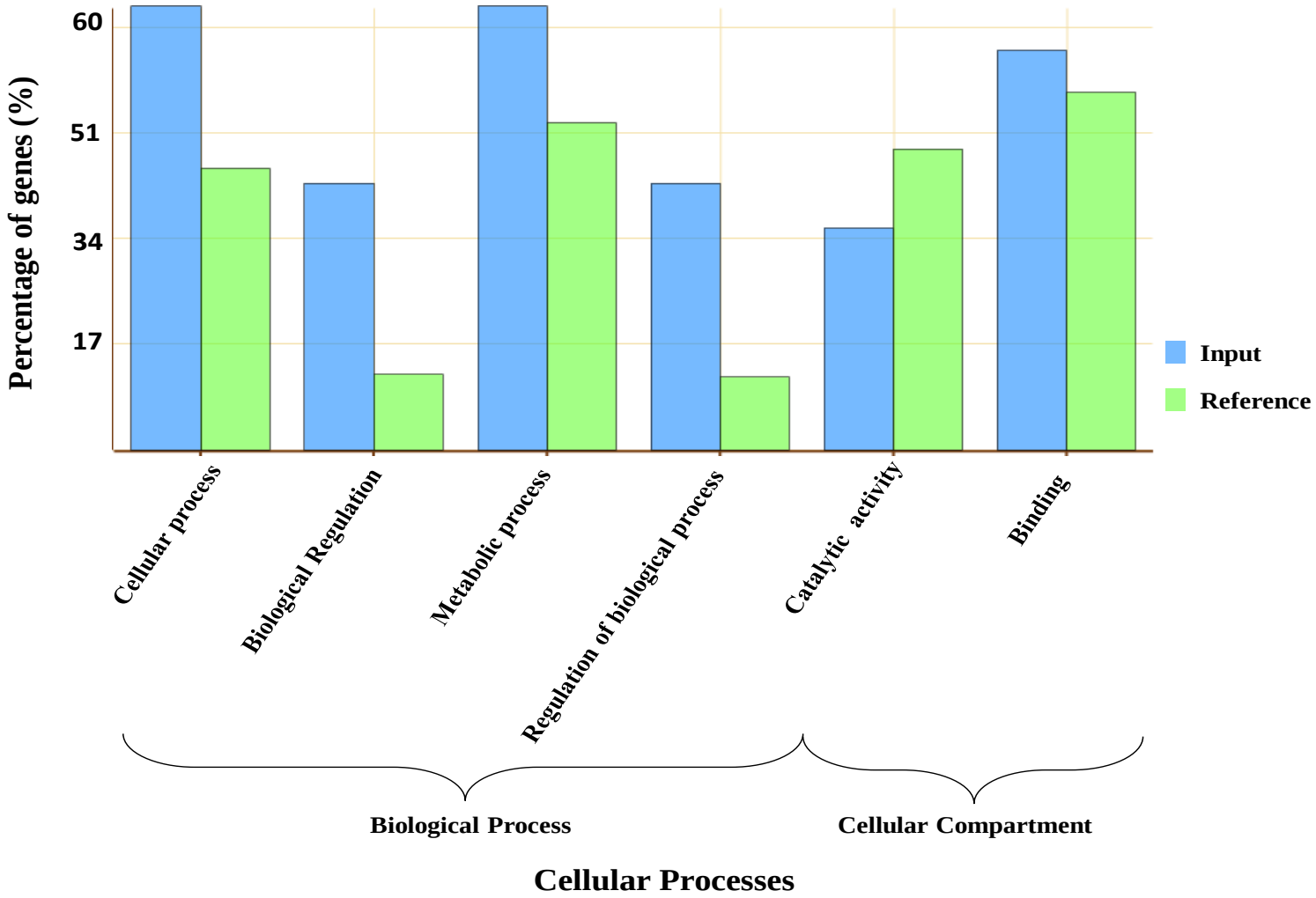


Fig. 3.8: GO terms of over-represented genes with a significant increase in expression at DS 60 % VWC *vs* DNS (60 %) VWC samples (data set 5).

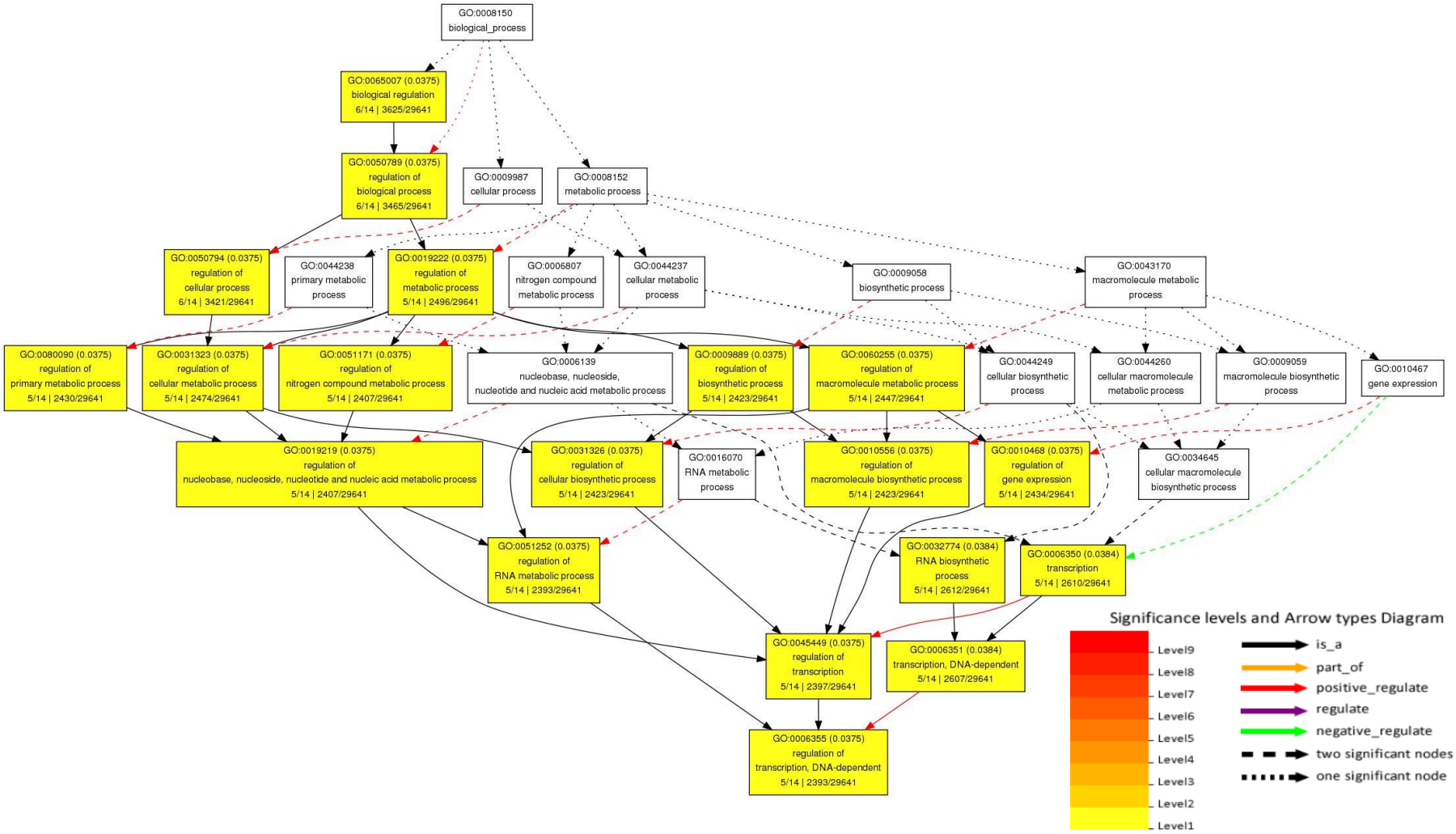


Fig. 3.9: GO terms of over-represented genes with a significant increase in expression at DS 60 % VWC *vs* DNS (60 %) VWC (data set 5). No GO term for molecular function and cellular compartment was over-represented.

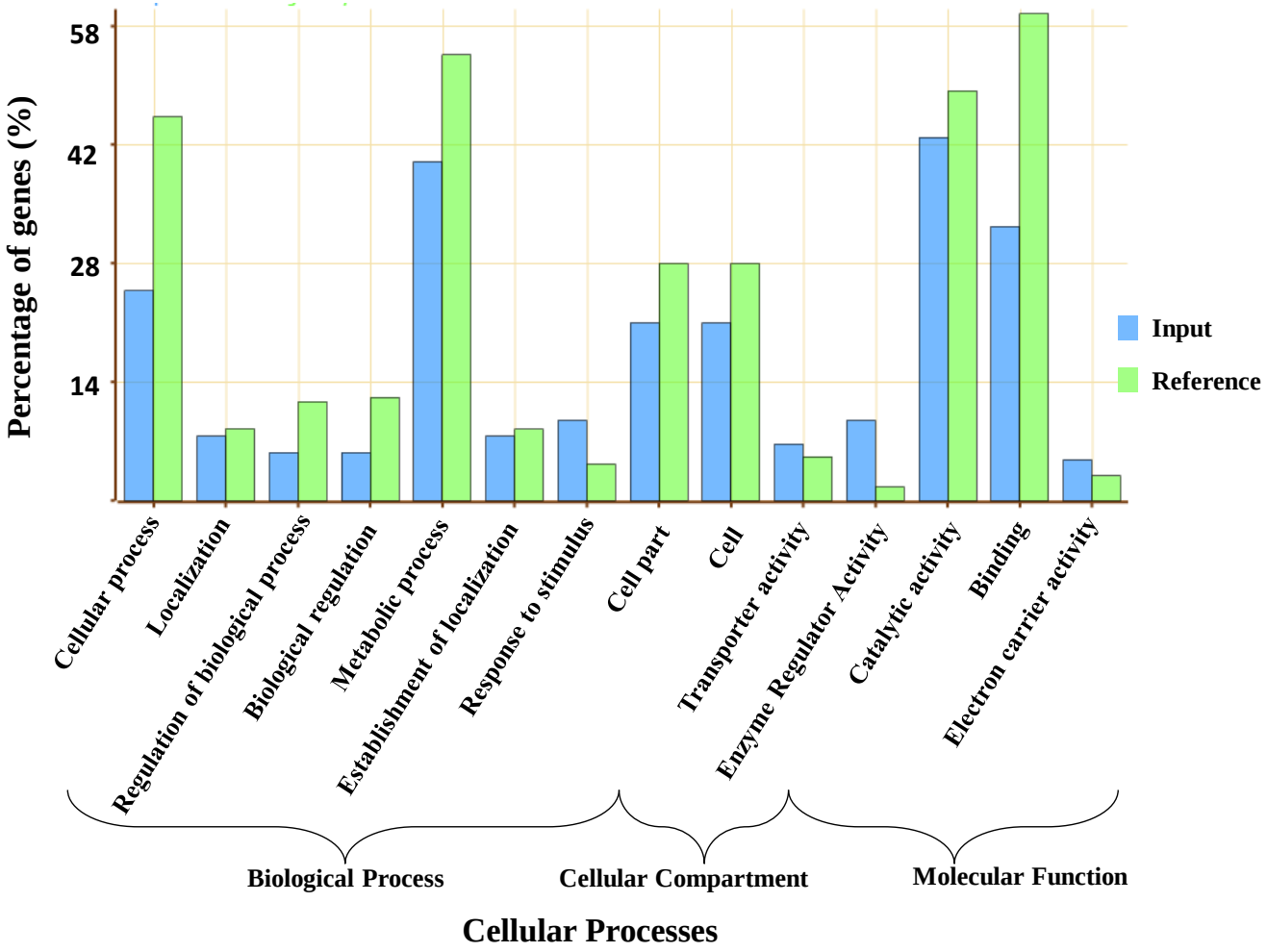


Fig. 3. 10: GO terms of over-represented genes with a significant increase in expression at DS 40 % VWC *vs* DNS (40 %) VWC (data set 5).

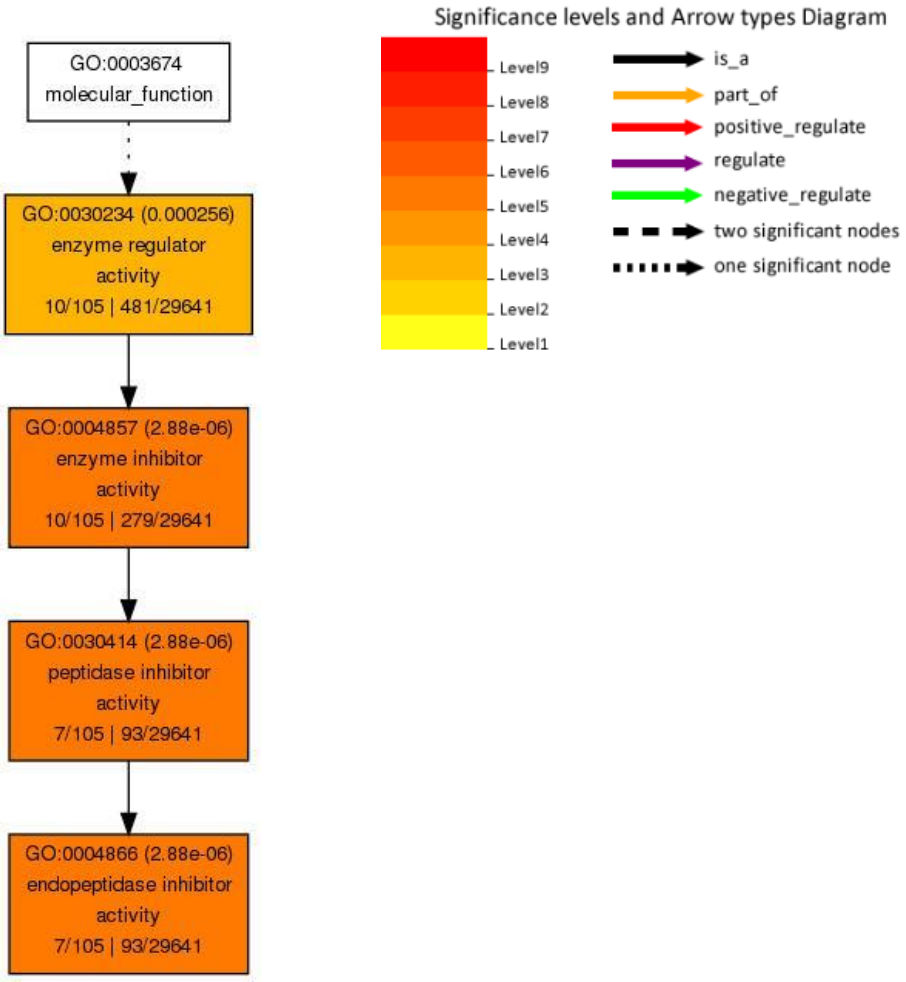


Fig. 3.11: GO terms of over-represented genes with a significant increase in expression from the DS 40 % VWC *vs* DNS (40 %) VWC (data set 5). No GO terms were over-represented for biological processes or cellular compartments.

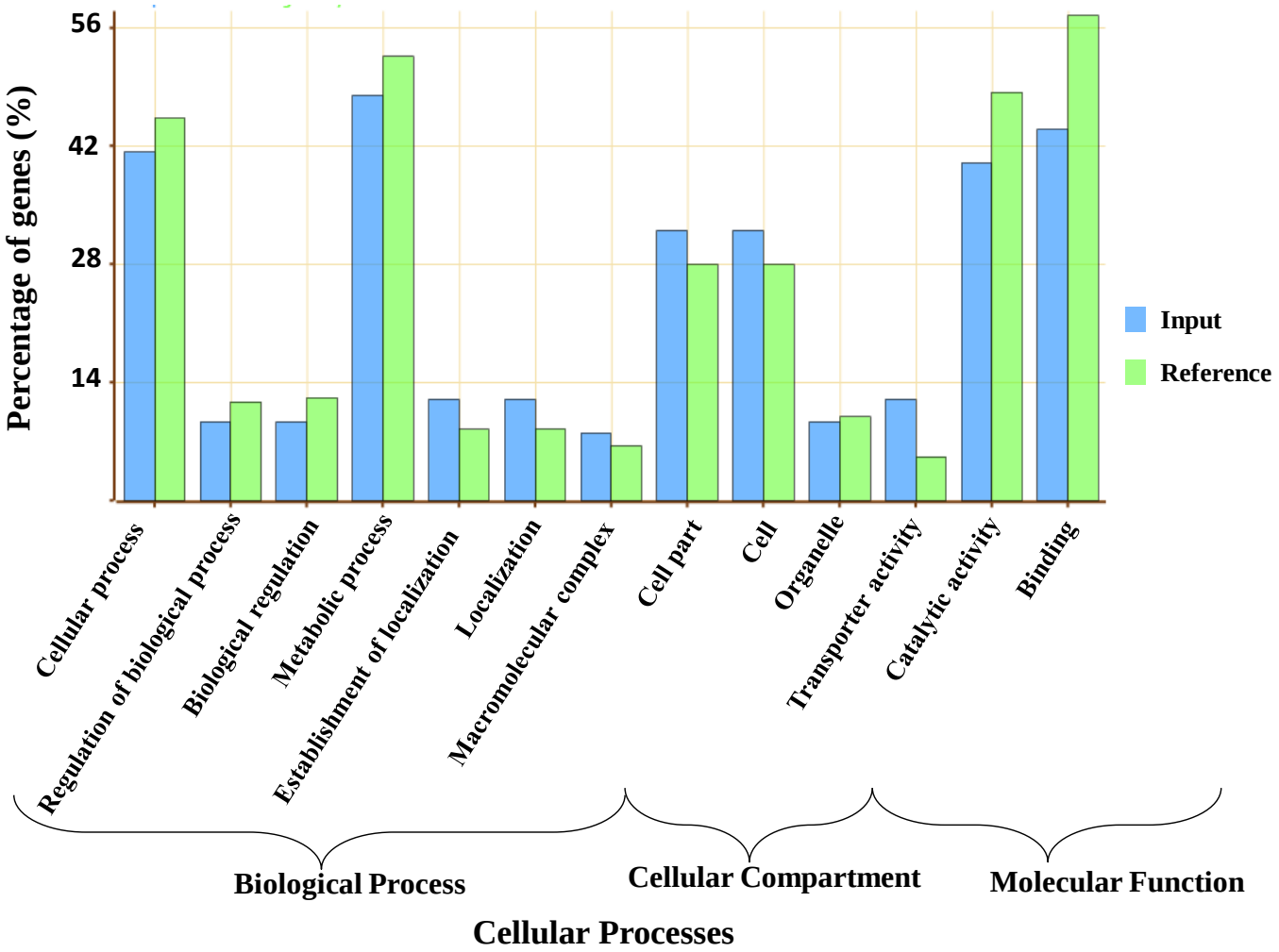


Fig. 3.12: GO terms of over-represented genes with a significant increase in expression at DS 30 % VWC *vs* DNS (30 %) VWC (data set 5).

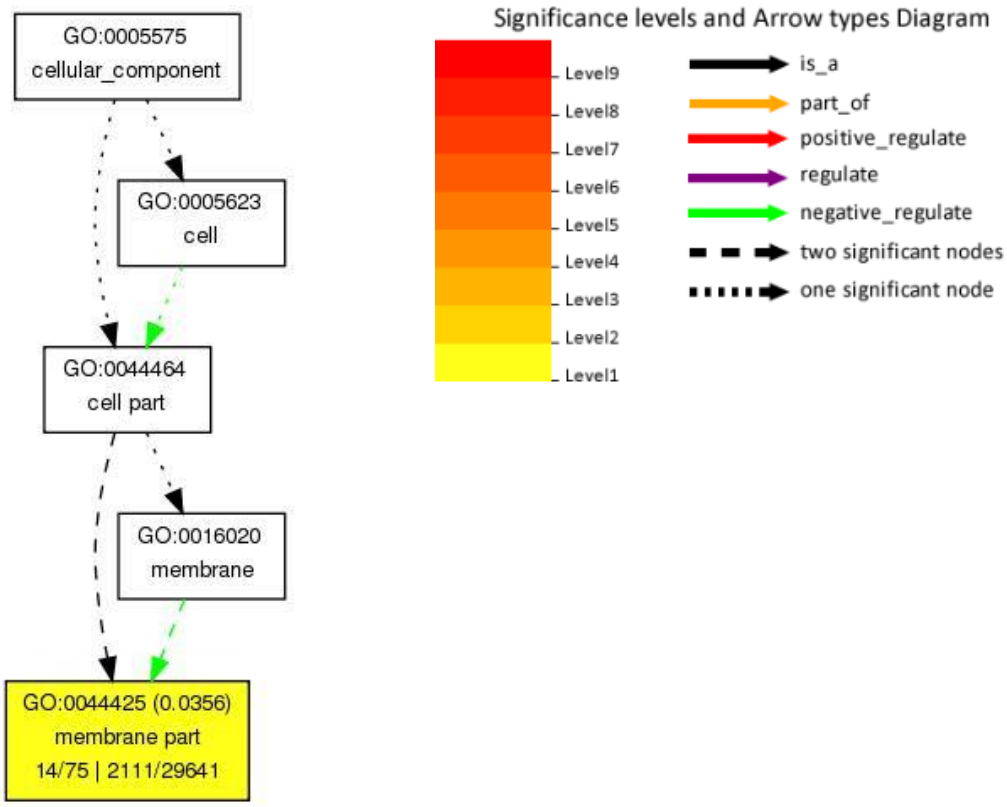


Fig. 3.13: GO terms of over-represented genes with a significant increase in expression at DS 30 % VWC *vs* DNS (30 %) VWC (data set 5). No GO terms were over-represented for biological processes or molecular function.

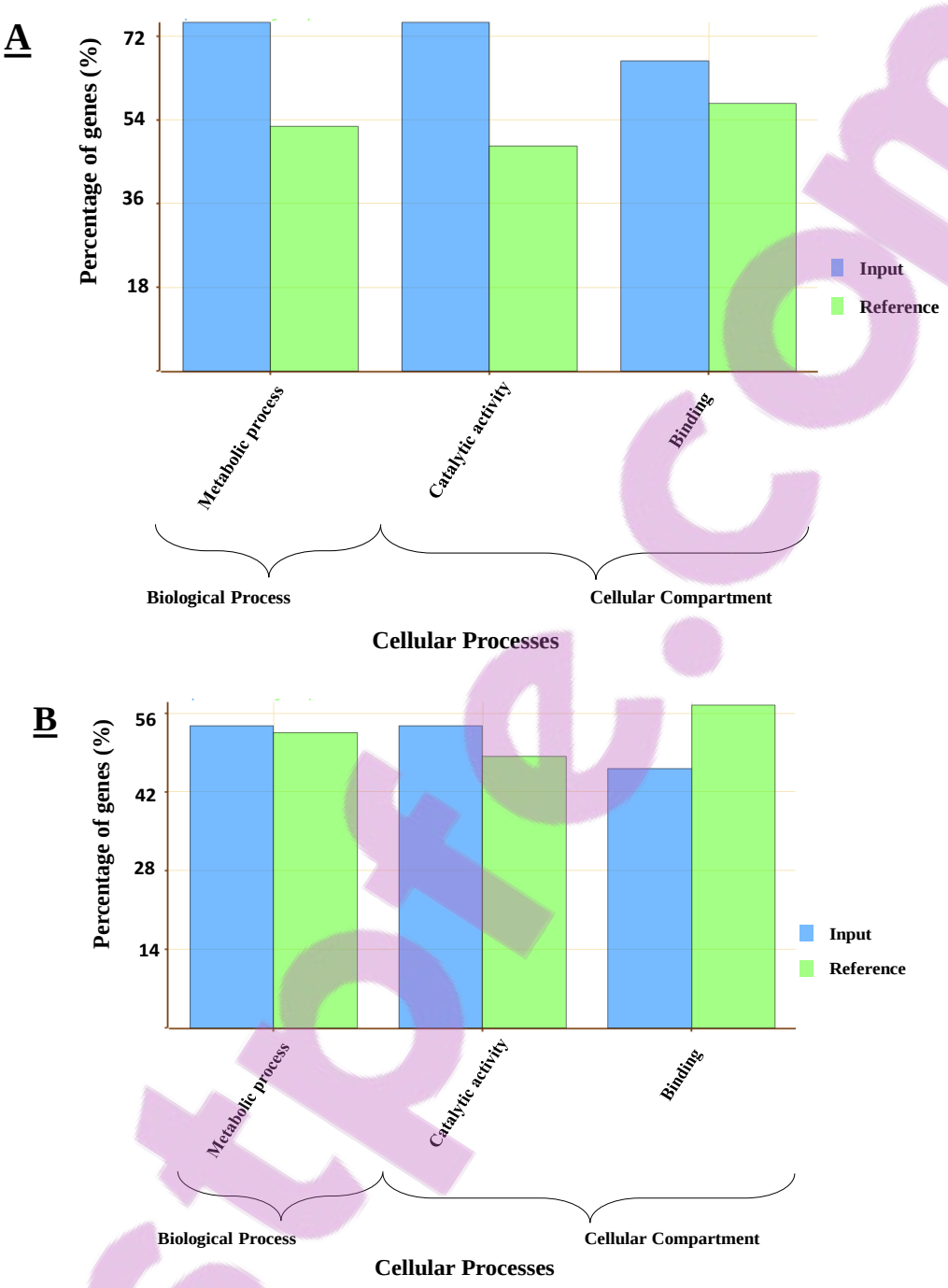


Fig. 3.14: GO terms of over-represented genes with a significant decrease from the data set 6. A) DS 60 % VWC *vs* DNS (60 %) VWC. B) DS 30 % VWC *vs* DNS (30 %) VWC which did not show to have a 2X $\log_2$-fold decrease in expression over all three of the DNS controls.

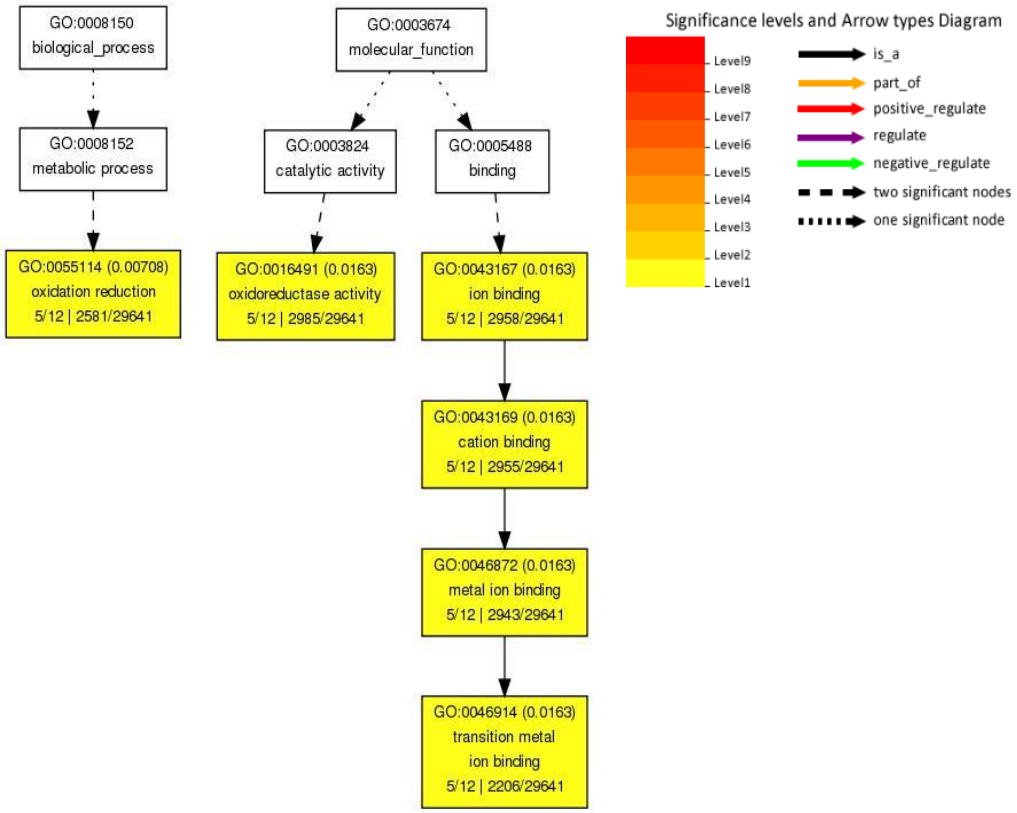


Fig. 3.15: GO terms of over-represented genes with a significant decrease at DS 60 % VWC *vs* DNS (60 %) VWC (data set 6). No GO terms were over-represented for cellular compartment.

3.3.2.3 Gene functional annotation

Further, gene annotation of important and physiological relevance genes that showed to be induced by drought stress were investigated. Ontological classification as found on the Soybase and the Phytozome database were used in conjunction with the MapMan software package to identify a possible gene functions.

As illustrated in Table 3.3 (data set 2 and 4), there were 13 genes that were up-regulated and two gene transcripts that were down-regulated more than 2X log_2-fold in DS 60 % VWC *vs* DS 40 %. The expression changes in these genes were significant and do not have the same effect in the DNS nodules. The gene transcript, *Glyma.20G205800* were the most up-regulated gene in 60 % VWC *vs* 40 % VWC in DS nodules, with a fold change of higher than 8X log_2-fold. This gene is a serine protease inhibitor called a Potato type I inhibitor. Unfortunately, the gene transcripts which expression was the most down-regulated (*Glyma.20G170800*) have not been annotated as of yet and no known homologues were available. The gene transcript *Glyma.20G206000*, a telomerase activator was down-regulated more than 2X log_2-fold. A telomerase caps chromosomes and activates damaged DNA responsive pathways (Fitzgerald *et al*., 1996). Fitzgerald *et al*., 1996 showed that the telomerase gene was expressed in soybean undifferentiated cells and was involved in cell division. Moreover, the strong down regulation of a telomerase activator could lead to inefficient cell division in nodules. The bin maps (Fig. 3.16) shows the distribution of the genes in Table 3.3 with differential gene regulation grouped into broad physiological functions, in this case, signalling, protein, RNA and secondary metabolism.

In total only three genes were significantly up-regulated and 18 genes significantly down-regulated from 40 % VWC *vs* 30 % VWC in DS nodules that did not have an effect in the DNS

samples (Table 3.4 (data set 2 and 4)). Included in these genes was the leghemoglobin gene, *Glyma.20G191200*, that showed a 2.55X log_2-fold down-regulation in expression supporting the loss of function of root nodules. Various transcription factors such as *Glyma.20G231800* and *Glyma.20G209700* were down-regulated during drought stress as well as *Glyma.20G180100*, a GATA transcription factor. This was in contrast to a study done on natural age related senescence where a clear increase in transcription factors was seen (Van Wyk *et al*., 2014). Chitinase (*Glyma.20G225200*), a gene usually associated with a defence response in plants were down-regulated in nodules. Gijzen *et al*. (2001) found that chitinase was up-regulated in senescence, ripening, and a response to pathogen infection. The bin maps (Fig. 3.17) of the above mentioned genes showed a down-regulation in secondary, lipid and amino acid metabolism.

Only eight genes were up-regulated more than 2X log_2-fold in DS 60 % *vs* DS 30 % VWC, whereas 29 were down-regulated (Table 3.5 (data set 2 and 4)). Again the most up-regulated gene was *Glyma.20G205800*, the serine protease inhibitor as described previously. The most down-regulated gene was a Hevein-like gene, *Glyma.20G225200*, which was down regulated more than 5-fold. This Hevein-like gene might be a putative chitinase. Other genes also mentioned above such as transcription factors and leghemoglobin were also down-regulated in this data set. The bin maps (Fig. 3.18) showed that most up-regulated genes were involved in secondary metabolism, protein metabolism and transport whereas the down-regulated genes were involved in the cell wall, lipid, secondary and nucleotide metabolism. Two of the genes were also involved in stress responses and redox.

When investigating genes that show differential expression compared to their respective DNS controls, 26 genes were up-regulated from DNS (60 %) *vs* DS 60 % (Table 3.6 (data set 5))

and 22 genes were down-regulated (Table 3. 6). Five of the genes that were up-regulated more than 2X log_2-fold were NAC-transcription factors (*Glyma.12G149100, Glyma.06G248900, Glyma.13G279900, Glyma.16G043200* and *Glyma.19G108800*). These transcription factors have been seen to be induced during various stress responses as they were components in the complex signalling progresses during plant stress. Over expression of these genes improved drought tolerance in potato and rice (Nuruzzaman *et al*., 2013). Another gene that was down-regulated more than 8X log_2-fold was an early nodulin gene (*Glyma.04G060600*), usually involved in root hair deformation as well as early nodule development. Most genes that were down-regulated were involved in development, and signalling (Fig. 3.19) whereas up-regulated genes were involved in hormone development and in the cell wall according to the bin maps.

One hundred and eighty transcripts were up-regulated when comparing DNS (40 %) *vs* DS 40 % and only 14 were significantly down-regulated in data set 10 and 12 (Table 3.7). A Defensin-like gene (*Glyma.13G27800*) was the most up-regulated gene in this data set with 17X log_2-fold change. This gene transcribes small cysteine-rich compounds (Lay and Anderson 2005). A further strongly up-regulated nodule gene (14X log_2-fold) coded for the *LEA*-D11 protein (*Glyma.05G112000*). The *LEA* (late embryogenesis abundant proteins) were produced for surviving environmental stresses. (Erikson and Harryson, 2011). According to the bin maps (Fig. 3.20), the most up-regulated genes were involved in stress responses, signalling, transport and development.

In the most severe drought stress conditions, DNS (30 %) *vs* DS 30 % VWC, 156 genes were up-regulated and 22 were down regulated (Table 3.8). Thirty of the up-regulated genes have not been annotated. Fifteen genes were associated with encoding a heat shock protein and six encode *LEA* proteins. Another gene of interest was an inositol polyphosphate 5-phosphatase

(*Glyma.14g077900*) which in lipid transfer. A cystatin was also seen to be significantly up-regulated more than 6X log_2-fold (*Glyma.18g003700*). The most prominent down-regulated nodule gene was *Glyma.03G132700*, a β 1,3 glucanse which is a pathogenesis-related protein that is responsible for the elicitation of host-defence responses, including glyceollin synthesis in root nodules. The bin map, Fig. 3.21 is very similar to Fig. 3.20. Twenty one genes were involved in a stress response.

Table 3.3: Differentially regulated genes comparing the stress treatments of DS 60 % VWC *vs* DS 40 % VWC after significance and FDR cut-offs have been applied. Data set 2 and 4 were used to construct the table.

	FKPM						
Gene ID	**DS 60%**	**DS 40%**	**Log_2-fold**	**Significant**	**Gene call**	**Annotation**	**PHYTOZOME Define**
Glyma.20G205800	4.54	1456.84	8.33	Yes	AT2G38870	Serine protease inhibitor, potato inhibitor I-type	
Glyma.20G241700	0.65	10.06	3.95	Yes	AT3G55120	Chalcone-flavanone isomerase	
Glyma.20G179700	15.17	117.52	2.95	Yes	AT2G37170	Plasma membrane intrinsic protein 2	Aquaporin pip2-1-related
Glyma.20G186100	0.33	2.13	2.68	Yes	AT1G25570	Di-glucose binding protein with Leucine-rich repeat domain	Serine-threonine protein kinase, plant-type
Glyma.20G231000	2.19	12.99	2.57	Yes	AT3G05140	ROP binding protein kinases 2	Serine-threonine protein kinase, plant-type
Glyma.20G240600	0.92	4.80	2.38	Yes	AT2G39420	Alpha/beta-Hydrolases superfamily protein	Phospholipase-related
Glyma.20G188800	1.05	5.34	2.35	Yes	AT1G08230	Transmembrane amino acid transporter family protein	
Glyma.20G180100	2.08	10.53	2.34	Yes	AT4G32890	GATA transcription factor 9	
Glyma.20G236900	13.73	67.91	2.31	Yes	AT3G05550	Hypoxia-responsive family protein	
Glyma.20G209700	0.64	3.03	2.24	Yes	AT3G23250	Myb domain protein	
Glyma.20G194000	2.60	11.44	2.14	Yes	AT5G04830	Nuclear transport factor 2 (NTF2) family protein	
Glyma.20G210100	27.92	119.11	2.09	Yes	AT1G03220	Eukaryotic aspartyl protease family protein	Aspartyl proteases
Glyma.20G203000	2.52	10.12	2.01	Yes		Not annotated	
Glyma.20G206000	1.88	0.47	-2.01	Yes	AT3G09290	Telomerase activator	
Glyma.20G170800	2.17	0.38	-2.50	Yes		Not annotated	

Fig. 3.16: MapMan generated Bin maps of drought stressed nodule samples DS 60 % VWC *vs* DS 40 % VWC of significant up- and down-regulated genes (data set 2 and 4). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

Table 3.4: Differentially regulated genes comparing the stress treatments of DS 40 % VWC *vs* DS 30 % VWC after significance and FDR cut-offs have been applied. Data set 2 and 4 were used to construct the table.

	FKPM						
Gene ID	**DS 40%**	**DS 30%**	**Log_2-fold**	**Significant**	**Gene call**	**Annotation**	**PHYTOZOME Define**
Glyma.20G170800	0.38	2.32	2.59	Yes		Not annotated	
Glyma.20G230100	1.30	6.19	2.26	Yes	PTHR20863	Acyl carrier protein 1, chloroplastic-related	
Glyma.20G183300	0.73	2.96	2.01	Yes		Not annotated	
Glyma.20G169200	68.83	17.17	-2.00	Yes	AT1G71695	Peroxidase superfamily protein	
Glyma.20G190100	2.58	0.63	-2.04	Yes	AT5G04490	Vitamin E pathway gene 5	Transmembrane protein 15-related
Glyma.20G231800	2.09	0.50	-2.07	Yes	AT1G72210	Basic helix-loop-helix (bhlh) DNA-bind	Transcription factor bhlh99
Glyma.20G223200	2.81	0.64	-2.14	Yes	AT1G08630	Threonine aldolase 1	
Glyma.20G220900	113.37	25.58	-2.15	Yes	AT3G05950	Rmlc-like cupins superfamily protein	
Glyma.20G240600	4.80	1.00	-2.26	Yes	AT2G39420	Alpha/beta-Hydrolases superfamily protein	
Glyma.20G180100	10.53	1.96	-2.43	Yes	AT4G32890	GATA transcription factor 9	
Glyma.20G233900	1.39	0.25	-2.48	Yes	AT1G09010	Glycoside hydrolase family 2 protein	
Glyma.20G191200	6026.27	1031.44	-2.55	Yes	AT3G10520	Haemoglobin 2	Non-symbiotic haemoglobin 2
Glyma.20G209700	3.03	0.48	-2.67	Yes	AT3G23250	Myb domain protein 15	
Glyma.20G241600	14.86	2.24	-2.73	Yes	AT3G55120	Chalcone-flavanone isomerase family protein	
Glyma.20G188800	5.34	0.76	-2.82	Yes	AT1G08230	Transmembrane amino acid transporter family protein	
Glyma.20G215000	5.04	0.63	-3.00	Yes	AT3G55620	Translation initiation factor IF6	
Glyma.20G224200	1.66	0.12	-3.79	Yes	AT4G14746	Interpro DOMAIN/s: EGF-like	
Glyma.20G177000	2.81	0.13	-4.41	Yes	PTHR10794	Alpha/beta-hydrolase domain-containing protein	
Glyma.20G192700	1.99	0.08	-4.60	Yes	PTHR11764	Lanosterol synthase	
Glyma.20G245200	3.12	0.10	-4.96	Yes		Not annotated	
Glyma.20G225200	3.65	0.10	-5.19	Yes	AT3G04720	Pathogenesis-related 4	Chitinase-related

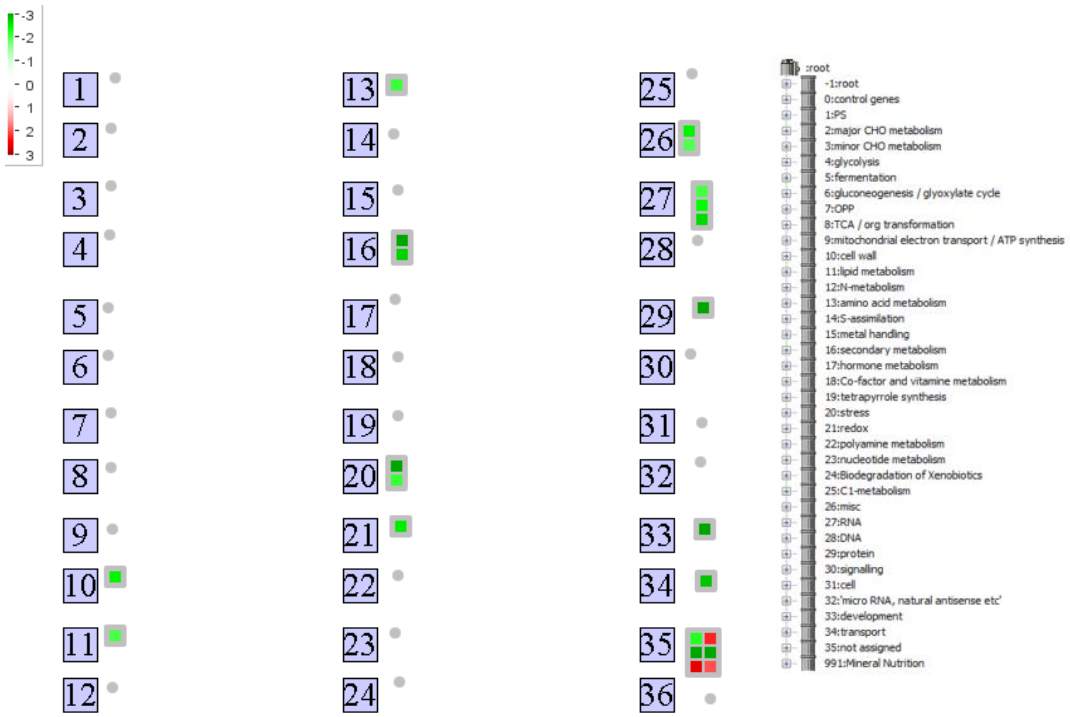


Fig. 3.17: MapMan generated Bin maps of drought stressed nodule samples DS 40 % VWC *vs* DS 30 % VWC of significant up- and down-regulated genes (data set 2 and 4). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

Table 3.5: Differentially regulated genes comparing the stress treatments of DS 60 % VWC *vs* DS 30 % VWC after significance and FDR cut-offs have been applied. Data set 2 and 4 were used to construct the table.

	FKPM						
Gene ID	**DS 60%**	**DS 30%**	**Log_2-fold**	**Significant**	**Gene call**	**Annotation**	**PHYTOZOME Define**
Glyma.20G205800	4,54	474,56	6,71	Yes	AT2G38870	Serine protease inhibitor, potato inhibitor I-type family protein	
Glyma.20G242000	1,23	5,01	2,03	Yes	AT1G05805	Basic helix-loop-helix (bhlh) DNA-binding superfamily protein	
Glyma.20G241700	0,65	26,73	5,36	Yes	AT3G55120	Chalcone-flavanone isomerase family protein	
Glyma.20G179700	15,17	84,29	2,47	Yes	AT1G31050	Basic helix-loop-helix (bhlh) DNA-binding superfamily protein	
Glyma.20G231000	2,19	20,23	3,20	Yes	AT3G12090	Tetraspanin6	
Glyma.20G236900	13,73	99,66	2,86	Yes	AT2G37690	Phosphoribosylaminoimidazole carboxylase, putative / AIR carboxylase	
Glyma.20G194000	2,60	16,04	2,62	Yes		Nuclear transport factor 2 (ntf2) family protein	Chalcone--flavonone isomerase
Glyma.20G210100	27,92	112,84	2,01	Yes	PTHR13683	Eukaryotic aspartyl protease family protein	
Glyma.20G206000	1.88	0.47	-2.01	Yes	PTHR26374	Tac1 (telomerase activator1)	
Glyma.20G217000	1.88	0.46	-2.02	Yes	KOG3399	Predicted Yippee-type zinc-binding protein	
Glyma.20G190100	2.58	0.63	-2.04	Yes	KOG4453	Predicted ER membrane protein	
Glyma.20G226300	44.71	10.69	-2.06	Yes	AT1G72210	Basic helix-loop-helix (bhlh) DNA-binding superfamily protein	
Glyma.20G231800	2.09	0.50	-2.07	Yes	AT5G04830	Nuclear transport factor 2 (NTF2) family protein	
Glyma.20G211100	10.87	2.57	-2.08	Yes	AT2G39000	Acyl-coa N-acyltransferases (NAT) superfamily protein	
Glyma.20G238000	25.40	6.01	-2.08	Yes	AT4G27745	Yippee family putative zinc-binding protein	
Glyma.20G248100	14.40	3.39	-2.08	Yes		Not annotated	
Glyma.20G205100	39.09	9.05	-2.11	Yes	PTHR31153	Histone h1flk-like protein-related	
Glyma.20G241300	4.67	1.08	-2.11	Yes	AT1G09010	Glycoside hydrolase family 2 protein	
Glyma.20G223200	2.81	0.64	-2.14	Yes		Not annotated	
Glyma.20G223600	10.57	2.40	-2.14	Yes	AT3G23090	TPX2 (targeting protein for Xklp2) protein family	Alpha- 1,6 - fucosyltransferase
Glyma.20G179800	1.64	0.36	-2.18	Yes	PTHR23091	Acyl-coa n-acyltransferases superfamily protein	

Glyma.20G188100	2.77	0.59	-2.24	Yes	KOG4282	Transcription factor GT-2 and related proteins, contains trihelix DNA-binding/SANT domain	
Glyma.20G216600	8.50	1.79	-2.25	Yes	AT3G23090	TPX2 (targeting protein for Xklp2) protein family	
Glyma.20G234000	1.50	0.30	-2.33	Yes	PTHR34213	Not annotated	
Glyma.20G178700	2.28	0.40	-2.50	Yes	PTHR32191:	Tetraspanin-5-related	
Glyma.20G191200	6026.27	1031.44	-2.55	Yes	PTHR22924	Non-symbiotic hemoglobin 2	Non-symbiotic haemoglobin 2
Glyma.20G187200	42.75	7.31	-2.55	Yes	PTHR24351	Ribosomal protein s6 kinase	Ribosomal protein s6 kinase
Glyma.20G184300	15.67	2.51	-2.64	Yes	PTHR16223	Transcription factor bhlh111	Transcription factor bhlh111
Glyma.20G241600	14.86	2.24	-2.73	Yes		Not annotated	
Glyma.20G239300	109.33	15.88	-2.78	Yes	AT5G04490	Vitamin E pathway gene 5	Exonuclease mut-7 homolog-related
Glyma.20G201100	1.51	0.22	-2.80	Yes		Not annotated	
Glyma.20G215000	5.04	0.63	-3.00	Yes	KOG3185	Translation initiation factor 6 (eif-6)	
Glyma.20G205400	1.69	0.13	-3.67	Yes	PTHR31358	Targeting protein for xklp2-like protein	
Glyma.20G224200	1.66	0.12	-3.79	Yes		Not annotated	
Glyma.20G192700	1.99	0.08	-4.60	Yes	PTHR11764	Lanosterol synthase	
Glyma.20G245200	3.12	0.10	-4.96	Yes		Trihydroxypterocarpan dimethylallyltransferase / Glyceollin synthase	
Glyma.20G225200	3.65	0.10	-5.19	Yes	PTHR22595	Hevein-like preproprotein	

Fig. 3.18: MapMan generated Bin maps of drought stressed nodule samples DS 60 % VWC *vs* DS 30 % VWC of significant up- and down-regulated genes (data set 2 and 4). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

Table 3.6: Differentially regulated genes comparing the stress treatments of DNS (60 %) *vs* DS 60 % VWC after significance and FDR cut-offs have been applied. Data set 5 and 6 were used in to construct the table.

Gene ID	FKPM		Log_2-fold	Significant	Gene call	Annotation	PHYTOZOME define
	DNS (60%)	DS 60%					
Glyma.14G211300	0.10	3.99	5.32	Yes	AT3G03270	Adenine nucleotide alpha hydrolases-like superfamily protein	
Glyma.15G151300	0.10	3.81	5.25	Yes		Not Annotated	
Glyma.12G149100	3.03	77.34	4.68	Yes	AT4G27410	NAC (No Apical Meristem) domain transcriptional regulator	
Glyma.06G248900	9.29	159.74	4.10	Yes	AT4G27410	NAC (No Apical Meristem) domain transcriptional regulator	
Glyma.13G279900	6.45	80.79	3.65	Yes	AT3G15500	NAC domain containing protein 3	
Glyma.10G151000	5.22	64.86	3.64	Yes	AT1G56220	Dormancy/auxin associated family protein	
Glyma.12G043500	11.76	144.18	3.62	Yes	AT1G63850	BTB/POZ domain-containing protein	
Glyma.14G209000	4.99	56.92	3.51	Yes	AT5G56550	Oxidative stress 3	Serine/threonine-protein kinase
Glyma.11G170000	3.87	43.63	3.49	Yes	AT2G30040	Mitogen-activated protein kinase kinase kinase 14	
Glyma.11G208700	16.75	173.71	3.37	Yes	AT5G56550	Oxidative stress 3	
Glyma.06G170700	2.31	23.54	3.35	Yes	AT3G48530	SNF1-related protein kinase regulatory	
Glyma.05G190100	3.61	36.17	3.32	Yes		Not Annotated	
Glyma.14G098200	11.56	112.53	3.28	Yes		Not Annotated	
Glyma.18G045200	11.55	110.25	3.25	Yes	AT5G56550	Oxidative stress 3	
Glyma.05G155600	6.38	52.92	3.05	Yes	AT1G15670	Galactose oxidase/kelch repeat superfamily protein	
Glyma.16G085600	5.83	46.79	3.00	Yes	AT3G07350	Protein of unknown function (DUF506)	
Glyma.07G139300	7.86	63.05	3.00	Yes	AT5G02020	Encodes a protein involved in salt tolerance	
Glyma.11G123400	16.05	121.33	2.92	Yes		Not Annotated	
Glyma.16G043200	7.70	53.47	2.80	Yes	AT3G04070	NAC domain containing protein 47	
Glyma.19G108800	10.17	62.39	2.62	Yes	AT3G04070	NAC domain containing protein 47	
Glyma.04G076900	6.99	41.16	2.56	Yes	AT4G31240	Protein kinase C-like zinc finger protein	Thioredoxin
Glyma.04G195300	11.58	66.30	2.52	Yes	AT3G48530	SNF1-related protein kinase regulatory subunit gamma	

Glyma.13G095000	21.08	118.06	2.49	Yes	AT4G25810	Xyloglucan endotransglycosylase 6	Secreted glucosidase-related
Glyma.11G170300	10.44	57.02	2.45	Yes	AT3G47340	Glutamine-dependent asparagine synthase 1	
Glyma.03G186300	4.97	26.00	2.39	Yes	AT3G09920	Phosphatidyl inositol monophosphate 5 kinase	
Glyma.16G011400	15.64	77.64	2.31	Yes	AT1G54070	Dormancy/auxin associated family protein	
Glyma.06G143300	109.63	19.19	-2.51	Yes	AT2G40610	Expansin A8	
Glyma.11G058100	77.72	12.68	-2.62	Yes	AT1G49570	Peroxidase superfamily protein	
Glyma.07G214000	36.53	5.70	-2.68	Yes	AT1G52950	Nucleic acid-binding, OB-fold-like protein	
Glyma.15G140200	23.75	3.31	-2.84	Yes	AT5G49460	ATP citrate lyase subunit B 2	
Glyma.16G014100	53.12	7.14	-2.89	Yes	AT3G14310	Pectin methylesterase 3	
Glyma.11G095100	151.94	20.03	-2.92	Yes	AT4G16260	Glycosyl hydrolase superfamily protein	
Glyma.12G161500	69.97	8.85	-2.98	Yes	AT5G53970	Tyrosine transaminase family protein	
Glyma.06G269700	41.23	5.00	-3.04	Yes	AT5G54770	Thiazole biosynthetic enzyme	
Glyma.02G245600	90.32	10.93	-3.05	Yes	AT1G74670	Gibberellin-regulated family protein	
Glyma.17G139700	42.26	5.00	-3.08	Yes	AT4G12490	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	
Glyma.06G275900	44.34	3.77	-3.55	Yes	AT5G05340	Peroxidase superfamily protein	
Glyma.10G200800	38.05	2.54	-3.91	Yes	AT2G45550	Cytochrome P450	
Glyma.08G329600	1.71	0.01	-7.41	Yes		Not Annotated	
Glyma.19G034500	1.86	0.01	-7.5	Yes	AT3G02310	K-box region and MADS-box transcription factor family protein	
Glyma.03G007700	2.20	0.01	-7.8	Yes	AT4G26010	Peroxidase superfamily protein	
Glyma.18G266200	3.34	0.01	-8.4	Yes	AT5G65360	Histone superfamily protein	
Glyma.04G060600	3.39	0.01	-8.4	Yes	AT2G25060	Early nodulin-like protein 14	
Glyma.10G056200	5.05	0.01	-9.0	Yes	AT3G09870	SAUR-like auxin-responsive protein family	
Glyma.14G158500	5.07	0.01	-9.0	Yes		Not Annotated	
Glyma.20G144900	5.76	0.01	-9.2	Yes		Not Annotated	
Glyma.11G134400	6.49	0.01	-9.3	Yes		Not Annotated	
Glyma.10G067600	58.50	0.01	-12.5	Yes		Not Annotated	

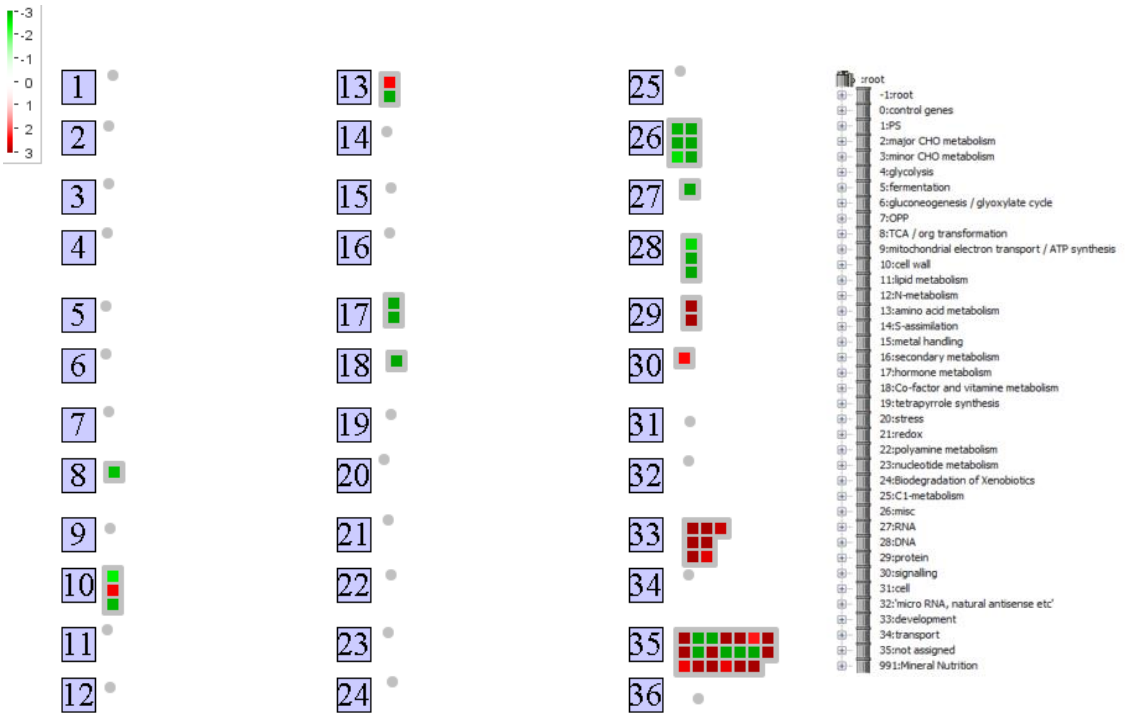


Fig. 3.19: MapMan generated Bin maps of drought stressed nodule samples DNS (60 %) VWC *vs* DS 60 % VWC of significant up- and down-regulated genes (data set 5 and 6). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

Table 3.7: Differentially regulated genes comparing the stress treatments of DNS (40 %) *vs* DS 40 % VWC after significance and FDR cut-offs have been applied. Data set 5 and 6 were used in to construct the table.

	FKPM						
Gene ID	**DNS (40%)**	**DS 40%**	**Log_2-fold**	**Significant**	**Gene call**	**Annotation**	**PHYTOZOME Define**
Glyma.13G278000	0.00	1741.52	17.41	Yes	PTHR33147	Defensin-like protein 13-related	
Glyma.09G185500	0.00	499.57	15.61	Yes	PF00257	Dehydrin (dehydrin	
Glyma.05G112000	0.00	225.19	14.46	Yes	AT1G32560	Late embryogenesis abundant protein, group 1 protein	
Glyma.06G220800	0.00	183.34	14.16	Yes	AT5G20230	Blue-copper-binding protein	
Glyma.14G145900	0.00	159.55	13.96	Yes	AT5G11580	Regulator of chromosome condensation (RCC1) family protein	
Glyma.05G126800	0.00	135.42	13.73	Yes	PF07712	Stress up-regulated Nod 19	
Glyma.08G156600	0.00	94.67	13.21	Yes		Not annotated	
Glyma.20G158500	0.00	91.29	13.16	Yes	AT5G52300	CAP160 protein	Serine/threonine protein kinase
Glyma.14G145800	0.00	70.28	12.78	Yes	AT5G11580	Regulator of chromosome condensation (RCC1) family protein	
Glyma.02G062900	0.00	70.12	12.78	Yes	AT2G23620	Methyl esterase 1	
Glyma.15G145600	0.00	69.02	12.75	Yes	PF00407	Pathogenesis-related protein Bet v I family (Bet_v_1)	
Glyma.18G033200	0.00	43.84	12.10	Yes	AT3G53980	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	
Glyma.11G180800	0.00	40.21	11.97	Yes	KOG0725	Reductases with broad range of substrate specificities	
Glyma.06G166800	0.00	40.08	11.97	Yes	AT5G23660	Homolog of Medicago truncatula MTN3	
Glyma.06G077000	0.00	40.07	11.97	Yes		Not annotated	
Glyma.13G324400	0.00	38.60	11.91	Yes	PTHR18952	Carbonic anhydrase	
Glyma.17G155000	0.00	37.26	11.86	Yes	AT1G32560	Late embryogenesis abundant protein	
Glyma.07G154700	0.00	36.40	11.83	Yes		Not annotated	
Glyma.03G056000	0.00	33.61	11.71	Yes	AT2G40170	Stress induced protein	Em-like protein gea6
Glyma.20G199300	0.00	32.80	11.68	Yes	AT4G13480	Myb domain protein	
Glyma.07G090400	0.00	29.38	11.52	Yes	PF00257	Dehydrin (dehydrin)	

Glyma.12G235800	0.00	29.20	11.51	Yes	PF00257	Dehydrin (dehydrin)
Glyma.02G216200	0.00	28.89	11.50	Yes		Not annotated
Glyma.08G210100	0.00	27.98	11.45	Yes		Not annotated
Glyma.04G208700	0.00	26.79	11.39	Yes		Not annotated
Glyma.05G065300	0.00	25.48	11.32	Yes	AT4G17030	Expansin-like B1
Glyma.13G081100	0.00	25.31	11.31	Yes	PF00892	Eama-like transporter family (eama)
Glyma.10G061800	0.00	23.61	11.21	Yes	AT1G20030	Pathogenesis-related thaumatin superfamily protein
Glyma.12G049100	0.00	23.22	11.18	Yes	PTHR22595	Chitinase-related
Glyma.06G157000	0.00	22.87	11.16	Yes		Not annotated
Glyma.17G147200	0.00	21.97	11.10	Yes	AT4G17030	Expansin-like B1
Glyma.16G038100	0.00	20.68	11.01	Yes		Not annotated
Glyma.06G111600	0.00	20.34	10.99	Yes	AT4G08290	Nodulin mtn21 /eama-like transporter family protein
Glyma.05G199300	0.00	20.17	10.98	Yes	AT5G07330	Not annotated
Glyma.14G186400	0.00	20.09	10.97	Yes		Not annotated
Glyma.18G153800	0.00	19.82	10.95	Yes	AT5G65660	Hydroxyproline-rich glycoprotein family protein
Glyma.10G215700	0.00	19.57	10.93	Yes	AT5G54160	O-methyltransferase 1
Glyma.20G201800	0.00	19.20	10.91	Yes	AT2G38870	Serine protease inhibitor, potato inhibitor I-type family protein
Glyma.11G196600	0.00	17.84	10.80	Yes	PTHR18952	Carbonic anhydrase
Glyma.18G273300	0.00	17.66	10.79	Yes	AT3G30210	Myb domain protein
Glyma.17G232800	0.00	17.35	10.76	Yes		Not annotated
Glyma.06G270500	0.00	16.77	10.71	Yes	AT4G27360	Dynein light chain type 1 family protein
Glyma.02G219100	0.00	16.09	10.65	Yes		Not annotated
Glyma.10G205900	0.00	15.55	10.60	Yes	AT1G04560	AWPM-19-like family protein
Glyma.12G230100	0.00	15.51	10.60	Yes		Not annotated
Glyma.07G043600	0.00	15.41	10.59	Yes	AT1G54050	HSP20-like chaperones superfamily protein
Glyma.19G025600	0.00	14.80	10.53	Yes	AT5G39530	Protein of unknown function
Glyma.19G189300	0.00	13.92	10.44	Yes	AT5G03795	Exostosin family protein
Glyma.09G005000	0.00	13.09	10.35	Yes		Not annotated

Glyma.10G176400	0.00	12.28	10.26	Yes	AT4G10250	HSP20-like chaperones superfamily protein	
Glyma.09G070500	0.00	12.22	10.26	Yes	AT3G51420	Strictosidine synthase-like	
Glyma.06G319700	0.00	12.16	10.25	Yes	AT1G33590	Leucine-rich repeat (LRR) family protein	
Glyma.08G342100	0.00	11.78	10.20	Yes	AT1G73260	Kunitz trypsin inhibitor 1	
Glyma.11G151500	0.00	11.76	10.20	Yes		Catechol oxidase / Tyrosinase	
Glyma.10G028900	0.00	11.63	10.18	Yes	AT3G51895	Sulfate transporter	
Glyma.18G003700	0.00	10.80	10.08	Yes	AT2G31980	Phytocystatin 2	
Glyma.16G212200	0.00	10.35	10.02	Yes	AT1G17860	Kunitz family trypsin and protease inhibitor protein	
Glyma.08G305000	0.00	10.34	10.01	Yes	AT2G42660	Homeodomain-like superfamily protein	
Glyma.12G237100	0.00	10.25	10.00	Yes		Not annotated	
Glyma.08G069000	0.00	10.18	9.99	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.08G068700	0.00	10.04	9.97	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.08G172800	0.00	9.90	9.95	Yes		Not annotated	
Glyma.07G139100	0.00	9.78	9.93	Yes	AT2G37900	Major facilitator superfamily protein	Oligopeptide transporter-related
Glyma.08G212000	0.00	9.70	9.92	Yes	AT1G52560	HSP20-like chaperones superfamily protein	
Glyma.13G162700	0.00	8.94	9.80	Yes		Not annotated	
Glyma.08G068800	0.00	8.76	9.77	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.09G150400	0.00	8.55	9.74	Yes	AT5G10530	Concanavalin A-like lectin protein kinase family protein	
Glyma.03G238800	0.00	8.40	9.71	Yes	AT1G54870	NAD(P)-binding Rossmann-fold superfamily protein	Glucose and ribitol dehydrogenase homolog 1-related
Glyma.20G167500	0.00	8.18	9.68	Yes	PF02496	ABA/WDS induced protein (ABA_WDS)	
Glyma.17G048500	0.00	8.12	9.67	Yes	AT2G05910	Protein of unknown function	
Glyma.16G180700	0.00	7.83	9.61	Yes	AT1G02260	Divalent ion symporter	Sugar transporter
Glyma.19G210900	0.00	7.79	9.61	Yes	AT1G02860	SPX (SYG1/Pho81/XPR1) domain-containing protein	
Glyma.06G048700	0.00	7.76	9.60	Yes		Not annotated	
Glyma.09G142200	0.00	7.39	9.53	Yes	AT3G48270	Cytochrome P450	
Glyma.09G060800	0.00	7.32	9.52	Yes		Not annotated	

Glyma.04G099200	0.00	7.21	9.49	Yes		Not annotated	
Glyma.11G059100	0.00	7.13	9.48	Yes	AT2G23640	Reticulan like protein B13	Zinc finger fyve domain containing protein
Glyma.17G011100	0.00	6.96	9.44	Yes	AT1G11925	Stigma-specific Stig1 family protein	
Glyma.18G273400	0.00	6.91	9.43	Yes	AT3G30210	Myb domain protein	
Glyma.15G077700	0.00	6.87	9.42	Yes	AT2G20560	Heat shock family protein	
Glyma.07G061500	0.00	6.86	9.42	Yes	AT2G46240	BCL-2-associated athanogene	
Glyma.18G203500	0.00	6.77	9.40	Yes	AT3G51810	Stress induced protein	
Glyma.12G200700	0.00	6.54	9.35	Yes	PF12609	Wound-induced protein	
Glyma.20G164200	0.00	6.05	9.24	Yes	AT5G24080	Protein kinase superfamily protein	
Glyma.13G219300	0.00	6.00	9.23	Yes		Not annotated	
Glyma.05G227200	0.00	5.91	9.21	Yes	AT2G34440	AGAMOUS-like 29	Mads box protein
Glyma.05G124000	0.00	5.70	9.16	Yes	AT5G06860	Polygalacturonase inhibiting protein 1	Serine-threonine protein kinase, plant-type
Glyma.07G237300	0.00	5.60	9.13	Yes	AT1G47960	Cell wall / vacuolar inhibitor of fructosidase	
Glyma.14G100000	0.00	5.45	9.09	Yes	AT5G12020	Heat shock protein	
Glyma.10G223100	0.00	5.31	9.05	Yes		Not annotated	
Glyma.06G220200	0.00	5.12	9.00	Yes	AT2G24130	Leucine-rich receptor-like protein kinase family protein	Serine-threonine protein kinase, plant-type
Glyma.09G115200	0.00	4.96	8.95	Yes	AT1G68390	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	
Glyma.06G249400	0.00	4.65	8.86	Yes	AT5G54400	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	
Glyma.08G250600	0.00	4.62	8.85	Yes	AT3G30210	Myb domain protein	
Glyma.06G167900	0.00	4.55	8.83	Yes	AT5G50720	HVA22 homologue E	
Glyma.14G051800	0.00	4.55	8.83	Yes	PTHR10891	Ef-hand calcium-binding domain containing protein	
Glyma.19G147200	0.00	4.39	8.78	Yes	AT5G06760	Late embryogenesis abundant 4-5	
Glyma.07G200500	0.00	4.38	8.77	Yes	AT1G07400	HSP20-like chaperones superfamily protein	

Glyma.13G042400	0.00	4.34	8.76	Yes	PTHR33735	Expressed protein	
Glyma.10G243800	0.00	4.29	8.74	Yes	AT3G22840	Chlorophyll A-B binding family protein	
Glyma.07G186000	0.00	4.27	8.74	Yes	AT1G50090	D-aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein	
Glyma.U035800	0.00	4.25	8.73	Yes	AT4G20970	Basic helix-loop-helix (bhlh)	
Glyma.13G242300	0.00	4.06	8.66	Yes	PTHR11474	5,6-dihydroxyindole-2-carboxylic acid oxidase	
Glyma.18G129800	0.00	4.04	8.66	Yes		Not annotated	
Glyma.11G121900	0.00	3.99	8.64	Yes		Not annotated	
Glyma.05G191100	0.00	3.94	8.62	Yes	PTHR11601	Cysteine desulfurylase	
Glyma.13G104500	0.00	3.93	8.62	Yes		Not annotated	
Glyma.16G212100	0.00	3.82	8.58	Yes	AT1G17860	Kunitz family trypsin and protease inhibitor protein	
Glyma.09G237600	0.00	3.79	8.57	Yes	AT1G32450	Nitrate transporter	
Glyma.17G151700	0.00	3.78	8.56	Yes	AT4G17030	Expansin-like B1	
Glyma.04G249600	0.00	3.77	8.56	Yes		Not annotated	
Glyma.04G039900	0.00	3.73	8.54	Yes	AT5G42830	HXXXD-type acyl-transferase family protein	
Glyma.06G183800	0.00	3.71	8.53	Yes	AT1G10560	Plant U-box 18	U-box domain-containing protein 18-related
Glyma.02G029900	0.00	3.67	8.52	Yes	AT1G24100	UDP-glucosyl transferase	
Glyma.15G198500	0.00	3.38	8.40	Yes		Not annotated	
Glyma.11G179100	0.00	3.29	8.36	Yes	PTHR31086	Aluminum-activated malate transporter 1-related	
Glyma.08G042000	0.00	3.28	8.36	Yes	AT1G73830	BR enhanced expression 3	Sterol regulatory element-binding protein
Glyma.10G190900	0.00	3.28	8.36	Yes	PF03330	Rare lipoprotein A (rlpa)-like double-psi beta-barrel	
Glyma.13G174300	0.00	3.27	8.35	Yes	PTHR13301	X-box transcription factor-related	
Glyma.02G123500	0.00	3.26	8.35	Yes		Not annotated	
Glyma.11G059800	0.00	3.25	8.34	Yes	AT2G23540	GDSL-like Lipase/Acylhydrolase superfamily protein	
Glyma.10G064400	0.00	3.19	8.32	Yes	AT2G36640	Embryonic cell protein 63	Late embryogenesis abundant (plants) lea-related

Glyma.04G179400	0.00	3.18	8.31	Yes	AT5G50100	Putative thiol-disulphide oxidoreductase DCC	
Glyma.09G163600	0.00	3.16	8.30	Yes	AT1G17860	Kunitz family trypsin and protease inhibitor protein	
Glyma.13G175700	0.00	3.16	8.30	Yes	PTHR11527	Small heat-shock protein (hsp20) family	
Glyma.U020300	0.00	3.09	8.27	Yes	AT2G28490	Rmlc-like cupins superfamily protein	
Glyma.08G306500	0.00	3.08	8.27	Yes	AT1G78990	HXXXD-type acyl-transferase family protein	
Glyma.06G178100	0.00	2.98	8.22	Yes		Not annotated	
Glyma.13G243200	0.00	2.97	8.21	Yes	PF02365	No apical meristem (NAM) protein (NAM)	
Glyma.16G063500	0.00	2.92	8.19	Yes	AT4G12560	F-box and associated interaction domains-containing protein	
Glyma.09G224500	0.00	2.88	8.17	Yes	AT5G66460	Glycosyl hydrolase superfamily protein	
Glyma.03G029100	0.00	2.77	8.12	Yes	AT2G45220	Plant invertase/pectin methylesterase inhibitor superfamily	
Glyma.11G018000	0.00	2.77	8.11	Yes	AT2G29380	Highly ABA-induced PP2C gene 3	
Glyma.06G174800	0.00	2.75	8.10	Yes	AT5G50260	Cysteine proteinases superfamily protein	
Glyma.12G174900	0.00	2.72	8.09	Yes	PTHR31223	Cytokinin riboside 5'-monophosphate phosphoribohydrolase log1	
Glyma.14G156300	0.00	2.67	8.06	Yes	AT1G21550	Calcium-binding EF-hand family protein	
Glyma.13G302400	0.00	2.59	8.02	Yes	PTHR10641	Myb-related	
Glyma.10G010000	0.00	2.57	8.01	Yes	AT1G02460	Pectin lyase-like superfamily protein	
Glyma.09G142500	0.00	2.55	7.99	Yes	AT3G48270	Cytochrome P450	
Glyma.01G143500	0.00	2.50	7.96	Yes	AT5G62020	Heat shock transcription factor B2A	
Glyma.20G171900	0.00	2.49	7.96	Yes	AT1G14550	Peroxidase superfamily protein	
Glyma.12G001600	0.00	2.46	7.94	Yes	PTHR23241	Late embryogenesis abundant (plants) *LEA*-related	
Glyma.16G141000	0.00	2.34	7.87	Yes	AT2G18480	Major facilitator superfamily protein	
Glyma.01G175500	0.00	2.34	7.87	Yes	AT3G50660	Cytochrome P450 superfamily protein	
Glyma.05G107600	0.00	2.27	7.82	Yes	AT5G65890	ACT domain repeat 1	
Glyma.20G117000	0.00	2.25	7.81	Yes	AT1G68320	Myb domain protein	Zinc finger, c3hc4 type
Glyma.03G005600	0.00	2.24	7.81	Yes	AT1G33110	MATE efflux family protein	
Glyma.08G241000	0.00	2.22	7.80	Yes	AT3G29200	Chorismate mutase 1	
Glyma.14G077900	0.00	2.20	7.78	Yes	AT1G47510	Inositol polyphosphate 5-phosphatase	
Glyma.10G027600	0.00	2.14	7.74	Yes	AT3G22490	Seed maturation protein	

Glyma.09G139600	0.00	2.12	7.73	Yes	AT5G52300	CAP160 protein	
Glyma.05G248100	0.00	2.10	7.72	Yes	AT1G73480	Alpha/beta-Hydrolases superfamily protein	
Glyma.17G165500	0.00	2.07	7.70	Yes	AT4G37390	Auxin-responsive GH3 family protein	
Glyma.03G189200	0.00	2.05	7.68	Yes	AT2G36640	Embryonic cell protein 63	Late embryogenesis abundant
Glyma.18G138000	0.00	2.05	7.68	Yes	AT1G70370	Polygalacturonase 2	
Glyma.19G175200	0.00	2.04	7.67	Yes	AT3G09520	Exocyst subunit exo70 family protein H4	
Glyma.10G236000	0.00	2.03	7.66	Yes	AT5G52300	CAP160 protein	
Glyma.17G180000	0.00	2.01	7.65	Yes	AT5G66900	Disease resistance protein (CC-NBS-LRR class) family	
Glyma.08G109100	0.00	1.98	7.63	Yes	AT3G23820	UDP-D-glucuronate 4-epimerase 6	
Glyma.11G125200	0.00	1.97	7.62	Yes	AT3G06160	AP2/B3-like transcriptional factor family protein	
Glyma.17G244000	0.00	1.75	7.45	Yes	AT2G26560	Phospholipase A 2A	
Glyma.08G321700	0.00	1.71	7.42	Yes		Not annotated	
Glyma.18G258100	0.00	1.63	7.35	Yes	AT3G49190	O-acyltransferase (WSD1-like) family protein	
Glyma.03G216000	0.00	1.57	7.29	Yes	AT1G02810	Plant invertase/pectin methylesterase inhibitor superfamily	
Glyma.20G138500	0.00	1.57	7.29	Yes	AT4G21410	Cysteine-rich RLK (RECEPTOR-like protein kinase)	
Glyma.17G147500	4.98	775.09	7.28	Yes	AT4G17030	Expansin-like B1	
Glyma.08G225800	0.00	1.55	7.27	Yes	AT1G78955	Camelliol C synthase 1	
Glyma.14G188000	0.00	1.55	7.27	Yes	AT5G43360	Phosphate transporter 1;3	
Glyma.19G027400	0.00	1.53	7.25	Yes	AT3G01060	Not annotated	
Glyma.07G030500	0.00	1.19	6.90	Yes	AT5G20190	Tetratricopeptide repeat (TPR)-like superfamily protein	
Glyma.13G208000	4.26	372.95	6.45	Yes	AT1G03220	Eukaryotic aspartyl protease family protein	Aspartyl protease
Glyma.15G071200	3.74	279.74	6.22	Yes	AT1G30700	FAD-binding Berberine family protein	
Glyma.02G148200	4.22	287.38	6.09	Yes	AT1G03220	Eukaryotic aspartyl protease family protein	
Glyma.11G198500	5.08	295.94	5.86	Yes	AT1G78380	Glutathione S-transferase	
Glyma.12G234800	3.30	184.16	5.80	Yes	AT1G17860	Kunitz family trypsin and protease inhibitor protein	
Glyma.18G205100	2.49	100.53	5.33	Yes	AT5G66770	GRAS family transcription factor	
Glyma.05G003900	3.03	119.80	5.31	Yes	AT5G40390	Raffinose synthase family protein	
Glyma.10G192900	6.83	263.43	5.27	Yes	AT3G09270	Glutathione S-transferase	

Glyma.10G145300	4.80	184.06	5.26	Yes	AT2G47180	Galactinol synthase	Inositol 3-alpha-galactosyl transferase
Glyma.10G180500	3.32	106.86	5.01	Yes	AT1G03220	Eukaryotic aspartyl protease family protein	Aspartyl proteases
Glyma.20G148900	2.73	84.96	4.96	Yes	AT3G22740	Homocysteine S-methyltransferase 3	
Glyma.06G021200	2.88	88.60	4.94	Yes	AT1G60420	DC1 domain-containing protein	Thioredoxin
Glyma.11G025600	6.54	194.16	4.89	Yes	AT4G11650	Osmotin	
Glyma.15G024600	2.02	58.24	4.85	Yes	AT3G13790	Glycosyl hydrolases family 32 protein	
Glyma.07G243500	23.02	635.90	4.79	Yes	PF00407	Pathogenesis-related protein Bet v I family (Bet_v_1)	
Glyma.20G150600	5.98	162.44	4.76	Yes	AT3G22840	Chlorophyll A-B binding family protein	
Glyma.15G134300	5.04	134.38	4.74	Yes	AT1G30700	FAD-binding Berberine family protein	
Glyma.11G051800	2.75	70.78	4.68	Yes	AT4G37340	Cytochrome P450	
Glyma.14G162100	3.61	72.86	4.33	Yes	AT2G29380	Highly ABA-induced PP2C gene	
Glyma.13G316100	6.65	116.82	4.13	Yes	AT2G27480	Calcium-binding EF-hand family protein	Calpain
Glyma.15G222200	2.03	0.00	-4.34	Yes		Not annotated	
Glyma.13G305100	2.40	0.00	-4.59	Yes	AT5G06860	Polygalacturonase inhibiting protein 1	
Glyma.14G177600	2.58	0.00	-4.69	Yes	AT5G65360	Histone superfamily protein	Serine-threonine protein kinase
Glyma.10G180100	2.59	0.00	-4.70	Yes		Fla (fasciclin-like arabinogalactan protein)	
Glyma.12G069300	2.92	0.00	-4.87	Yes	PTHR10984	Cell redox homeostasis protein-related	
Glyma.20G205700	3.04	0.00	-4.93	Yes	AT1G74670	Gibberellin-regulated family protein	
Glyma.19G142700	3.69	0.00	-5.21	Yes	AT5G03170	FASCICLIN-like arabinogalactan-protein 11	
Glyma.03G131700	3.91	0.00	-5.29	Yes	AT4G14550	Indole-3-acetic acid inducible 14	
Glyma.17G169200	4.27	0.00	-5.41	Yes	PTHR31444:	Protein irregular xylem 15-related	
Glyma.13G288600	4.53	0.00	-5.50	Yes	AT1G72230	Cupredoxin superfamily protein	
Glyma.08G169100	5.90	0.00	-5.88	Yes		Not annotated	
Glyma.15G209300	6.56	0.00	-6.04	Yes	PF00560	Leucine rich repeat (lrr_1)	
Glyma.11G134500	7.38	0.00	-6.21	Yes		Not annotated	
Glyma.U006400	7.98	0.00	-6.32	Yes		Not annotated	

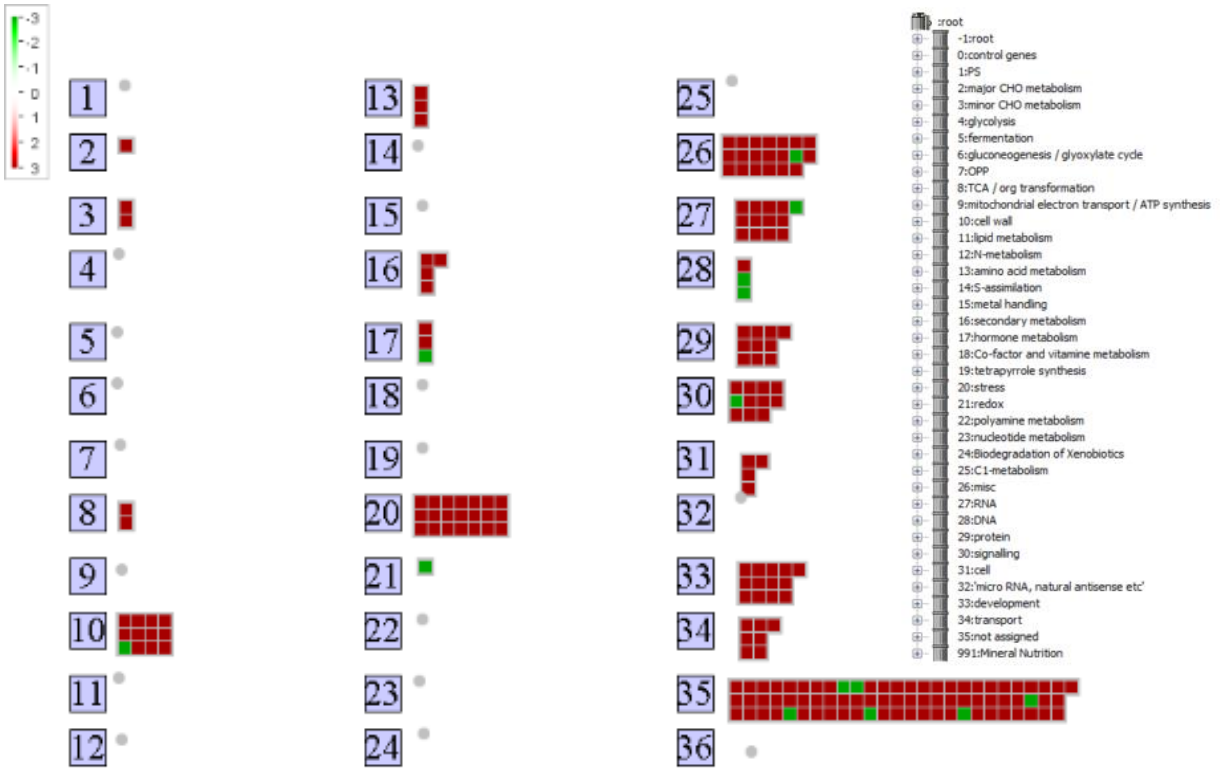


Fig. 3.20: MapMan generated Bin maps of drought stressed nodule samples DNS (40 %) VWC *vs* DS 40 % VWC of significant up- and down-regulated genes (data set 5 and 6). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

Table 3.8: Differentially regulated genes comparing the stress treatments of DNS (30 %) *vs* DS 30 % VWC after significance and FDR cut-offs have been applied. Data set 5 and 6 were used in to construct the table.

Gene ID	FKPM		Log_2-fold	Significant	Gene call	Annotation	PHYTOZOME Define
	DNS (30%)	DS 30%					
Glyma.08G160800	0.00	232.31	11.18	Yes		Not annotated	
Glyma.08G156600	0.00	212.98	11.06	Yes		Not annotated	
Glyma.08G068700	0.00	165.03	10.69	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.08G172800	0.00	147.28	10.52	Yes		Not annotated	
Glyma.06G220800	0.00	137.97	10.43	Yes	AT5G20230	Blue-copper-binding protein	
Glyma.17G224900	0.00	133.84	10.39	Yes	AT5G12020	Heat shock protein	
Glyma.08G068800	0.00	116.31	10.18	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.06G077000	0.00	109.96	10.10	Yes		Not annotated	
Glyma.07G154700	0.00	103.20	10.01	Yes		Not annotated	
Glyma.10G176400	0.00	102.55	10.00	Yes	AT4G10250	HSP20-like chaperones superfamily protein	
Glyma.02G216200	0.00	87.18	9.77	Yes		Not annotated	
Glyma.06G157000	0.00	86.36	9.75	Yes		Not annotated	
Glyma.07G200500	0.00	84.04	9.71	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.03G144400	0.00	80.68	9.66	Yes	AT5G06760	Late embryogenesis abundant 4-5	
Glyma.03G056000	0.00	80.67	9.66	Yes	AT2G40170	Stress induced protein	
Glyma.07G200700	0.00	78.57	9.62	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.07G043600	0.00	76.62	9.58	Yes	AT1G54050	HSP20-like chaperones superfamily protein	
Glyma.08G069000	0.00	72.66	9.50	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.04G208700	0.00	71.61	9.48	Yes		Not annotated	
Glyma.18G203500	0.00	65.33	9.35	Yes	AT3G51810	Stress induced protein	Short-chain dehydrogenases/reductase
Glyma.11G180800	0.00	64.31	9.33	Yes	AT1G52340	NAD(P)-binding Rossmann-fold superfamily protein	
Glyma.14G063800	0.00	63.11	9.30	Yes	AT1G53540	HSP20-like chaperones superfamily protein	

Glyma.20G201800	0.00	47.47	8.89	Yes		Not annotated	
Glyma.12G235800	0.00	45.65	8.83	Yes	AT4G39130	Dehydrin family protein	
Glyma.18G273300	0.00	45.22	8.82	Yes	AT3G30210	Myb domain protein	
Glyma.14G100000	0.00	43.74	8.77	Yes	AT5G12020	Heat shock protein	
Glyma.14G134900	0.00	41.91	8.71	Yes		Not annotated	
Glyma.16G030000	0.00	41.88	8.71	Yes	AT2G46240	BCL-2-associated athanogene 6	
Glyma.12G205700	0.00	41.01	8.68	Yes	AT1G03790	Zinc finger C-x8-C-x5-C-x3-H type family protein	
Glyma.10G236000	0.00	39.12	8.61	Yes	AT5G52300	CAP160 protein	
Glyma.14G099900	0.00	38.20	8.58	Yes	AT5G12020	Heat shock protein	
Glyma.11G151500	0.00	31.26	8.29	Yes		Not annotated	
Glyma.02G219100	0.00	30.89	8.27	Yes		Not annotated	
Glyma.17G147600	0.00	29.39	8.20	Yes	AT4G17030	Expansin-like B1	
Glyma.06G270500	0.00	28.11	8.13	Yes	AT4G27360	Dynein light chain type 1 family protein	
Glyma.08G359500	0.00	25.87	8.02	Yes		Not annotated	
Glyma.18G121000	0.00	25.59	8.00	Yes		Not annotated	
Glyma.15G077700	0.00	25.58	8.00	Yes	AT2G20560	DNAJ heat shock family protein	
Glyma.13G138300	0.00	25.06	7.97	Yes	AT2G37130	Peroxidase superfamily protein	
Glyma.12G067100	0.00	24.38	7.93	Yes	AT5G06900	Cytochrome P450,	
Glyma.01G141900	0.00	24.30	7.92	Yes		Not annotated	
Glyma.18G129800	0.00	24.11	7.91	Yes		Not annotated	
Glyma.14G186400	0.00	22.73	7.83	Yes		Not annotated	
Glyma.18G205400	0.00	22.71	7.83	Yes		Not annotated	
Glyma.12G237100	0.00	20.61	7.69	Yes		Not annotated	
Glyma.03G238900	0.00	20.04	7.65	Yes	AT1G54870	NAD(P)-binding Rossmann-fold superfamily protein	
Glyma.17G040800	0.00	19.18	7.58	Yes	AT2G18340	Late embryogenesis abundant domain-containing protein	
Glyma.11G196600	0.00	18.94	7.56	Yes	AT1G08080	Alpha carbonic anhydrase 7	Alcohol dehydrogenase related
Glyma.01G021000	0.00	18.89	7.56	Yes	AT4G37990	Elicitor-activated gene 3-2	

Glyma.07G200000	0.00	17.71	7.47	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.16G206200	0.00	17.25	7.43	Yes	AT4G10250	HSP20-like chaperones superfamily protein	
Glyma.13G209600	0.00	17.11	7.42	Yes	PTHR16007	Epididymal membrane protein e9-related	
Glyma.01G222600	0.00	17.10	7.42	Yes	PTHR10233	Translation initiation factor eif-2b	
Glyma.01G036000	0.00	16.46	7.36	Yes	AT1G24100	UDP-glucosyl transferase 74B1	
Glyma.07G200300	0.00	15.97	7.32	Yes	AT1G07400	HSP20-like chaperones superfamily protein	
Glyma.01G119600	0.00	15.29	7.26	Yes	AT2G40170	Stress induced protein	
						Hydrogen ion transporting ATP synthases, rotational mechanism;zinc	
Glyma.17G185600	0.00	15.00	7.23	Yes	ATMG00640	ion binding	
Glyma.18G278700	0.00	13.63	7.09	Yes	AT5G53820	Late embryogenesis abundant protein (*LEA)* family protein	
Glyma.16G212200	0.00	13.11	7.03	Yes	AT1G17860	Kunitz family trypsin and protease inhibitor protein	
Glyma.11G067900	0.00	12.97	7.02	Yes		Not annotated	
Glyma.09G060800	0.00	12.91	7.01	Yes		Not annotated	
Glyma.08G250600	0.00	12.28	6.94	Yes	AT3G30210	Myb domain protein 121	
Glyma.03G189200	0.00	11.63	6.86	Yes	AT2G36640	Embryonic cell protein 63	Late embryogenesis abundant (plants)
Glyma.14G167800	0.00	11.55	6.85	Yes		Not annotated	
Glyma.08G239400	0.00	11.51	6.85	Yes	AT1G22600	Late embryogenesis abundant protein (*LEA*)	
Glyma.11G194800	0.00	11.20	6.81	Yes	AT2G28085	SAUR-like auxin-responsive protein family	
Glyma.05G107600	0.00	11.08	6.79	Yes	AT5G65890	ACT domain repeat	
Glyma.11G059800	0.00	10.59	6.73	Yes	AT2G23540	GDSL-like Lipase/Acylhydrolase superfamily protein	Zinc finger fyve domain containing protein
Glyma.09G067100	0.00	10.13	6.66	Yes		Not annotated	
Glyma.18G273400	0.00	10.02	6.65	Yes	AT3G30210	Myb domain protein	
Glyma.13G119400	0.00	9.86	6.62	Yes	AT3G53040	Late embryogenesis abundant protein, putative	
Glyma.11G121900	0.00	9.83	6.62	Yes		Not annotated	
Glyma.13G202600	0.00	9.61	6.59	Yes		Not annotated	
Glyma.04G082300	0.00	9.48	6.57	Yes	AT4G32770	Tocopherol cyclase, chloroplast / vitamin E deficient 1	

Glyma.20G164200	0.00	9.36	6.55	Yes	AT5G24080	Protein kinase superfamily protein	
Glyma.15G073100	0.00	9.16	6.52	Yes	AT4G27450	Aluminium induced protein with YGL and LRDR motifs	Asparagine synthetase
Glyma.18G003700	0.00	9.14	6.51	Yes	AT2G31980	Phytocystatin 2	
Glyma.13G186400	0.00	9.00	6.49	Yes	AT5G13750	Zinc induced facilitator-like 1	
Glyma.19G025600	0.00	8.98	6.49	Yes		Not annotated	
Glyma.10G223100	0.00	8.91	6.48	Yes		Not annotated	
Glyma.16G038000	0.00	8.47	6.40	Yes		Not annotated	
Glyma.15G117600	0.00	8.29	6.37	Yes		Not annotated	
Glyma.02G252800	0.00	8.27	6.37	Yes	AT1G53540	HSP20-like chaperones superfamily protein	
Glyma.20G013600	0.00	8.03	6.33	Yes	AT1G71000	Chaperone dnaj-domain superfamily protein	
Glyma.10G064400	0.00	7.87	6.30	Yes	AT2G36640	Embryonic cell protein 63	Late embryogenesis abundant (plants)
Glyma.19G190400	0.00	7.83	6.29	Yes	PTHR12626	Programmed cell death 4	
Glyma.19G159000	0.00	7.69	6.26	Yes	AT3G51895	Sulfate transporter	
Glyma.03G238800	0.00	7.47	6.22	Yes	AT1G54870	NAD(P)-binding Rossmann-fold superfamily protein	Short-chain dehydrogenases/reductase
Glyma.16G141000	0.00	7.44	6.22	Yes	AT2G18480	Major facilitator superfamily protein	Sugar transporter
Glyma.15G138300	0.00	7.33	6.20	Yes		Not annotated	
Glyma.04G198600	0.00	7.10	6.15	Yes	AT5G50800	Nodulin mtn3 family protein	
Glyma.07G015200	0.00	6.99	6.13	Yes	AT3G16500	Phytochrome-associated protein 1	
Glyma.16G039600	0.00	6.79	6.08	Yes	AT2G19070	Spermidine hydroxycinnamoyl transferase	
Glyma.13G247200	0.00	6.78	6.08	Yes	AT4G21440	MYB-like 102	
Glyma.06G078700	0.00	6.49	6.02	Yes	AT2G25890	Oleosin family protein	
Glyma.01G181600	0.00	6.25	5.96	Yes	AT2G23540	GDSL-like Lipase/Acylhydrolase superfamily protein	
Glyma.13G235100	0.00	6.05	5.92	Yes		Not annotated	
Glyma.08G342100	0.00	6.00	5.91	Yes	AT1G73260	Kunitz trypsin inhibitor 1	
Glyma.19G009900	0.00	5.93	5.89	Yes	AT5G62850	Nodulin mtn3 family protein	
Glyma.03G254900	0.00	5.68	5.83	Yes	AT4G01130	GDSL-like Lipase/Acylhydrolase superfamily protein	
Glyma.05G161100	0.00	5.64	5.82	Yes		Not annotated	

Glyma.13G105700	0.00	5.62	5.81	Yes	AT2G26150	Heat shock transcription factor A2	
Glyma.10G243800	0.00	5.58	5.80	Yes	AT3G22840	Chlorophyll A-B binding family protein	
Glyma.10G214600	0.00	5.57	5.80	Yes	AT5G61660	Glycine-rich protein	
Glyma.06G183800	0.00	5.40	5.75	Yes	AT1G10560	Plant U-box 18	
Glyma.04G099200	0.00	5.25	5.71	Yes		Not annotated	
Glyma.17G186400	0.00	5.23	5.71	Yes	ATCG00710	Photosystem II reaction center protein H	
Glyma.09G132200	0.00	5.20	5.70	Yes	AT4G25700	Beta-hydroxylase 1	
Glyma.04G179400	0.00	4.98	5.64	Yes	AT5G50100	Putative thiol-disulphide oxidoreductase DCC	
Glyma.02G131400	0.00	4.95	5.63	Yes	PF04043	Plant invertase/pectin methylesterase inhibitor (PMEI)	
Glyma.05G065300	0.00	4.79	5.58	Yes	AT4G17030	Expansin-like B1	
Glyma.13G302400	0.00	4.58	5.52	Yes	AT4G21440	MYB-like 102	
Glyma.06G255300	0.00	4.55	5.51	Yes	AT4G27310	B-box type zinc finger family protein	
Glyma.17G076100	0.00	4.48	5.49	Yes	AT4G19810	Glycosyl hydrolase family protein with chitinase insertion domain	
Glyma.14G156300	0.00	4.33	5.44	Yes	AT1G21550	Calcium-binding EF-hand family protein	Chitinase
Glyma.16G162400	0.00	4.28	5.42	Yes	AT1G34060	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein	
Glyma.08G109100	0.00	4.21	5.39	Yes	AT3G23820	UDP-D-glucuronate 4-epimerase 6	
Glyma.15G057700	0.00	4.20	5.39	Yes	AT4G33580	Beta carbonic anhydrase 5	Nad dependent epimerase/dehydratase
Glyma.05G156300	0.00	4.15	5.38	Yes	AT4G25650	ACD1-like	
Glyma.10G027600	0.00	4.10	5.36	Yes	AT3G22490	Seed maturation protein	Iron-sulfur domain containing protein
Glyma.12G014400	0.00	4.08	5.35	Yes	AT1G64980	Nucleotide-diphospho-sugar transferases superfamily protein	
Glyma.13G326700	0.00	4.00	5.32	Yes	AT3G53410	RING/U-box superfamily protein	
Glyma.02G255900	0.00	3.86	5.27	Yes	PTHR34271	Nucleolar histone methyltransferase-related protein	
Glyma.14G188000	0.00	3.84	5.26	Yes	AT5G43360	Phosphate transporter 1;3	
Glyma.13G152100	0.00	3.79	5.25	Yes	AT5G03670	Not annotated	Sugar transporter
Glyma.12G199600	0.00	3.76	5.23	Yes	AT4G21440	MYB-like 102	

Glyma.04G198400	0.00	3.75	5.23	Yes	AT5G50800	Nodulin mtn3 family protein	
Glyma.03G101200	0.00	3.73	5.22	Yes	AT1G02205	Fatty acid hydroxylase superfamily	
Glyma.16G032200	0.00	3.70	5.21	Yes	AT3G61510	ACC synthase 1	Sterol desaturase
Glyma.10G083400	0.00	3.52	5.14	Yes	AT5G06470	Glutaredoxin family protein	
Glyma.19G254300	0.00	3.50	5.13	Yes	AT1G01430	Trichome birefringence-like 25	
Glyma.14G077900	0.00	3.41	5.09	Yes	AT1G47510	Inositol polyphosphate 5-phosphatase 11	
Glyma.04G238200	0.00	3.37	5.07	Yes	AT5G24120	Sigma factor E	
Glyma.12G117700	0.00	3.30	5.04	Yes	AT2G20570	GBF's pro-rich region-interacting factor 1	
Glyma.11G233400	0.00	3.29	5.04	Yes	AT2G40370	Laccase 5	
Glyma.08G106700	0.00	3.28	5.04	Yes	AT5G48570	FKBP-type peptidyl-prolyl cis-trans isomerase family protein	
Glyma.05G248100	0.00	3.11	4.96	Yes	AT1G73480	Alpha/beta-Hydrolases superfamily protein	
Glyma.07G056800	0.00	3.06	4.94	Yes	AT3G61680	Alpha/beta-Hydrolases superfamily protein	
Glyma.20G176200	0.00	2.84	4.83	Yes		Not annotated	
Glyma.08G060900	0.00	2.77	4.79	Yes	AT1G11340	S-locus lectin protein kinase family protein	
Glyma.03G230500	0.00	2.74	4.78	Yes	AT1G61040	Plus-3 domain-containing protein	
Glyma.10G246500	0.00	2.66	4.73	Yes	AT3G22640	Cupin family protein	
Glyma.17G180000	0.00	2.40	4.58	Yes	AT5G66900	Disease resistance protein (CC-NBS-LRR class) family	
Glyma.06G154200	0.00	2.35	4.56	Yes	AT2G13620	Cation/hydrogen exchanger 15	Leucine-rich repeat-containing protein
Glyma.08G204500	0.00	2.34	4.55	Yes	AT1G79820	Major facilitator superfamily protein	
Glyma.13G153200	0.00	2.29	4.52	Yes	AT2G36270	Basic-leucine zipper (bzip) transcription factor family protein	Sugar transporter
Glyma.19G245800	0.00	2.03	4.35	Yes	AT5G28680	Malectin/receptor-like protein kinase family protein	Camp-response element binding protein-related
Glyma.19G034500	0.00	1.93	4.27	Yes	AT3G02310	K-box region and MADS-box transcription factor family protein	Serine/threonine protein kinase
Glyma.12G101600	0.00	1.87	4.23	Yes	PTHR10811	Beta-1,3-n-acetylglucosaminyltransferase radical fringe	
Glyma.10G192800	0.00	1.59	3.99	Yes	AT1G04650	Not annotated	

Glyma.08G159900	1.77	0.00	-4.14	Yes	AT1G10670	ATP-citrate lyase A-1	
Glyma.06G257500	1.81	0.00	-4.18	Yes	AT4G27290	S-locus lectin protein kinase family protein	
Glyma.08G329900	2.53	0.00	-4.66	Yes	AT5G55730	FASCICLIN-like arabinogalactan 1	
Glyma.17G244200	2.90	0.00	-4.86	Yes	AT2G26560	Phospholipase A 2A	Serine-threonine protein kinase, plant-type
Glyma.06G294100	2.95	0.00	-4.88	Yes	AT1G03670	Ankyrin repeat family protein	
Glyma.13G348600	3.01	0.00	-4.91	Yes	PTHR33109	Epidermal patterning factor-like protein 3	
Glyma.02G130400	4.07	0.00	-5.35	Yes	AT5G13930	Chalcone and stilbene synthase family protein	
Glyma.19G239800	4.22	0.00	-5.40	Yes	AT3G06350	Dehydroquinate dehydratase, putative / shikimate dehydrogenase, putative	
Glyma.15G274200	4.39	0.00	-5.46	Yes	AT1G70890	MLP-like protein 43	Hydroxymethylglutaryl-coa synthase
Glyma.14G064200	4.93	0.00	-5.62	Yes	AT5G59100	Subtilisin-like serine endopeptidase family protein	
Glyma.20G142100	5.08	0.00	-5.67	Yes	AT2G33460	ROP-interactive CRIB motif-containing protein 1	
Glyma.02G064200	5.12	0.00	-5.68	Yes	AT2G02990	Ribonuclease 1	
Glyma.20G223800	5.87	0.00	-5.88	Yes	AT5G56750	N-MYC downregulated-like 1	
Glyma.13G008700	6.73	0.00	-6.07	Yes	PF13668	Ferritin-like domain (Ferritin_2)	
Glyma.01G180500	6.88	0.00	-6.10	Yes	AT2G23690	Not annotated	
Glyma.07G096700	8.03	0.00	-6.33	Yes	AT3G18280	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	Transforming growth factor
Glyma.05G188900	9.57	0.00	-6.58	Yes	AT1G49230	RING/U-box superfamily protein	
Glyma.02G112900	10.35	0.00	-6.69	Yes	AT2G03350	Not annotated	
Glyma.15G099500	10.94	0.00	-6.77	Yes	AT5G18840	Major facilitator superfamily protein	
Glyma.13G262800	20.34	0.00	-7.67	Yes	AT3G14710	RNI-like superfamily protein	
Glyma.03G132700	151.40	0.00	-10.56	Yes	AT3G57270	Beta-1,3-glucanase 1	

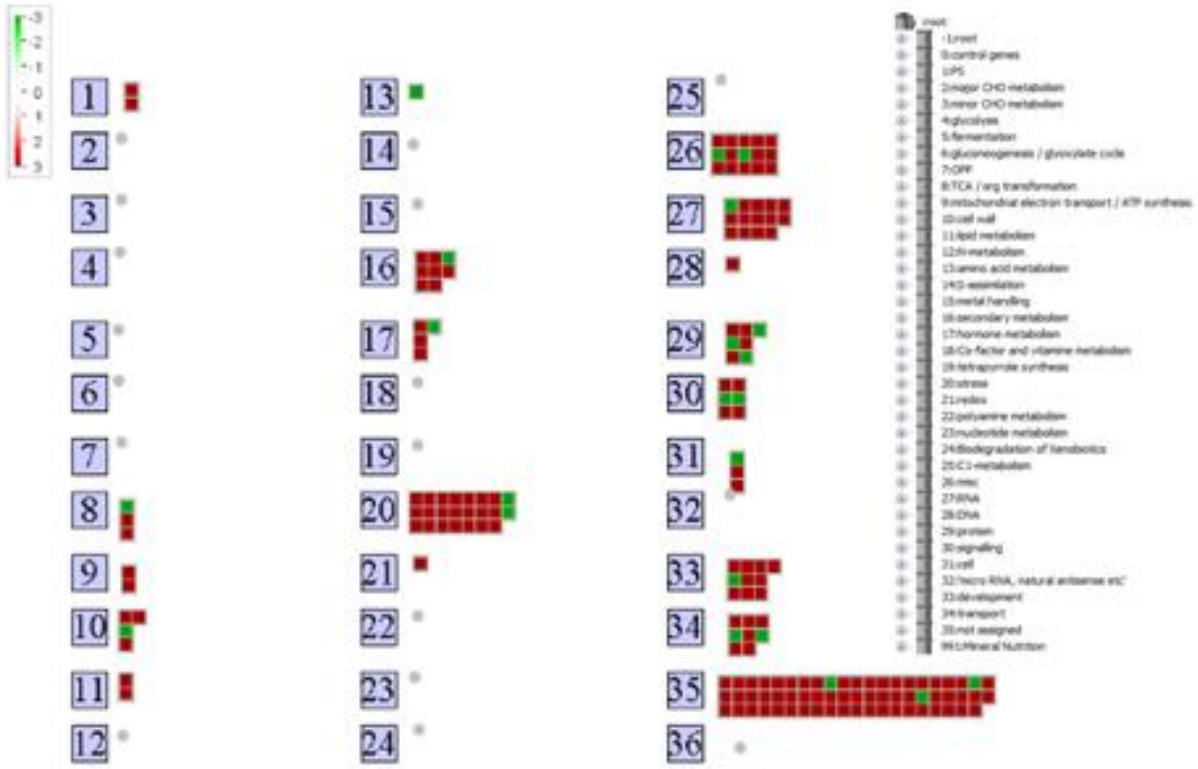


Fig. 3.21: MapMan generated Bin maps of drought stressed nodule samples DNS (30 %) VWC *vs* DS 30 % VWC of significant up- and down-regulated genes (data set 5 and 6). The degree of change is depicted based on the colour scale, where red indicates up-regulated genes and green indicated down-regulated genes.

3.3.3 Identification of drought-induced proteases and cysteine proteases

The Phytozome and Merops database were used by using their BLAST functions to identify proteases expressed in soybean nodules that had a homology of 1E < -1.0 with the aligned reads. The most dominant groups of proteases to be expressed in both DNS as well as DS nodules were the metallo proteases at 39 % and serine proteases with 27 %, of total nodule proteases expressed (Fig. 3.22 A1). Cysteine proteases, the protease family of interest, contributed only with 17 % in DNS nodules and 19 % DS nodules, respectively. Drought stress did not have a major effect on the percentage of protease groups when DNS and DS nodules were compared (Appendix B Table B.6). The representation of the different protease groups was consistent with that of the entire soybean transcriptome (14 tissue types at 20-25 days after inoculation) under non-stressed conditions (Severin *et al*., 2010).

Eighteen different groups of cysteine proteases were identified in nodules exposed to drought stress as well in DNS nodules (Fig. 3.22 B). The number of cysteine proteases which was actively expressed differs slightly between the different levels of drought stress and their DNS controls. Hundred and five cysteine proteases were active in both 40 % and 30 % VWC stressed nodules and 104 in 60 % VWC stressed nodules. In the respective DNS controls there were 103 gene transcripts at 60 % VWC, 99 gene transcripts at 40 % VWC and 101 gene transcripts for 30 % VWC. In comparison, 221 gene transcripts were previously identified in the whole soybean transcriptome under DNon-stressed conditions (Severin *et al*., 2010). Appendix B, Table B7 shows the expression of different cysteine proteases groups across the different drought treatments. The most prominent group across all treatments both drought stressed and DNon-stressed was the group of C1 cysteine proteases. This group contains papain-like cysteine proteases, homologues to the papaya (*Carica papaya)* cysteine protease, a model

representative for the C1A cysteine protease family. When an E-value cut-off of 1E-1.0 was applied to identify homologous gene sequences, 28 C1 papain-like proteases were identified in nodules under drought stress of which 24 was active in the nodules in at least one time point (Appendix B, Table B.4). Eight of these C1 cysteine proteases were induced by drought. Five of these proteases had a 2X log_2-fold expression increase after five days treatment at 30 % VWC when compared to the respective DNS controls (Table 3.9). However, none of these C1 cysteine proteases changed in their expression due to natural nodule senescence (± 8-11 weeks) under DNS conditions. *Glyma.14G085800* was the most prominent cysteine protease expressed after five days at 30 % VWC followed by *Glyma.10G207100* and *Glyma.12G039400*. Further, two C1 cysteine proteases (*Glyma.06G174800* and *Glyma.06G283100*) were expressed after five days at 40 % VWC but no expression, or only very little, was found in the respective DNS control nodules.

Eight C13 legumain-like cysteine proteases (VPEs) were also expressed in nodules during drought stress (Appendix B, Table B.5). Only three VPEs, *Glyma.17G230700*, *Glyma.14G092800* and *Glyma.05G055700*, had a 2X log_2-fold expression increase when compared to the respective DNS control nodules of the same age. However, expression of all three of these proteases increased due to natural nodule senescence (± 8-11 weeks) as well as due to drought stress (Table 3.9). The most prominent VPE expressed was *Glyma.17G230700* at 30 % VWC. Expression of this VPE was also the highest in older DNS nodules (week 11). However, there was no VPE uniquely expressed under drought.

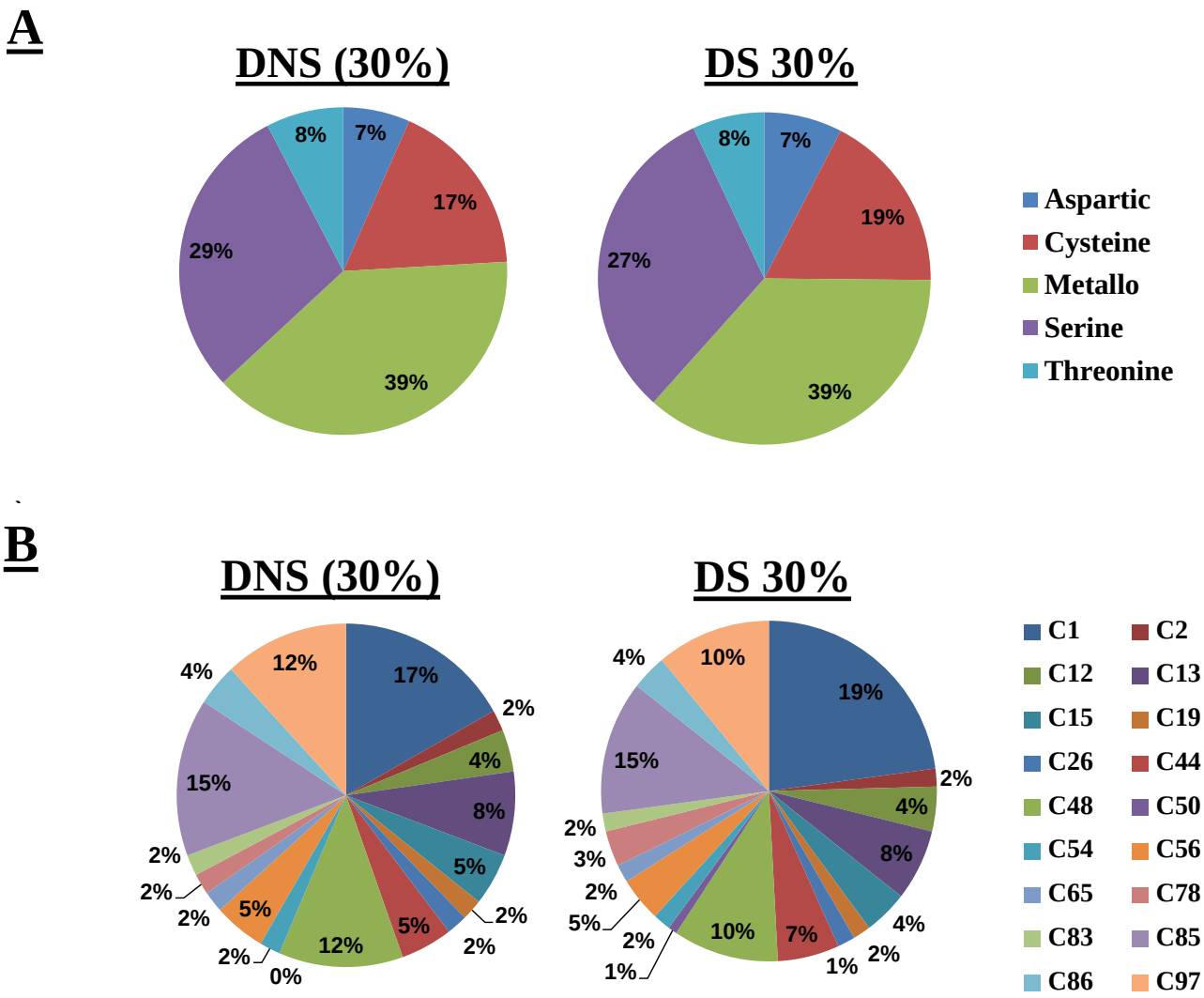


Fig. 3.22: A) Expression of different protease families as part of total proteases expressed and B) expression of different individual cysteine protease groups as part of total cysteine proteases expressed in crown nodules at 30 % VWC DS and in DNS nodules.

Table 3.9: Expression changes of Papain-like (C1) and Legumain-like (C13) cysteine proteases in drought stressed and DNon-stressed nodules.

	DNS			DS			
Gene Identifier	**(60 %)**	**(40 %)**	**(30 %)**	**60 %**	**40 %**	**30 %**	
Glyma.05G096800	**1.79**	**1.90**	**0.65**	**1.39**	**0.40**	**3.43**	**C1**
Glyma.04G014800	**1.11**	**0.16**	**1.54**	**1.17**	**2.87**	**0.82**	
Glyma.14G085800	**121.79**	**117.03**	**122.54**	**134.13**	**415.86**	**685.72**	
Glyma.12G039400	**4.21**	**4.82**	**4.15**	**3.36**	**14.01**	**23.14**	
Glyma.04G028300	**4.75**	**4.38**	**4.17**	**4.47**	**14.01**	**18.98**	
Glyma.10G207100	**2.47**	**2.28**	**1.09**	**1.50**	**7.07**	**35.83**	
Glyma.06G283100	**0.00**	**0.00**	**0.00**	**0.00**	**2.66**	**0.61**	
Glyma.06G174800	**0.00**	**0.00**	**0.17**	**0.00**	**2.81**	**0.29**	
Glyma.17G230700	**28.12**	**17.44**	**226.87**	**97.14**	**309.08**	**367.77**	**C13**
Glyma.14G092800	**8.18**	**6.05**	**39.87**	**24.86**	**45.61**	**78.01**	
Glyma.05G055700	**8.23**	**11.94**	**41.50**	**28.07**	**63.34**	**221.44**	
	56 days	**67 days**	**75 days**	**56 days**	**67 days**	**75 days**	

Expression determined as FPKM (transcript abundances in fragments per kilo base of exon per million fragments mapped).
Darker colours indicate a higher expression level.

3.3.4 Identification of drought-induced cystatins

Protease inhibitors were also identified in the nodule transcriptome (Appendix B, Table B.8). There were seven different types of protease inhibitors, with the most prominent type being the Kunitz trypsin inhibitors and cystatins which makes up 30 % and 22.2 % of inhibitors identified

in the soybean genome (Severin *et al*., 2010). Both these two types were also the most prominent groups in all the DS samples and DNS controls (Fig. 3.23). Twelve cystatins have been annotated in soybean nodules out of a possible 20 cystatins identified with the model I25B cystatin *OC-I*. The abundance of cystatins in DNS nodules were 26 % compared to DS 29 %. Appendix B, Table B.9 shows the representation of cystatins over different drought stress treatments as well as the different DNS controls.

Ten cystatins were transcriptionally active in root nodules. However only four of these were induced by drought stress (Table 3.10). Three (*Glyma.05G149800, Glyma.13G189500* and *Glyma.18G003700*) showed a 2X $\log_2$-fold change from drought stress to their DNS control at 30 % VWC. The other cystatin (*Glyma.18G103700*) showed a 2X $\log_2$-fold change at 40 % VWC.

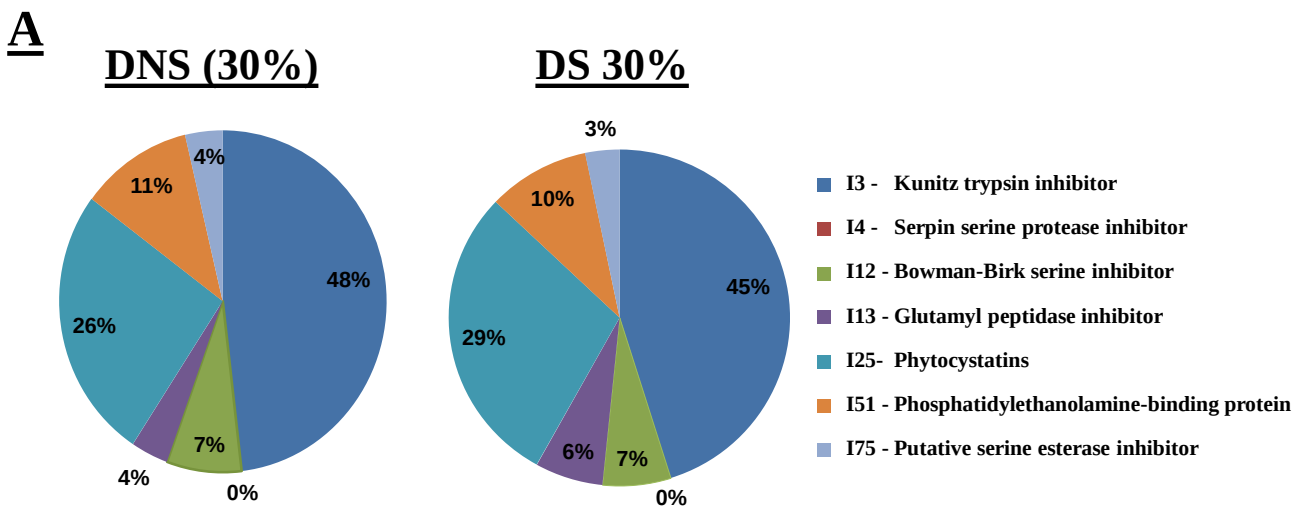


Fig. 3.23: Expression of different protease inhibitor families as part of total inhibitors expressed in crown nodules at 30 % VWC DS and DNS nodules.

Table 3.10: Expression changes of cystatins in drought stressed and DNon-stressed nodules.

	DNS			DS		
Gene Identifier	**(60 %)**	**(40 %)**	**(30 %)**	**60 %**	**40 %**	**30 %**
Glyma.05G149800	**59.85**	**55.12**	**45.54**	**53.83**	**132.73**	**357.98**
Glyma.13G189500	**50.21**	**49.94**	**42.76**	**44.17**	**115.12**	**216.87**
Glyma.18G103700	**0.00**	**0.29**	**0.45**	**0.14**	**3.78**	**0.15**
Glyma.18G003700	**0.00**	**0.00**	**0.00**	**0.00**	**10.80**	**9.14**
	56 days	**67 days**	**75 days**	**56 days**	**67 days**	**75 days**

Expression determined as FPKM (transcript abundances in fragments per kilo base of exon per million fragments mapped).
Darker colours indicate a higher expression level.

3.3.5 RNA-Seq validation

As RNA-Seq was only used as a gene discovery technique due to cost, it was necessary to validate the correctness of the expression patterns of the investigated gene transcripts with RT-qPCR (Fig. 3.24). Nodule tissue from a second independent harvest was used for validation. Three biological replicates and three technical replicates were used. The trends of C1, C13 and cystatins that were found in RNA-Seq where comparable to those found by RT-qPCR. This confirmed the fidelity of our RNA-Seq approach to identify gene transcripts that were induced by drought stress.

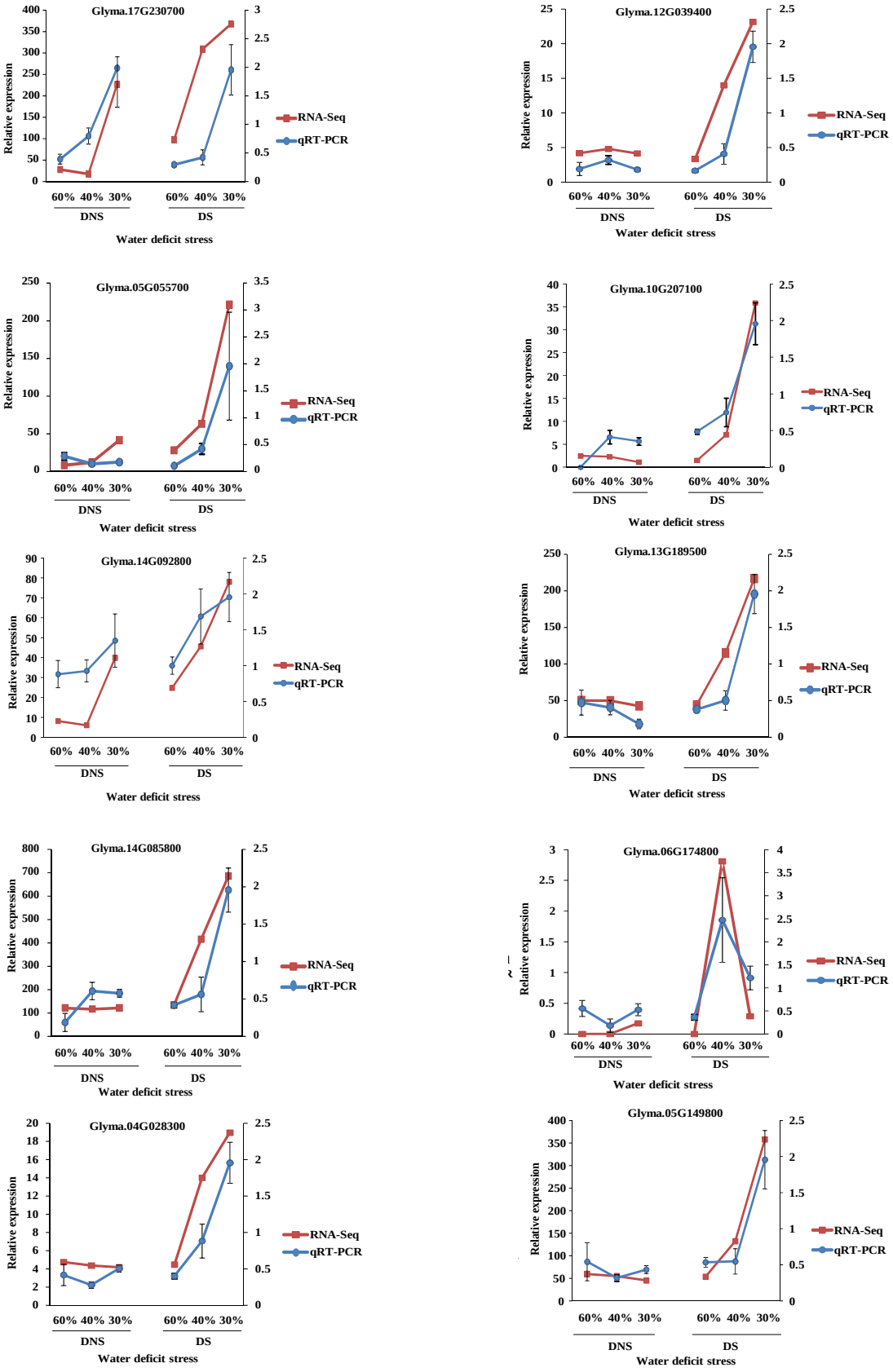


Fig. 3.24: RNA-Seq validation was done by RT-qPCR on C1 and C13 cysteine proteases and cystatins derived from drought stressed crown nodules compared to FPKM values of transcripts calculated from RNA-Seq analysis.

3.4 DISCUSSION

RNA-Seq was used as an approach to establish gene expression profiles of soybean root nodules under drought stress conditions. This study is the first analysis of gene expression profiles of crown root nodules. Previously, only the transcriptome of developing non-stressed crown nodules have been investigated (Van Wyk *et al*., 2014). The use of RNA-Seq technologies simplifies our understanding of eukaryotic transcriptomes and is a far more precise measurement to measure the level of gene transcripts than other methods (Wang *et al*., 2009). RNA-Seq was the best option to utilise, to identify possible drought markers and candidate genes for enhanced drought tolerance in the whole nodule genome and also study all expressed cysteine proteases and cystatins. Comparing legume species with non-legume species, such as *Arabidopsis*, to possibly identify genes involved in drought tolerance, are not reliable due to the complex symbiotic nitrogen fixation process, unique to legume root nodules. (Benedito *et al*., 2008).

Two comparisons were made between drought stress treatments. First comparing the three drought stress levels to each other and then comparing each drought stress level to its same age control. It was evident that a lot more genes were induced more than 2X $\log_2$-fold when drought stressed treatments were compared to their same age control. Because each treatment had its own control this comparison eliminated the age factor between the treatments. This provided evidence that drought does cause dramatic changes in gene transcript levels in root nodules. Changes in expression in a high number of genes under severe drought conditions have recently been reported for soybean roots (Song *et al.*, 2016). A first insight about any up-regulation of nodule genes following drought treatment is also presented here. These genes might be candidates for future and more detailed investigations to improve root nodule drought

tolerance. The most prominent drought-induced up-regulated gene over all treatments and in DNS (40 %) *vs* DS 40 %, was a gene encoding a Defensin-like protein (*Glyma.13G27800*), which is normally expressed in soybean flowers and roots and only seen to be very lowly expressed in nodules (Severin *et al.*, 2004). Plant genomes consist of up to 300 defensin genes, which are induced by pathogen inoculation or environmental stresses (Graham *et al.*, 2007). The second most up-regulated nodule gene over all treatments was a *LEA*-D11 protein (*Glyma.05G112000*). *LEA* (Late Embryogenesis Abundant proteins) genes are divided into six groups and *LEA*-D11 group 2 proteins, also called dehydrins, are plant-specific proteins. Various other *LEA* and dehydrin genes were highly up-regulated when comparing drought stress treatments with their same age controls. Soybean varieties differ in their *LEA*-D11 sequence. Such sequence variation was found when leaves of different drought-sensitive and drought-tolerant soybean varieties were compared (Savitri *et al*., 2013). These LEA proteins and dehydrins, produced to survive stressful environments like plant growth under low temperature or drought, possibly stabilize membranes, proteins or other cellular structures under stress (Erikson and Harryson, 2011). *LEA* and defensin-like genes will be good drought molecular markers for nodule induced senescence.

Various other genes were also highly up-regulated during drought stress such as *Glyma.12G182500* which are associated with the Cytochrome C oxidase. Preisig *et al*. (1993) showed that a mutation in this gene leads to a decrease in nitrogen fixation. Another gene, *Glyma.13G081100* which is a MtN21 Nodulin gene/EamA transporter gene, is proposed to transport O-acetylserine and cysteine proteins (Denancé *et al*., 2014). The *Glyma.13G193800* gene transcribes a VQ-motif that was up-regulated during plants experiencing nitrogen stress due to low levels of nitrogen fixation (Wanga *et al*., 2014).

Leghemoglobins, genes required for active SNF (Downie, 2005) in root nodules, were seen to be mostly down-regulated during drought stress. Leghemoglobin is responsible for the regulation of oxygen in the nodule which is required in low levels by the bacterial symbiosome (Appleby, 1984). The down-regulation of leghemoglobin genes could explain the change of colour from an active pink nodule to an inactive greenish nodule as seen in Chapter 2. Although five leghemoglobin genes of the active six in the soybean nodules (*Glyma.10G198800, Glyma.10G199000, Glyma.10G199100, Glyma.11G121800,* and *Glyma.20G191200*) were strongly down-regulated when drought was initiated, one increased in expression (*Glyma.11G121700*). The down-regulation of these five genes was also seen in developmental senescence (Van Wyk, unpublished results, 2015) which is usually associated with symbiotic leghemoglobins. However, one leghemoglobin, a non-symbiotic leghemoglobin, expression increased during drought stress. Non-symbiotic leghemoglobins possibly detoxify nitric oxide, which activates the plants' defence responses (Downie, 2005).

Drought stress also induced a higher expression of various cysteine proteases and cystatins. The majority of cysteine proteases, expressed in root nodules, belonged to the C1 cysteine protease (papain-like) family. C1 cysteine proteases represent about 50 % of all cysteine proteases in the soybean transcriptome (Severin *et al*., 2010). The expression of C1 cysteine proteases increases during drought stress conditions. Only the expression of one C1 cysteine protease, *Glyma.14G085800*, and two C13 cysteine proteases, *Glyma.17G230700* and *Glyma.05G055700,* were highly up-regulated under our most severe drought condition (30 % VWC). However, although expressed in non-stressed nodules, expression of the C1 cysteine protease, *Glyma.14G085800,* did not increase in our study when nodules senesced. The C1 cysteine protease, *Glyma.14G085800*, is also highly expressed in all other investigated soybean tissues (Severin *et al*., 2010; Van Wyk *et al*., 2014) and is further responsive to phosphorus

deficiency (Sha *et al*., 2016). *Glyma.14G085800* functions as an early indicator for drought stress and this protease specific function under drought stress should be further investigated. The C1 cysteine protease *Glyma.10G207100* showed an increase of more than 2X log_2-fold across the three drought stress treatments although the transcript level was much lower than *Glyma.14G085800* and showed a decrease in expression during natural senescence. This protease is only highly expressed in roots, pods and nodules (Severin *et al*., 2010). Although *Glyma.10G207100* is seen to be involved in nodules response to drought stress, Quian *et al*. (2015) found that this cysteine proteases were not involved in the adaptation to nitrogen limitation. The involvement of *Glyma.10G207100* in drought stress was also seen by Márquez-Garcia *et al*. (2014) after 21 days of drought stress. Both *Glyma.12G039400* and *Glyma.04G028300* were also induced slightly by drought stress compared to their constant expression during natural senescence. Both these genes have been found to be highly expressed in seeds and pods (Severin *et al*., 2010). *Glyma.04G028300* was seen to increase in expression during normal developmental senescence at 14 weeks (Van Wyk *et al*., 2014) and seems to be nitrate dependant due to its decrease in activity in soybean leaves and roots in low nitrate conditions (Quain *et al*., 2015). All three identified C13 cysteine proteases expression increased in both drought stressed and developmental senescence. This could further support the hypothesis that C13 cysteine proteases are also responsible for the activation of other enzymes such as other cysteine proteases (Roberts *et al*., 2012).

The expression of two C1 cysteine proteases, *Glyma.06G283100* and *Glyma.06G174800*, changed only slightly under drought even though the two proteases were similar to the *Arabidopsis* senescence-related *SAG12* gene with 62 % similarity by *Glyma.06G283010* and 58 % to *Glyma.06G174800*. The senescence-specific cysteine protease *SAG12* is involved in developmental senescence, specifically cell death and does not accumulate, for example, until

a leaf develops chlorosis (Weaver *et al*., 1998; Gepstein *et al*., 2003). However, except for some very low expression due to drought at 40 % VWC, both were not greatly expressed in our study under drought conditions.

Two cystatins likely located in the endoplastic reticulum (ER), involved in the secretion pathway (Van Wyk *et al*., 2014), *Glyma.05G149800* and *Glyma.13G189500*, were highly induced by drought stress (40 % and 30 % VWC). Due to the ER's high protein storage capacity genes such as cystatins might contribute to a low proteolytic activity in this organelle (Ivessa *et al*., 1999). Changes in cellular proteins are needed for plants to adjust to different abiotic stresses such as drought which can be achieved through regulation of proteolysis (Kunert *et al*., 2016). Both these cystatins also have signal peptides which are able to interact with C1 cysteine proteases (Grudkowska *et al*., 2004). The manipulation of cysteine protease activity, by altering cystatins expression, has shown enhanced tolerance to drought in other parts of soybean (Quain *et al*., 2015).

A serine proteases inhibitor (*Glyma.20G205800*) was highly up-regulated from one drought treatment to another. This protease is a Type I potato serine protease inhibitor ($OCPI_1$). The expression of an $OCPI_1$ inhibitor was also strongly induced by dehydration stress in rice (Huang *et al*., 2007). Huang *et al*. (2007) also found that rice overexpressing $OCPI_1$ had a higher grain yield, seed rate setting and a higher protein content during drought stress. Dunse *et al*. (2010) suggested that the co-expression of Type I and Type II serine potato protease inhibitors protect plants against biotic stress factors. The co-expression of this serine protease inhibitor (*Glyma.20G205800*), together with a cystatin for improved drought tolerance due to their similarity in physiological roles in the plant should be investigated.

Results from this transcriptome analysis to establish gene expression profiles of drought stressed crown nodules delivered expression profiles for cysteine protease and cystatins of root nodules. Different genes were also identified that could possibly be used to see if they will be effective candidate genes for an increase in the lifespan of crown nodules and improved SNF under drought conditions.

CHAPTER 4

The ability of soybean root nodules and the soybean cysteine protease gene expression profiles to recover from drought stress after rehydration.

ABSTRACT

Selecting possible candidate genes for breeding drought tolerant soybean cultivars with enhanced symbiotic nitrogen fixation (SNF) activity, is needed to ensure food security. The ability for drought-induced genes to recover after rehydration is important when selecting candidate genes for drought tolerance. A potted drought trial with three levels of drought stress (60 %, 40 %, and 30 % vermiculite water content (VWC)) were conducted, where after drought stressed plants were rehydrated for five days. Irreversible damage in plant growth, vegetative growth and nodule number were observed as drought intensified. Moisture content of all plant organs were also not able to return to pre-stress levels except for roots which moisture content recovered at both 40 % and 30 % VWC at time of measurement. Although nodule water potential was able to return to pre-stress levels, irreversible damage in nodule tissue and nitrogen fixation ability were observed. Seven C1 cysteine proteases, three C13 cysteine proteases and two cystatins' expression levels were able to return to pre-stress levels supporting their role in drought stress. C1 and C13 cysteine proteases initiated protein remobilization during drought stress which caused irreparable damage in nodule tissue as seen in the greenish/brown colour in nodule cross sections after rehydration. It was concluded that cysteine proteases and cystatins are good candidate genes for molecular breeding of drought tolerant soybean cultivars.

4.1 INTRODUCTION

Drought stress periods, as seen in temperate climatic regions, severely limit plant growth and productivity of important agricultural crops (Simova-Stoilova *et al*., 2010). One third of the world's current population lives in areas affected by drought (Kunert *et al*., 2015). Marker-assisted crop selection with improved drought resistance is of importance to improve food security (Francia *et al*., 2005). Plants' response and recovery to drought stress periods, are an important aspect to consider in crop selection. Therefore, a better understanding of drought tolerance and identifying molecular markers is needed for crop improvement (Kunert *et al*., 2015).

A few studies have been done on legumes and their nodules to investigate if they can recover after short drought stress periods. Nodules that do not lose more than 75 % of water have been seen to recover after rehydration (Fellows *et al*., 1987). Larrainzar *et al*. (2009) showed that leaf and nodule water potential are able to recover after six days of drought stress. They also showed that proteins that accumulated due to drought stress were able to be reversed. Even though plants and nodule water status recover to some extent, the recovery of SNF activity is very limited especially following severe drought stressed periods (Larrainzar *et al*., 2009, Fellows *et al*., 1987). If SNF can be maintained during drought stress, it is likely that plants could have a higher yield (Pimratch *et al*., 2008).

The activation of proteases to facilitate remobilisation of nitrogen reserves via proteolysis is an adaptive mechanism of plants during stress conditions (Simova-Stoilova *et al.*, 2010; Kidrič *et al*., 2014). Nodule cysteine proteases, as mentioned, are found in senescent nodules (Kardailsky

and Brewin, 1996; Vorster *et al*., 2013; Van Wyk *et al*., 2014) and their expression affect SNF activity (Van de Velde *et al*., 2006) due to the degradation of leghemoglobin and other proteins.

Cystatins, an important role player in stress-induced senescence, and programmed cell death (PCD) are involved in protein regulation (Benchabane *et al*., 2010) by providing a balanced interplay between the proteoltic activity, cysteine proteases and cystatins (Grudkowska and Zagdanska 2004). Cystatin gene transcripts also accumulated during drought stress conditions (Chapter 3, Diop *et al*., 2004). It has been suggested that cystatins with an increased expression during rehydration contribute to the recovery of plants and nodules after drought stress (Pinheiro *et al*., 2005). This is because they limit cell senescence. The ability of drought induced C1 and C13 cysteine proteases and cystatins, to return to pre-stress levels needs to be investigated. This will indicate whether these drought-induced genes are only involved in the induced senescence process or do they help the plant to recover during rehydration. Thereafter these genes can be considered for crop improvement.

The objective of this part of the study was to investigate whether soybean plants and nodule development are able to recover after different levels of drought stress, upon rehydration. C1 and C13 cysteine proteases and cystatins that were induced by drought, as seen with RNA-Seq, were further analysed with RT-qPCR and ddPCR to compare gene expression of drought stressed and rehydrated nodules.

4.2 MATERIALS AND METHODS

4.2.1 Plant material and vegetative development

In an independent experiment, soybean (*Glycine max* L. Merr) seeds of the Prima 2000 cultivar (Pannar Seed, South Africa) were grown, as mentioned previously (Chapter 2), until all plants reached a plastochron index of 3.6. Drought stress was initiated and vermiculite water content was lowered to 60 %, 40 % and 30 % VWC and kept at the respective VWC percentages for five days. Plants were then rehydrated by increasing the VWC until 100 % after the fifth day of drought stress. Plants were then kept at 100 % VWC for another five days and were harvested at the end of the fifth day. The rehydration treatment (R) also has a non-stressed same age control that was kept at 100 % VWC throughout the trial, which will be called RNon-stressed (RNS) hereafter. The ages of the plants after rehydration at the time of harvest were 61 days (60 % VWC), 72 days (40 % VWC) and 80 days (30 % VWC) after sowing.

Five plant organs (young leaves, old leaves, root tips, shoots crown and lateral nodules) from seven different plants were harvested. Four samples from all five organs were used in subsequent growth analysis and three samples of nodules were flash frozen with liquid nitrogen.

4.2.1.1 Moisture content

Fresh mass (FM) was determined using a AB104-5 Mettler-Tolendo balance for the two youngest trifoliate leaves (not fully expanded), remaining trifoliate leaves, roots, shoots, crown and lateral nodules. Samples were then dried in an oven at 60 ºC for approximately 48 h until a constant mass were obtained to determine the dry mass (DM). The moisture content was then

determined on a wet basis as follows: Moisture content = (Fresh mass-Dry mass)/Fresh mass X 100.

4.2.1.2 Nodule number, leaf and nodule water potential

Leaf water potential (Ψ_{Leaf}) was determined predawn on the day of harvest of DS, R and RNS samples using a pressure bomb (Model 3005 ITC International Australia) as explained in Chapter 2.

Crown nodules were collected by hand and counted. Nodule water potential (Ψ_{Nod}) was determined immediately after the harvest commenced at 09:00, using 0.1 g of crown nodules together with a WP4 Dew Point Potential meter (Decagon, USA) as described by Guerin *et al.* (1990).

4.2.2 Nitrogenase activity

For determination of symbiotic N_2 fixation activity, plants were removed from pots and a 4 cm root sample where crown nodules were attached, was placed in an air-tight Erlenmeyer flasks with a 52 ml capacity. Five ml of air was extracted from the flask with an air tight syringe and replaced with 5 ml acetylene. After a 15 min incubation period, at room temperature, ethylene production was measured by extracting 1 ml of gas from the headspace of each flask and injecting it into the gas chromatograph (GC 2025; Shimadzu., Japan) according to the method described by Turner and Gibson (1980). A flame ionization detector was used with an oven temperature of 200 °C, column temperature of 130 °C a gas flow of: air (400 kPa), H_2 (40 kPa), He carrier gas (40 kPa) and running time was three min. Acetylene reduced to ethylene was

calculated using a standard curve that was made by injecting pure ethylene in different volumes ranging from 1 µl to 9 µl into the gas chromatography. Acetylene reduced was calculated by nmol per 100 µl/drymass (g)/incubation time of 15 minutes.

4.2.3 Gene expression studies

4.2.3.1 RNA Extraction and cDNA synthesis

Crown nodules were harvested from DS, R as well as RNS plants from each of the 60 %, 40 % and 30 % VWC treatments and were immediately flash frozen in liquid nitrogen. RNA were extracted and cDNA synthesised as previously mentioned (Chapter 3). For gene expression, three biological plant replicates were used for each treatment.

4.2.3.2 Gene expression of rehydrated nodule samples

Gene expression validation on rehydrated nodules was done using RT-qPCR. The same primers designed for RNA-Seq confirmation were used to investigate whether cysteine proteases and inhibitors induced by drought stress expression levels could be down-regulated to pre-stress levels. Primer sequences as well as amplicon product sizes can be found in Appendix B, Tables B.10. Thermo cycling was carried out with a Bio-Rad CX5 Thermo cycler (Bio-Rad, USA) and a KAPA SYBR® FAST qPCR Kit (Kapa Biosystems, USA). Normalization of expression values was carried out with housekeeping genes ribosomal 40S protein subunit S8 (NCBI - XM_003532110) (40S) and Elongation factor 1-beta (ELF1) (NCBI - XM_003545405). All standard curves of genes of interest had a PCR efficiency of between 90 %-110 % with a R^2 value higher than 0.9.

4.2.3.3 Gene expression studies of expressed genes with a low copy number

Gene expression validation for genes with low expression levels was done using a Digital Droplet PCR (Bio-Rad, USA). Primer sequences as well as amplicon product sizes can be found in Appendix B, Tables B.10. Manufacturer instructions were followed by applying 25-100 ng/ul of cDNA together with QX200™ ddPCR™ EvaGreen Supermix. Three biological replications were used for each treatment and as well as three technical replications. The reaction mixture and QX200™ Droplet Generation Oil were loaded into the DG8™ cartridges. Droplets were generated using a QX200™ droplet generator (Bio-Rad, Germany). Thereafter PCR amplifications were carried out with a T100™ Bio-Rad Thermal Cycler. The reactions were set up at 95 °C for 5 min followed by cycling at 95 ºC for 30 sec and 60 °C for 1 min over 40 cycles. The reaction was then cooled down to 4 ºC for 5 min and ramped up to 90 °C for 5 min to stabilize the signal. A ramp rate of 2 °C/second was used for all cycling steps. Droplets formed were then read with a QX200™ Droplet Reader. The QuantaLife software package was applied for data analysis. Normalization of expression values was carried out with housekeeping genes ribosomal 40S protein subunit S8.

4.2.4 Statistical analysis

Statistical analysis of expression changes were determined after rehydration of DS, R and RNS plants and nodules. A one-way ANOVA was conducted across treatments (RNS (60 %) *vs* DS 60 % *vs* R 60 %) as well as within a treatment (DS 60 % *vs* DS 40 %, *vs* DS 30 %), if results showed a normal distribution over residual, a Duncan post-test was used in a next step. A two way unpaired t-test was performed to determine significance in gene expression studies.

SPSS© Version 23 and IBM © Software was applied. A P-value of $P \leq 0.05$ were seen as significantly different.

4.3 RESULTS

4.3.1 Recovery in plant growth

The ability of soybean plants to resume growth and SNF after exposure to drought stress were investigated as a next step. The vegetative growth of plants were irreversibly affected by drought stress as there was no difference in plant growth between DS and R plants (Fig. 4.1, Fig. 4.2 A and B). This was confirmed by measuring the vegetative growth (PI index) where plant growth could not resume growth to result in a significant difference in PI compared to DS samples. (Fig. 4.2 A). DS samples did significantly differ ($P \leq 0.05$) from RNS controls in both 40 % and 30 % VWC, although drought stress treatments do not differ significantly over the different levels of stress (7.35 ± 0.1, 6.61 ± 0.2 and 7.26 ± 0.35 at 60 %, 40 % and 30 % VWC respectively). A significant difference ($P \leq 0.05$) can be seen at 30 % VWC (12.13 ± 0.5 30 % VWC) in RNS controls compared to 60 % and 40 % VWC (7.94 ± 0.4 and 8.46 ± 0.6 at 60 % and 40 % respectively).

Crown nodules were also dissected (Fig. 4.3) and it was clear that, as the levels of drought stress got more severe, nodule tissue was not able to recover to its active red/pink nitrogen fixing state but became more inactive with a brown/greenish interior indicating continuing degradation of leghemoglobin.

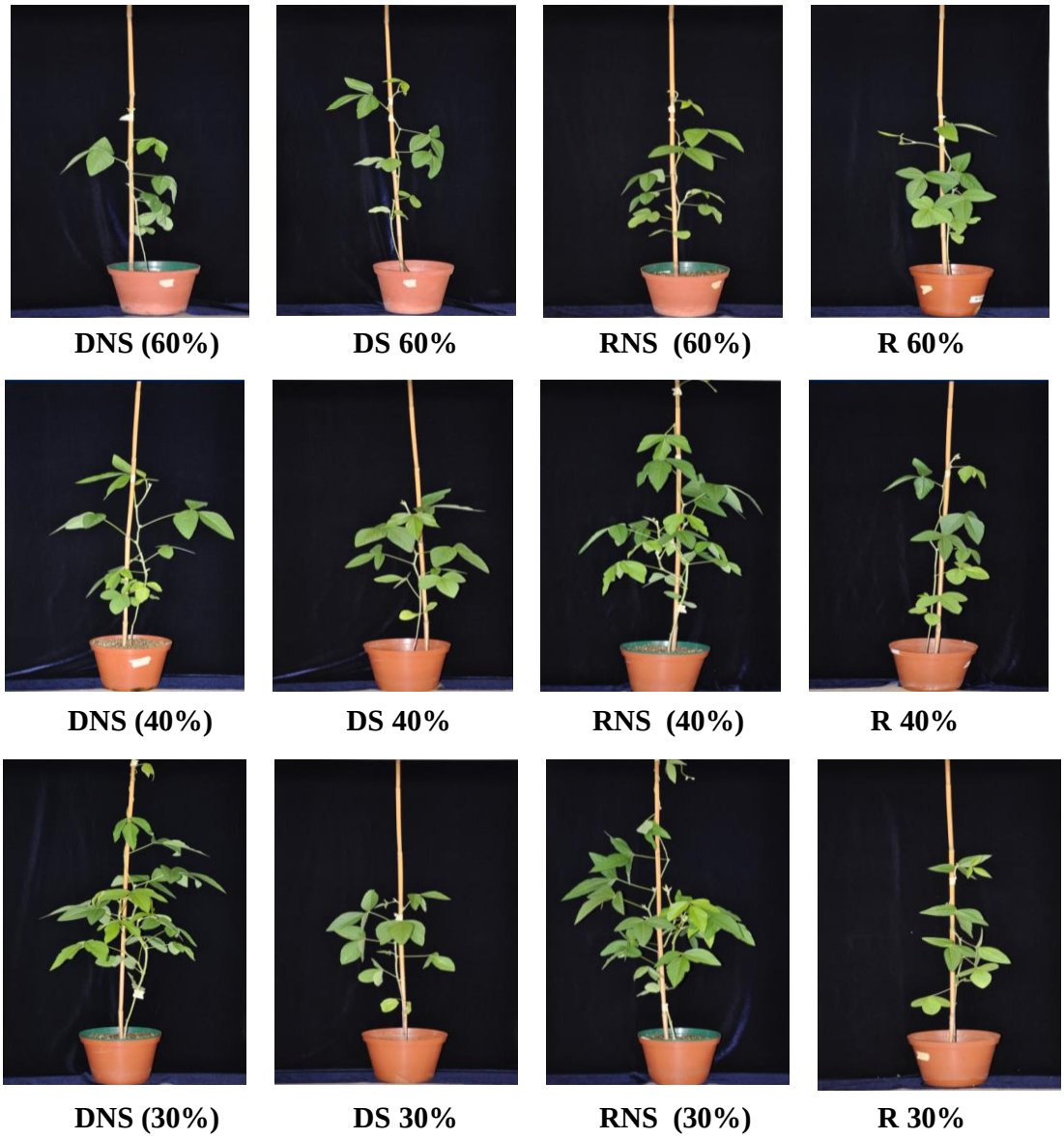
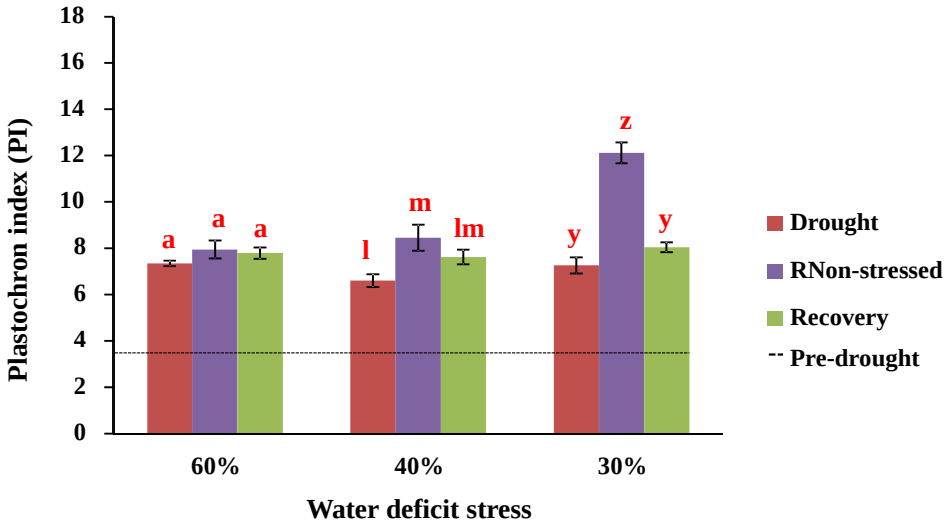


Fig. 4.1: Soybean growth under non-stress (DNon-stressed and RNon-stressed), drought stressed and recovery conditions applied as different percentage of VWC representing different levels of drought (60 %, 40 % and 30 % VWC) and the recovery of the drought stressed plants to 100 % VWC. The ages of the plants after rehydration at harvest were 61 days (60 % VWC), 72 days (40 % VWC) and 80 days (30 % VWC) after sowing (five days older than drought stressed plants).

A



B

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	A
RNon-stressed	L	L	M
Recovery	Z	Z	Z

Fig. 4.2: A) Plastochron index (PI) of plants under drought stress (DS) conditions and after recovery of each treatment including RNon-stressed. B) Comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC) recovery as well as across RNon-stressed controls. Pre-drought plants had a Plastochron index of 3.6. Data represents the mean ± SE of the PI of five individual plants.

[a-c] Significant differences of treatments at 60 % VWC

[l-n] Significant differences of treatments at 40 % VWC

[y-z] Significant differences of treatments at 30 % VWC

[A-C] Significant differences of the three DS treatments were compared to each other.

[L-N] Significant differences of the three RNon-stressed treatments were compared to each other.

[Y-Z] Significant differences of the three R treatments were compared to each other.

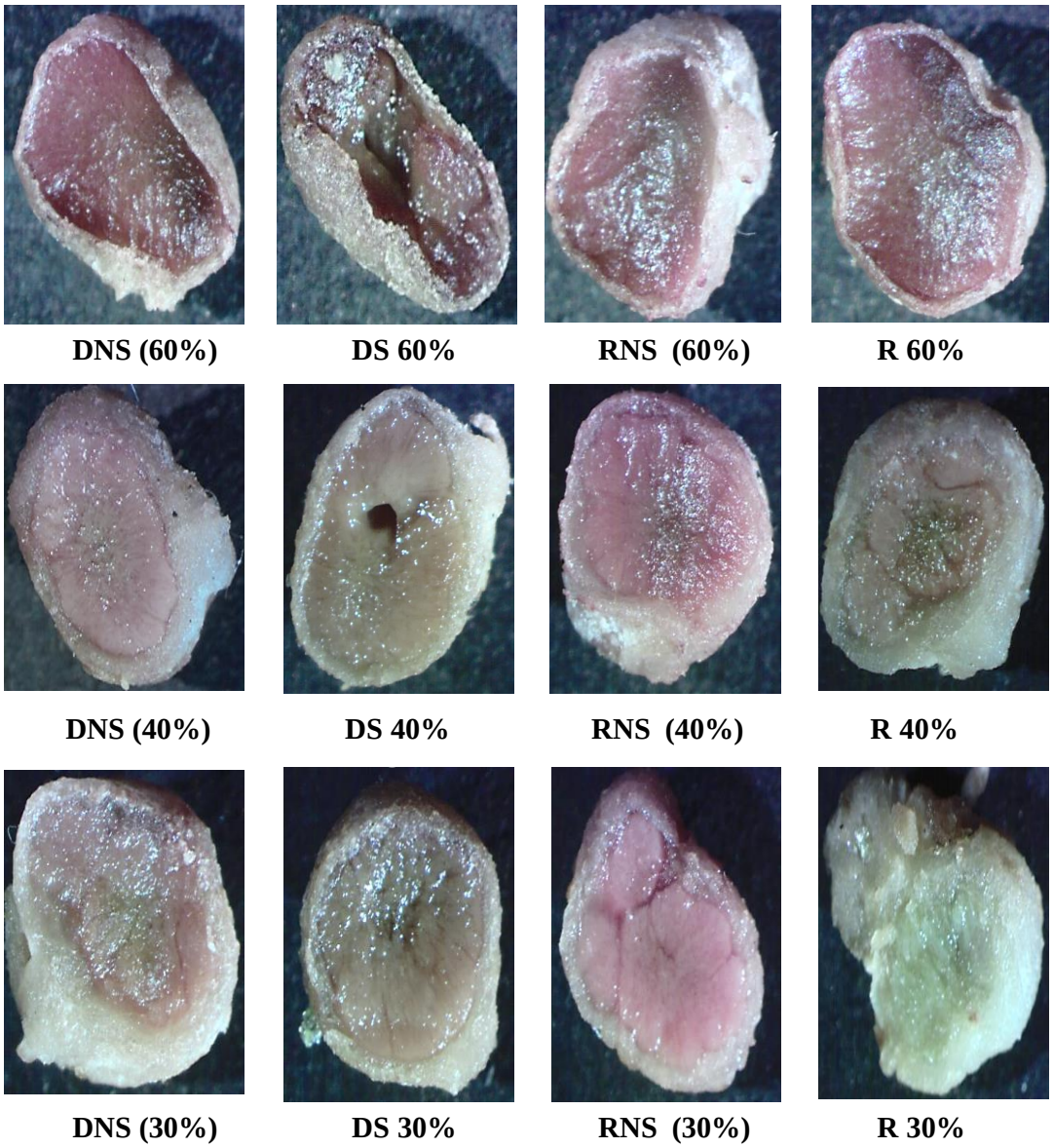
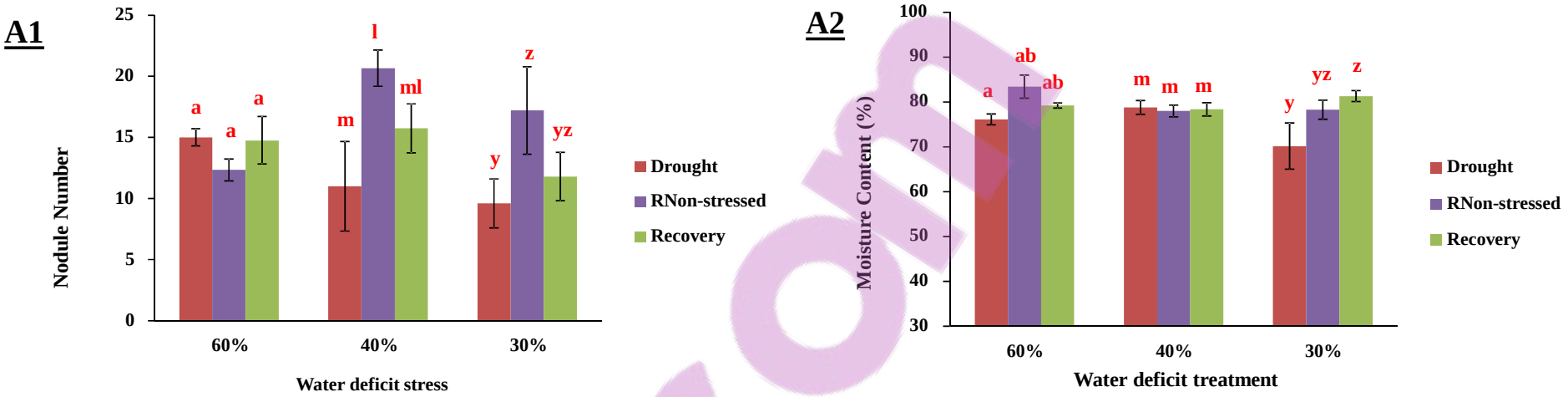


Fig. 4.3: Crown nodule cross sections under non-stress (DNon-stressed and RNon-stressed), drought stressed and recovery conditions applied as different percentage of VWC representing different levels of drought (60 %, 40 % and 30 % VWC) and the recovery of the drought stressed plants to 100 % VWC. The ages of the plants after rehydration at harvest were 61 days (60 % VWC), 72 days (40 % VWC) and 80 days (30 % VWC) after sowing (five days older than drought stressed plants).

4.3.2 Moisture content of soybean plant organs and nodule number after rehydration

The interest in whether crown nodules can recover after drought stress was further investigated by counting the nodules by hand as previously seen and confirmed yet again that drought stress prevents further nodule formation. (Fig. 2.7 and Fig. 4.4 A and B). No significant difference ($P \leq 0.05$) was seen in nodule numbers between DS (Fig. 4.4 A2) nodules (15 ± 0.9, 11 ± 3.67 and 9.60 ± 2.01 at 60 %, 40 % and 30 % VWC respectively) and nodules that were rehydrated (14.75 ± 1.9, 15.75 ± 2.0 and 11.8 ± 2.0 at 60 %, 40 % and 30 % VWC respectively). A significant difference was however seen between DS and the RNS control at 40 % and 30 % VWC with the nodule number of 20.66 ± 0.9 at 40 % VWC and 17.77 at 30 % VWC.

Crown nodule moisture content was able to recover to some extent as a significant difference from DS nodules (70.15 % ± 5.14) to rehydrated nodules (81.25 % ± 1.2) were seen at 30 % VWC (Fig 4.4 A2). There was also no significant difference between nodules at 30 % VWC that have been rehydrated and in RNS (78.25 % ± 2.12). Young leaves (Fig. 4.5 A1 and B1), old leaves (Fig. 4.5 A2 and B2), and shoots (Fig. 4.6 A2 and B2), were not able to recover from severe stress (30 % VWC) after rehydration as there were no significant difference between the DS (30 % VWC) treatments and R was seen at the time of the measurement. However, root moisture content was severely affected by drought stress (Fig 4.6 A1 and B1) in both the 40 % and 30 % VWC (50.43 % ± 3.8 at 40 % VWC and 41.96 % ± 34.3 at 30 % VWC) but recovered completely after rehydration (80.78 % ± 1.59 at 40 % VWC and 80.84 % ± 3.37 at 30 % VWC.



B1

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	A
RNon-stressed	M	M	M
Recovery	Z	Z	Z

B2

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	A
RNon-stressed	M	M	M
Recovery	Z	Z	Z

Fig. 4.4: A1) Crown nodule number of nodules and A2) nodule moisture under drought stressed conditions and after recovery of each treatment including its RNon-stressed control. B1 and B2) comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC), recovery as well as RNon-stressed controls.

Data represent the mean ± SE of nodules from four individual plants.

[a-c] Significant differences of treatments at 60 % VWC

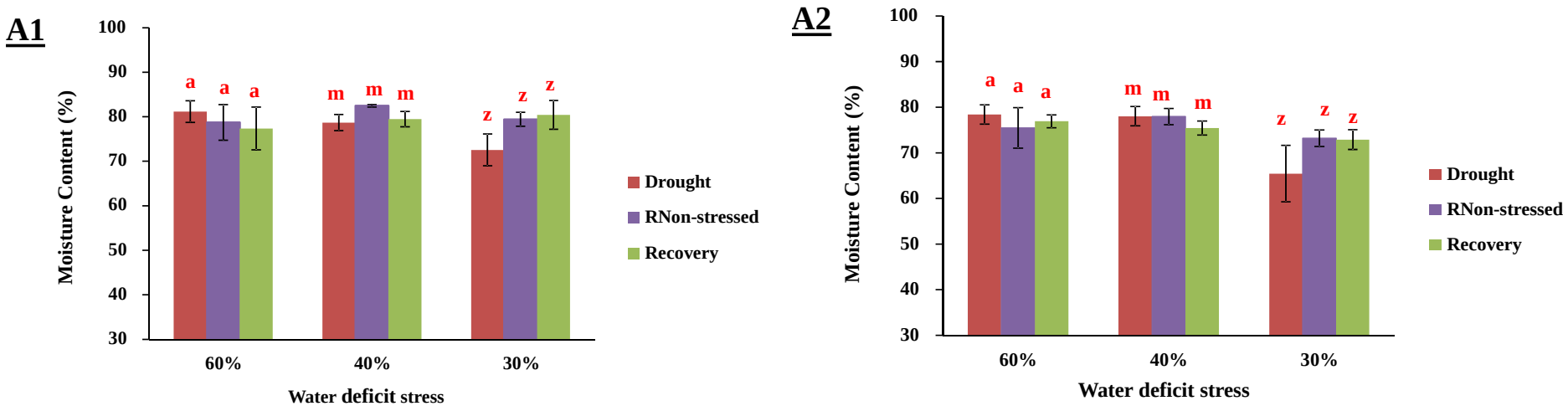
[l-n] Significant differences of treatments at 40 % VWC

[y-z] Significant differences of treatments at 30 % VWC

[A-C] Significant differences of the three DS treatments were compared to each other.

[L-N] Significant differences of the three RNS treatments were compared to each other.

[Y-Z] Significant differences of the three R treatments were compared to each other.



B1

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	AB	AB
RNon-stressed	M	M	M
Recovery	Z	Z	Z

B2

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	B
RNon-stressed	M	M	M
Recovery	Y	Y	Z

Fig. 4.5: A1) Moisture content of young leaves and A2) old leaves under drought stressed conditions and after recovery of each treatment including its RNon-stressed treatment. B1 and B2) comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC), recovery as well as across RNon-stressed controls.

Data represent the mean ± SE of leaves from four individual plants.

[a-c] Significant differences of treatments at 60 % VWC

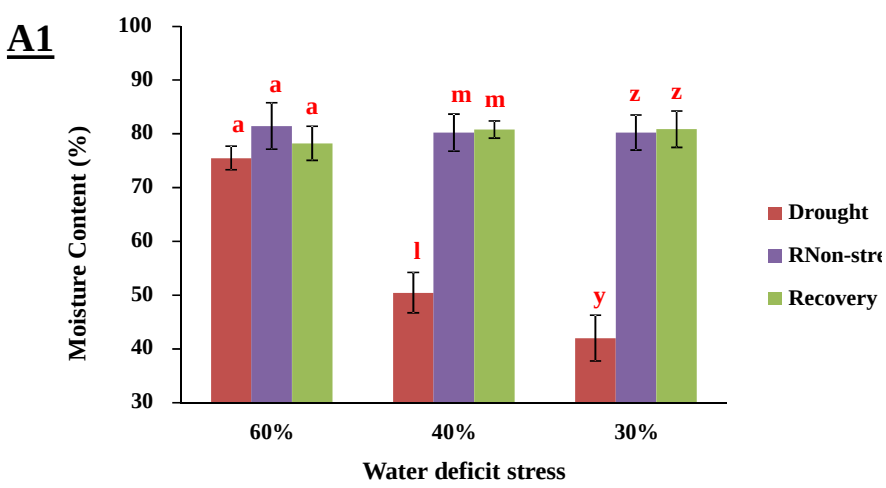
[l-n] Significant differences of treatments at 40 % VWC

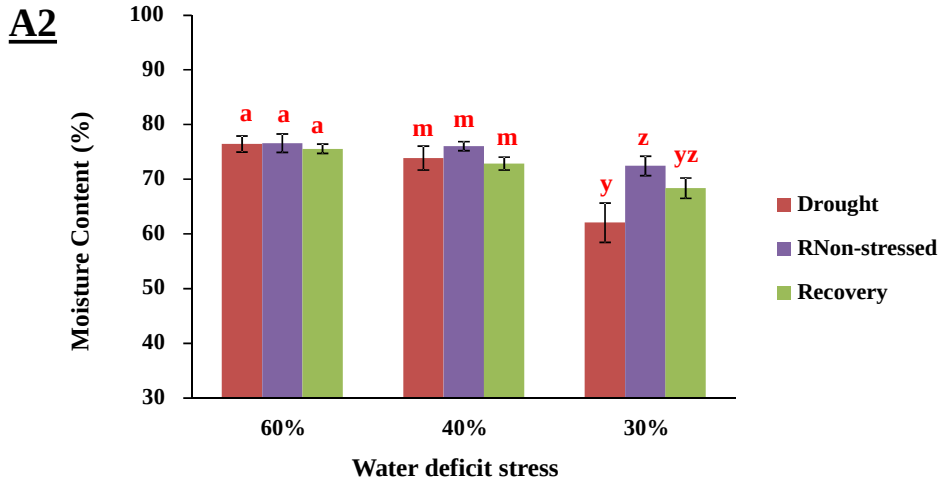
[y-z] Significant differences of treatments at 30 % VWC

[A-C] Significant differences of the three DS treatments were compared to each other.

[L-N] Significant differences of the three RNS treatments were compared to each other.

[Y-Z] Significant differences of the three R treatments were compared to each other.





B1

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	B	B
RNon-stressed	M	M	M
Recovery	Z	Z	Z

B2

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	B
RNon-stressed	M	M	M
Recovery	Z	Z	Z

Fig. 4.6: A1) Moisture content of roots and A2) shoots under drought stressed conditions after recovery of each treatment including RNon-stressed controls. B1 and B2) comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC), recovery as well as across RNon-stressed controls.

Data represent the mean ± SE of roots and shoots from four individual plants.

[a-c] Significant differences of treatments at 60 % VWC

[l-n] Significant differences of treatments at 40% VWC

[y-z] Significant differences of treatments at 30 % VWC

[A-C] Significant differences of the three DS treatments were compared to each other.

[L-N] Significant differences of the three RNS treatments were compared to each other.

[Y-Z] Significant differences of the three R treatments were compared to each other.

4.3.3 Leaf and nodule water potential

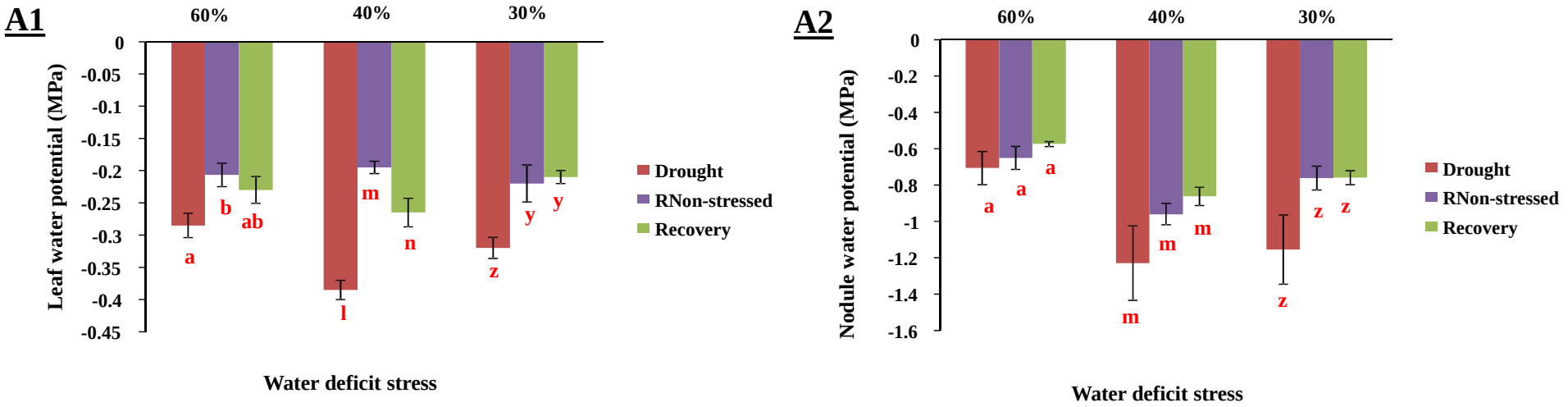
Leaf and nodule water potential were measured to explore to what extent these organs were able to recover after rehydration. Leaf water potential were able to recover slightly at 40% and completely 30 % VWC (Fig. 4.7 A1). A significant difference ($P \leq 0.05$) between the respective drought stress samples and their rehydrated counterparts with a Ψ_{Leaf} of -0.36 MPa ± 0.015 MPa at 40 % VWC that were able to recover to -0.27 MPa ± 0.02 MPa. A significant difference was seen between R and RNS (-0.19 MPa ± 0.01 MPa). A complete recovery was seen at 30 % VWC with a Ψ_{Leaf} of -0.32 MPa ± 0.02 MPa that recovered to -0.21 MPa ± 0.01 MPa.

Nodule water potential which was significantly ($P \leq 0.05$) affected by drought stress in both the 40 % and 30 % VWC treatments were both able to recover completely (Fig. 4.7 A2 and B2). Nodule water potential increased from -1.40 MPa ± 0.16 MPa to -0.86 MPa ± 0.05 MPa at 40 % VWC and in 30 % VWC from -1.34 MPa ± 0.03 MPa to -0.76 MPa ± 0.4 MPa.

4.3.4. Nitrogenase activity

Nitrogenase activity was measured to examine whether SNF were able to recover after the termination of drought stress. The nitrogenase activity was significantly lower ($P \leq 0.05$) due to drought stress as seen in DS 40 % VWC and DS 30 % VWC nodules compared to DS 60 % (Fig. 4.8 A and B). No recovery was seen in nitrogenase activity as there was a significant lower ($P \leq 0.05$) amount of acetylene reduced between drought stress treatments and their recovery treatment in both 60 % and 30 % VWC. Age is a major contributing factor in acetylene reduction as can be seen by the significant decrease in ($P \leq 0.05$) acetylene reduced RNS

control samples which showed a decline ($P \leq 0.05$) in nitrogenase activity at 40 % VWC and 30 % VWC compared to 60 % VWC.



B1

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	B	A
RNon-stressed	M	M	M
Recovery	YZ	Y	Z

B2

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	A	A
RNon-stressed	L	L	LM
Recovery	y	Z	Z

Fig. 4.7: A1) Leaf water potential and A2) nodule water potential under drought stress conditions and after recovery of each treatment including RNon-stressed controls. B1 and B2) comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC) recovery as well as across RNon-stressed controls.

Data represent the mean ± SE of leaves and nodules from four individual plants.

[a-c] Significant differences of treatments at 60 % VWC

[l-n] Significant differences of treatments at 40 % VWC

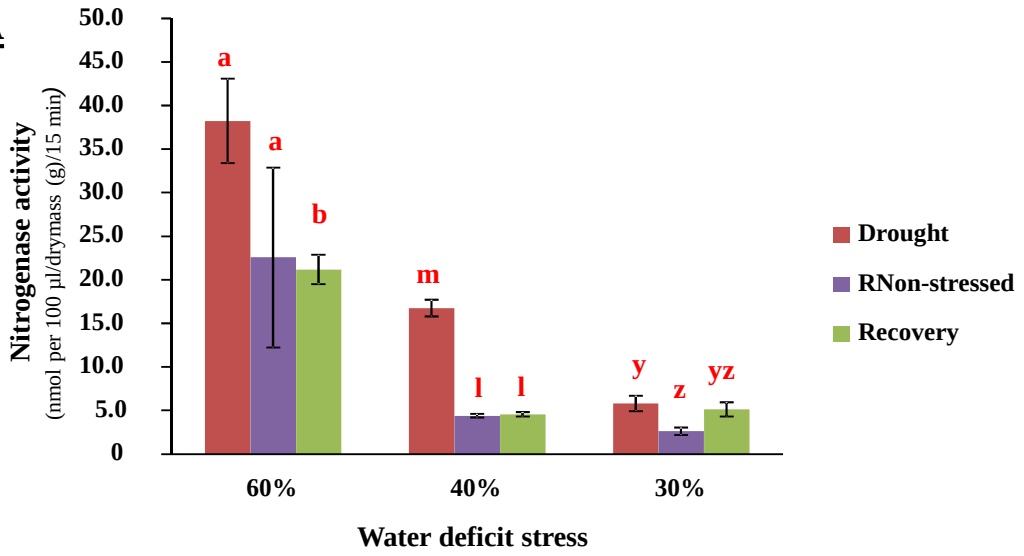
[y-z] Significant differences of treatments at 30 % VWC

[A-C] Significant differences of the three DS treatments were compared to each other.

[L-N] Significant differences of the three RNS treatments were compared to each other.

[Y-Z] Significant differences of the three R treatments were compared to each other.

A



B

Treatment	Vermiculite water content		
	60%	40%	30%
Drought stress	A	B	C
RNon-stressed	M	L	L
Recovery	y	Z	Z

Fig. 4.8: A) Nitrogenase assay (acetylene reduction nmol per 100 µl/drymass (g)/incubation time (0.25) of nodules under drought stressed conditions and after recovery of each treatment including RNon-stressed controls. B1 and B) comparison of data across different levels of drought stress treatments (60 %, 40 % and 30 % VWC) recovery as well as across RNon-stressed controls.

Data represent the mean ± SE of leaves and nodules from three individual plants.

$^{a\text{-}c}$ Significant differences of treatments at 60 % VWC

$^{l\text{-}n}$ Significant differences of treatments at 40 % VWC

$^{y\text{-}z}$ Significant differences of treatments at 30 % VWC

$^{A\text{-}C}$ Significant differences of the three DS treatments were compared to each other.

$^{L\text{-}N}$ Significant differences of the three RNS treatments were compared to each other.

$^{Y\text{-}Z}$ Significant differences of the three R treatments were compared to each other.

4.3.5 Gene expression analysis

Genes induced by drought stress, as identified after comparative transcriptomics that had a 2X log_2-fold change, were further analysed with either RT-qPCR or in the case of genes with a low transcript abundance with ddPCR. It was evaluated whether their expression levels normalised after plants were treated with the different levels of DS (60 %, 40 % and 30 % VWC) were rehydrated to 100 % VWC for five days. Eight C1 cysteine protease genes, three C13 cysteine proteases and four cystatins were induced by drought stress (Chapter 3). The effect rehydration had on five of these C1 cysteine proteases genes (Fig. 4.9) (*Glyma.14G085800, Glyma.04G028300, Glyma.12G039400, Glyma.10G207100 and Glyma.06G174800*), were investigated with RT-qPCR. *Glyma.04G028300, Glyma.12G039400* and *Glyma.10G207100* were significantly ($P \leq 0.01$) induced more than 2X log_2-fold in DS 30 % VWC (Fig. 4.9 B) in nodules compared to RNS and also showed a return to pre-stress expression levels. Genes returning to pre-stressed levels will be referred to as genes that recovered. Although *Glyma.14G085800* did normalised slightly after drought stress ($P \leq 0.05$) it did not show a 2X log_2-fold induction or recovery to pre-stress expression levels (Fig. 4.9 B1). A peak in gene expression of *Glyma.06G174800* was seen earlier in DS 40 % and not DS 30 % compared to proteases mentioned above. When taking this into consideration, it seems that the proteases did recover although not significantly ($P \geq 0.05$). *Glyma.06G174800* were the only gene that showed a higher expression at 30 % VWC after rehydration in R samples, compared to DS samples but the changes in gene expression were also not significant ($P \geq 0.05$). Changes in two C1 cysteine proteases, gene expression levels after rehydration were confirmed with ddPCR rather than RT-qPCR due to their low copy number (Fig. 4.11 A1-2). In this case a low copy number refers to genes that have expression levels below the detection levels that can be quantified with RT-qPCR or had a very high CT value. *Glyma.05G096800* recovered after 30 % VWC drought although this was not significant

($P \geq 0.05$). *Glyma.04G014800* were induced more than 2X $\log_2$-fold at 40 % VWC in DS nodules and also showed no significant difference ($P \geq 0.05$) after rehydration. One C1 cysteine protease (*Glyma.06G283100*) could not be quantified with RT-qPCR or ddPCR due to too low copy number. Its ability to recover after rehydration should be investigated in the future.

For all three C13 cysteine proteases (*Glyma.17G230700, Glyma.14G092800* and *Glyma.05G055700*) that were seen to be induced by drought stress, transcript abundance also normalised after rehydration (Fig. 4.10 A1-A3). *Glyma.14G092800* and *Glyma.05G055700* expression was down-regulated more than 2X $\log_2$-fold in 30 % VWC DS samples compared to rehydrated samples 30 % VWC and the down-regulation was significant ($P \leq 0.05$) in *Glyma.17G230700* (Fig. 4.10 B1-B3). The same was also true for the measurement of two cystatins quantified by RT-qPCR (*Glyma.05G149800* and *Glyma.13G189500*), which transcript abundances normalised after rehydration (Fig. 4.10 A4-A5). Both these cystatins showed a higher than 2X $\log_2$-fold increase ($P \leq 0.05$) in expression in DS 30 % VWC nodule samples compared to their respective same age controls as well as a lower than 2X $\log_2$-fold down-regulation ($P \leq 0.05$) in expression from DS nodules to R nodules (Fig. 4.10 B4-B5). Two cystatins (*Glyma.18G103700* and *Glyma.18G003700*) expression levels after rehydration were also investigated with ddPCR. Both these cystatins were induced by drought at 40 % VWC and not 30 % VWC. *Glyma.18G103700* showed a full significant recovery after rehydration ($P \leq 0.01$) with a down-regulated expression of more than 2X $\log_2$-fold at 40 % VWC (Fig. 4.11 A3 and B3). *Glyma.18G003700* did not show any significant changes in expression levels (Fig. 4.11 A4 and B4).

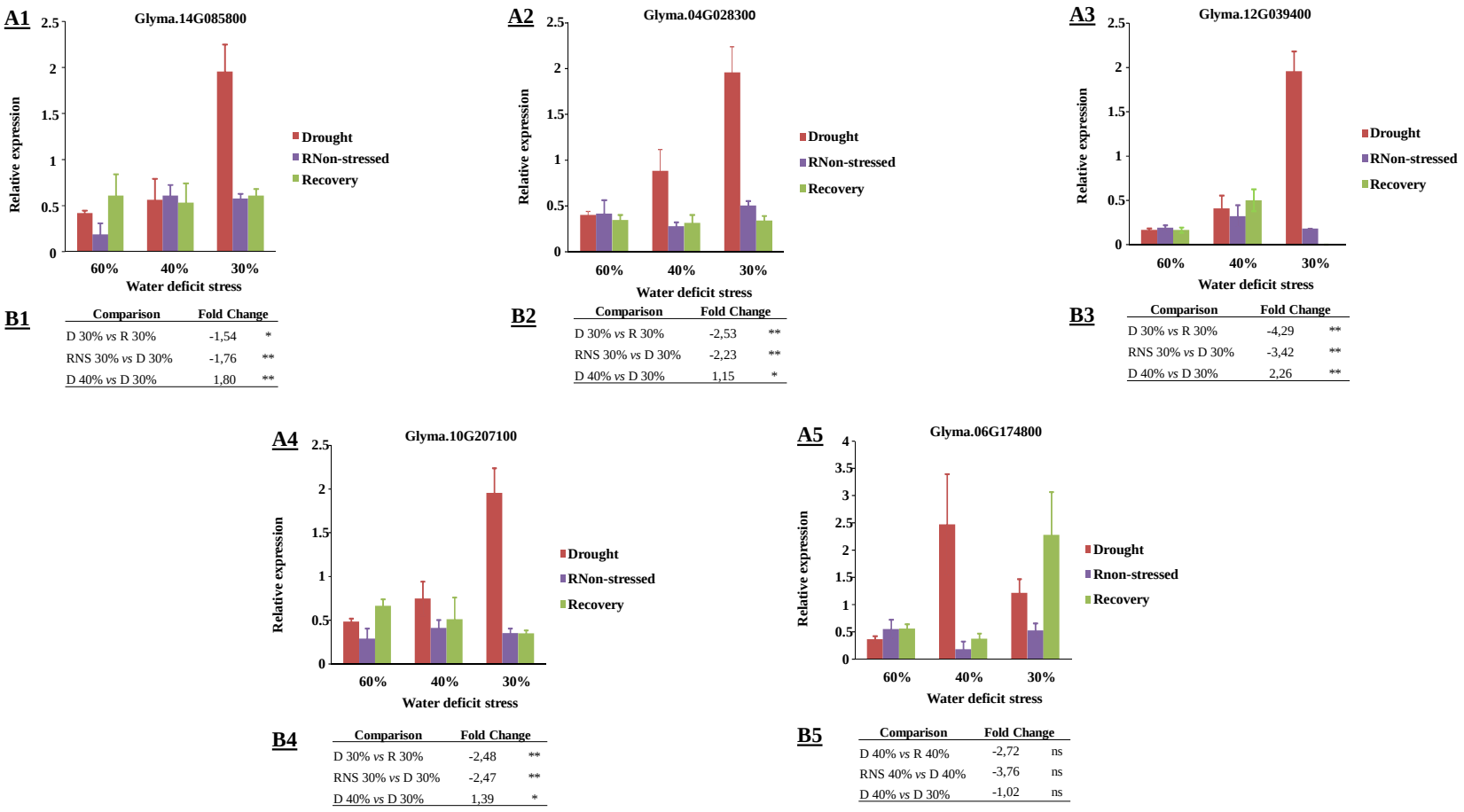
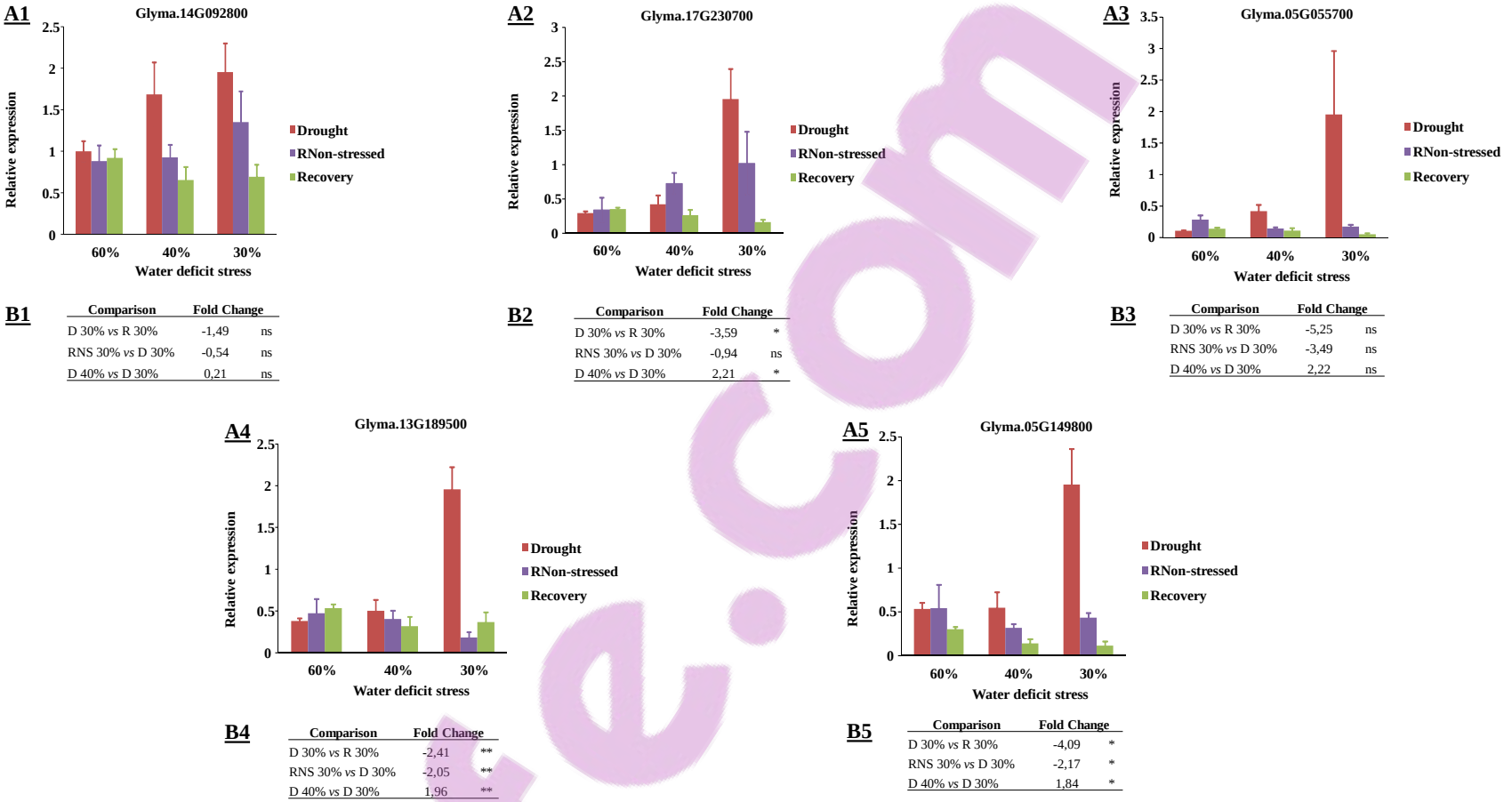


Fig. 4.9: The ability of C1 cysteine proteases to recover after rehydration A1-A5) expressed as relative expression, measured with RT-qPCR and the B1-5) fold changes of interest.

* - Significance between time points were $P \leq 0.05$

** - Significance between time points were $P \leq 0.01$

ns - Non-significant



B1

Comparison	Fold Change	
D 30% *vs* R 30%	-1,49	ns
RNS 30% *vs* D 30%	-0,54	ns
D 40% *vs* D 30%	0,21	ns

B2

Comparison	Fold Change	
D 30% *vs* R 30%	-3,59	*
RNS 30% *vs* D 30%	-0,94	ns
D 40% *vs* D 30%	2,21	*

B3

Comparison	Fold Change	
D 30% *vs* R 30%	-5,25	ns
RNS 30% *vs* D 30%	-3,49	ns
D 40% *vs* D 30%	2,22	ns

B4

Comparison	Fold Change	
D 30% *vs* R 30%	-2,41	**
RNS 30% *vs* D 30%	-2,05	**
D 40% *vs* D 30%	1,96	**

B5

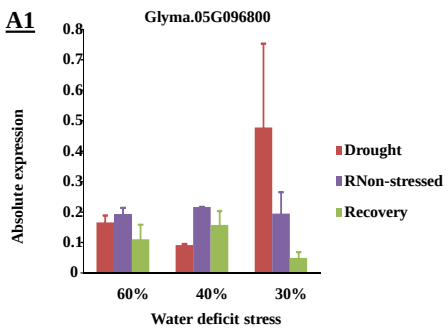
Comparison	Fold Change	
D 30% *vs* R 30%	-4,09	*
RNS 30% *vs* D 30%	-2,17	*
D 40% *vs* D 30%	1,84	*

Fig. 4.10: The ability of A1-A3) C13 cysteine proteases and A4-A5) cystatins to recover after rehydration expressed as relative expression, measured with RT-qPCR and the B1-5) fold changes of interest.

* - Significance between time points were $P \leq 0.05$

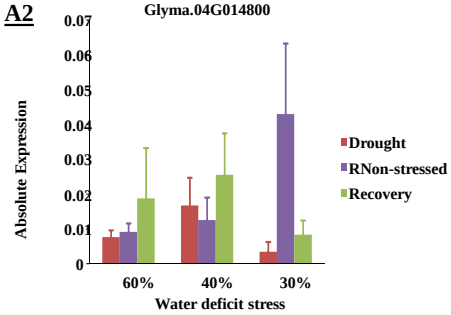
** - Significance between time points were $P \leq 0.01$

ns - Non-significant



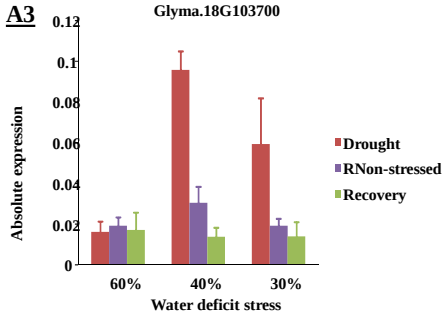
B1

Comparison	Fold Change	
D 30% *vs* R 30%	-3,27	ns
RNS 30% *vs* D 30%	-1,29	ns
D 40% *vs* D 30%	2,38	ns



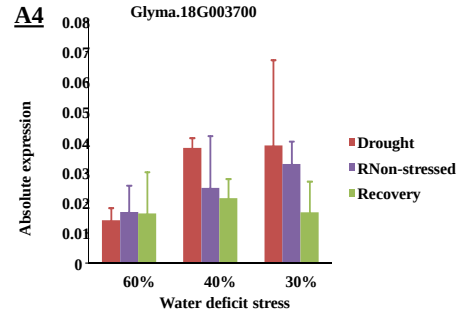
B2

Comparison	Fold Change	
D 40% *vs* R 40%	0,61	ns
RNS 40 % *vs* D 40%	-0,41	ns
D 40% *vs* D 30%	-2,25	ns



B3

Comparison	Fold Change	
D 40% *vs*R 40%	-2,79	**
RNS 40 % *vs* D 40%	-1,65	**
D 40% *vs* D 30%	-0,69	ns



B4

Comparison	Fold Change	
D 40% *vs* R 40%	-0,83	ns
RNS 40 % *vs* D 40%	-0,61	ns
D 40% *vs* D 30%	0,03	ns

Fig. 4.11: The ability of A1-A2) C1 cysteine proteases and A3-A4) cystatins to recover after rehydration expressed as absolute expression, measured with ddPCR and the B1-B4) fold changes of interest.

* - Significance between time points were $P \leq 0.05$

** - Significance between time points were $P \leq 0.01$

ns - Non-significant

4.4 DISCUSSION

There is an urgent need for a new generation of crop cultivars that are more drought tolerant and able to recover after drought as one-third of the world's population lives in areas which is subjected to regular drought stress periods (Kunert *et al*., 2015). Most studies looked at the effect that drought had on soybean growth and nodules but few focused on the process of recovery after rehydration.

The ability of soybean plants and root nodules to recover after induced senescence following rehydration were investigated. It was seen that irreversible inhibition occurred in vegetative plant growth, nodule numbers and function as well as nitrogenase activity. The ability for leaf and nodule water potential to recover after severe drought stress, as seen in this study, mirrors the recovery seen in leaf and nodule water potential after 30 day old soybean were drought stressed for five days and rehydrated for one day (Fellows *et al*., 1986). A similar recovery in leaf and nodule water potential was seen after soybean was drought stressed for 19 days and rehydrated for two days (Song *et al*., 2016).

Although root nodules water potential and moisture content were able to recover even after severe stress, nitrogenase activity did not recover. Serraj *et al*. (1999) showed that nitrogenase activity can recover if there is a sufficient amount of available O_2 in nodules (Serraj *et al*., 1999). However, due to the fact that reduction in nitrogenase activity in drought stressed plants are usually associated with a limitation in O_2, due to a structural change in the oxygen diffusion barrier, less oxygen are available to the symbiosome (Serraj and Sinclair, 1996). This could result in nitrogenase not being able to recover. Determinate plant nodules, as found in soybean, are more sensitive to stress and soil dehydration due to limited meristematic activity when

compared to indeterminate nodules resulting in a low capacity to recover after drought stress (Serraj *et al*., 1999).

Genomic-assisted breeding could be used to select plants that will survive short periods of drought stress. Using cysteine proteases and cystatins as possible molecular markers were proposed due to their presence in soybean root nodules (Kardailsky and Brewin 1996, Espinosa-Victoria *et al*., 2000; Lee *et al*., 2004; Oh *et al*., 2004) and during abiotic stress tolerance (Salas *et al*., 2008; Martínez *et al*., 2012).

Expression of all the C1 and C13 nodule cysteine proteases, induced by drought stress, returned to pre-stress levels. This was also seen in cowpea where three papain-like cysteine proteases showed a full recovery after drought-induced induction (Khanna-Chopra *et al*., 1999), where expression levels returned to pre-stress levels. The recovery of these proteases confirmed their involvement in initiating induced senescence caused by drought stress. The C1 cysteine protease (*WCP2*) in wheat leaves had a lower expression in drought tolerant cultivars during drought stress and non-drought tolerant cultivars, suggesting that the decrease in activity of certain cysteine proteases can be an indication of drought tolerance (Simova-Stoilova *et al*., 2010). Although cysteine protease expression were able to recover to pre-stress expression levels, the degradation as seen in the browning of active nitrogen fixing pink nodule tissue colour, suggest that the induced senescence process were not halted after rehydration. As the tested C1 and C13 cysteine proteases were induced by drought stress only and not rehydration. With future research it will be possible to see if the down-regulation of these genes during severe stress can possibly produce more drought tolerant cultivars by delaying the onset of senescence. One C1 cysteine protease, *Glyma.06G174800*, did recover after drought induction as it was induced at DS, 40% VWC, but the expression increased at R, 30% VWC which was

not seen in any other C1 cysteine protease. *Glyma.06G174800* have 58% similarity to the *Arabidopsis SAG12* gene. The senescence-specific cysteine protease (*SAG12*) is involved in developmental senescence specific cell death and does not accumulate for example until a leaf develops chlorosis (Weaver *et al*., 1998; Gepstein *et al*., 2003). The continued increase in expression after rehydration could be due to its involvement in senescence.

All four tested drought-induced cystatins could recover after their initial drought stress induction. This is contrast to the findings of Pinheiro *et al*. (2005) that found that re-watering lupins increased cystatin expression even further. Pinheiro *et al*. (2005) also suggested that this protein might be involved in protection during recovery. Soybean plants' overexpressing *OCI*, the rice cystatin gene, were also able to recover to the level of photosynthesis before drought more rapidly after drought stress periods (Kunert *et al*., 2015). This opens up the possibility that the over-expression of the drought-induced cystatins, especially in the recovery phase, could lead to limited protein degradation and help plants and nodules recover after drought stress.

CHAPTER 5

Functional analysis of C13 cysteine proteases

ABSTRACT

C13 cysteine proteases (vaculoar processing enzymes, VPE*s*) are involved in senescence and programmed cell death by causing the collapse of the vacuole and the release of different proteases and enzymes. VPEs have also been proposed to be involved in the activation of C1 cysteine proteases. The possible function of two C13 vacuole processing enzymes (*a^{-1}-VPEs*), which was highly expressed in nodules under drought (Chapter 2), was further investigated in *Arabidopsis* mutant plants. Five mutant *Arabidopsis* lines with non-functional mutations in each of the VPE genes, including a null mutation, were subjected to drought stress. Drought stress was simulated as osmotic stress where PEG-8000 was used to lower the water potential of the media. Two mutant plants, *a^{-1}-VPE* and *γ^{-1}-VPE*, had a lower C1 cysteine protease activity than wild type plants. These mutants also had a higher biomass and protein content compared to the WT *Arabidopsis β^{-1}-VPE* and *δ^{-1}-VPE* mutants. It was suggested that nodule C1 cysteine proteases and the C13 cysteine protease, *Glyma.14G085800* and *Glyma.17G230700*, might be a useful indicators for drought-induced premature senescence of soybean root nodules and that localised down regulation of VPEs might be a strategy to reduce the consequences of drought stress in root nodules.

5.1 INTRODUCTION

C13 cysteine proteases, also called vacuolar processing enzymes (VPEs), are enzymes that are similar to mammalian caspases and are involved in senescence as well as programmed cell death. Although very little is still known about the exact function of VPE's, research has shown that they are responsible for the collapse of the vacuole membrane which leads to the release of different enzymes, including proteases, into the cytoplasm (Hara-Nishimura *et al*., 2005). The *Arabidopsis* genome has four VPE genes, two vegetative VPE's (*α-VPE* and *γ-VPE*), a seed type VPE, called *β-VPE* and a VPE that does not fall into one of those categories called δ-*VPE* (Muntz *et al*., 2002). A null mutation of all four VPE genes in *Arabidopsis* stopped lesion formation on leaves after exposure of a mycotoxin, supporting their involvement in programmed cell death (PCD) and senescence (Kuroyanagi *et al*., 2002). The transcript abundance of the *α-VPE* and *γ-VPE* also increased in senescent leaves as well as in plants stressed with jasmonate (Kuroyanagi *et al*., 2005). Albertini *et al*. (2014) showed that plants with a non-functional mutation in the legumain *γ-VPE* were more tolerant to drought stress.

VPE*s* are also implicated in the activation of cysteine proteases due to their ability to remove the I19 inhibitory domain of pre-proteases (Roberts *et al*., 2012). A study in *V. mungo* seeds also showed that a VPE was responsible for the post translational processing of a cysteine protease. This method of activation is different than the method of activation by an autocatalytic mechanism usually seen in Cathepsin-L and Cathepsin-B proteases (Okamoto and Minamikawa, 1995).

During the gene expression profile analysis of C13 cysteine proteases in soybean root nodules (Chapter 2), three VPE proteases were found to be induced by drought stress. A VPE,

Glyma.17G230700 was highly expressed under drought and during senescence, as well as the VPE *Glyma.05G055700,* although this VPE was also expressed in natural senescence as seen in DNS controls (Chapter 3) and in developmental senescence (Van Wyk *et al*., 2014).

Even though it has been suggested that VPE proteases activate other proteases such as C1 cysteine proteases (Roberts *et al*., 2012), it is unclear if this is also the case for soybean nodule C13 cysteine proteases. The objective of this study was to investigate whether mutations in C13 genes causing a loss of function mutation had an influence on C1 cysteine proteases activity under induced senescence conditions. To achieve this, *Arabidopsis* mutants subjected to osmotic stress, to simulate induced senescence, was used to evaluate whether C1 cysteine proteases activity was affected by the non-functional mutation in the different VPE genes. Thereafter, nodule C13 proteases were aligned with *Arabidopsis* C13 cysteine proteases to establish which type of VPE protease they were. This could assist in establishing whether VPE genes, as well as which type of VPE genes would be suitable candidate genes in soybean nodules for silencing and, as a result, improving drought tolerance by delaying soybean root nodules senescence.

5.2 MATERIALS AND METHODS

5.2.1 Plant material

Wild-type *Arabidopsis thaliana* (ecotype, Columbia) and five VPE mutants' (*a^{-1}-VPE, β^{-1}-VPE, δ^{-1}-VPE, γ^{-1}-VPE, null^{-1}-VPE* mutant) seeds (Kuroyanagi *et al.*, 2005) were purchased from the *Arabidopsis* seed stock organisation (https://www.TAIR.org/). Seeds were treated for three days at a temperature of -4 ^{o}C, then surface-sterilised with 100 % bleach and 70 % EtOH. Seeds were germinated on 30 ml of 1/2 strength Murashige and Skoog(MS) *(*Sigma-Aldrich, Germany) agar medium containing 6 mM 4-Morpholineethanesulfonic acid (MES) *(*Sigma-Aldrich, Germany) buffer and 0.03 M sucrose. Seedlings were grown for six weeks (42 days) at 22/18 $^{\circ}$C, 14/10 h light/dark cycle at 150 µmol $m^{-2}s^{-1}$ PAR. Plates with 1/2 strength MS-media were then prepared in the same manner as above but were over-laid with a PEG-8000 solution overnight to facilitate osmotic stress. The remaining PEG solution was carefully removed where after plants were then transferred onto the PEG-8000 overlaid plates for seven days (Verslues and Bray, 2004). The water potential of the agar was -0.09 MPa ± 0.01 for Non-stressed (NS) samples and – 0.5 MPa ± 0.04 for osmotic stressed (OS) samples as measured with a WP4 Dew Point Potential meter (Decagon, USA).

Plants were harvested on the seventh day and were weighed on an AB104-5 Mettler-Tolendo balance to determine the fresh mass (FM). Thermal images were produced with a Testo 800 thermal imager to compare the mutant's tolerance to osmotic stress. The average temperature of the rosettes were determined using the Testo IRSoft Software by calculating the temperature over the rosettes horizontally, vertically and at random (Appendix D, Fig. D1 and Table D1). The *Arabidopsis* plants were then dried in an oven at 60 ºC for 48 hours where after dry mass

(DW) was determined. Whole rosettes of four biological replicates for each mutant were then flash frozen in liquid nitrogen for subsequent analysis

5.2.2 Electron conductivity measurements

A rosette was submerged in distilled water (18.2 MΩ·cm at 25 ºC) for 20 min where after cellular ion leakage was measured with an electrical conductivity meter (Kawai-Yamada *et al*., 2004). Three technical replicates were used for each treatment.

5.2.3 Protein extraction and determination

For *Arabidopsis* characterization, total protein were extracted from rosettes of four plants from all mutants and the wild type. The rosettes were ground into a fine powder using a mortar and pestle and liquid nitrogen. The powder was dissolved in 1 ml extraction buffer (50 mM Tris-HCl, pH 7.5). The suspension was centrifuged at 12 000 rpm for 10 min at 4 °C using an Eppendorf centrifuge. The supernatant containing the protein was then used to quantify the amount of protein.

The protein content of each individual rosette was determined using a protein determination kit (Bio-Rad, UK) based on the method described by Bradford (1976). Bovine serum album (Protein Standard I; Bio-Rad, South Africa) was used to set up a standard curve (Appendix D Fig. D2). A FluoStar plate reader (BMG, Germany) and micro-titre plate was used to determine the absorbance at 595 nm. All reactions were performed in triplicate for the four biological replicates.

5.2.4 Enzyme activity assays

Cathepsin-L like activity was determined in a 50 mM sodium phosphate buffer, pH 6.0, containing 10 mM L-cysteine (Sigma-Aldrich, Germany), together with the 100 µM C1 cysteine protease specific substrate, Z-Phe-Arg-MCA (Sigma-Aldrich, Germany) dissolved in 1 ml of DMSO. For detection, 30 µg total soluble protein was used per assay (Van Wyk *et al*., 2014). Hydrolysis of the fluorogenic substrate was monitored by measuring the released α-amino 4-methylcoumarin (MCA) that is able to fluoresce, using a hydrolysis progress curve as described by Salvesen and Nagase (1989). Fluorescence development was measured over 11 min at 25 °C in a FluoStar plate reader (BMG, Germany) with excitation and emission wavelengths of 360 nm and 450 nm, respectively. Cathepsin L activity provided by C1 cysteine proteases was determined by measuring the difference in fluorescence before and after addition of the cysteine protease inhibitor E-64 (10 nM), which prevents C1 cysteine protease activity. Negative controls were prepared in the same manner as above without the addition of protein extract. All reactions were performed in triplicate for the four biological replicates.

5.2.5 Statistical analysis

To determine statistical significant changes during growth as well as cathepsin-L like activity in *Arabidopsis* mutants, a one-way ANOVA was conducted followed by a Duncan post-test. SPSS© Version 23 and IBM © Software was applied. A P-value of $P \leq 0.05$ were seen as significantly different.

5.2.6 C13 cysteine protease homology search

Coding sequences of the three soybean nodule C13 cysteine proteases were found on the Phytozome database (version 11.0) [https://phytozome.jgi.doe.gov/pz/portal.html] (Schmutz *et al*., 2010). Homology searches were performed against the *Arabidopsis* genome using Blastn (https://blast.ncbi.nlm.nih.gov/) (Altschul *et al.,* 1990). The protein sequence of the VPE gene with the highest identity according to coverage and E-value were then used together with the C13 cysteine protease protein sequence in soybean nodules and were aligned with MAFFT (version 7.3) (http://mafft.cbrc.jp/alignment/server/) (Katoh *et al*., 2002), Appendix D, Fig. D4, to classify which type of VPE protein the two (*Glyma.17G230700* and *Glyma.05G055700)* soybean protein are.

5.3 RESULTS

5.3.1 Plant growth

A functional analysis of C1 cysteine proteases in VPE mutant *Arabidopsis* plants under stress-induced senescence were done using PEG-8000 to induce osmotic stress *in vitro*. Plant growth of all mutants (Fig. 5.1A) were affected by stress as OS plants had less leaves and were undergoing a slight colour change in rosette from a bright green to a pale yellowish green. Thermal images (Fig. 5.1) also showed that OS plants were observed to have a higher temperature due to less moisture than OS plants (Appendix D, Fig. D.1 and Table D1). No distinctive differences in growth could be seen between the different mutants although, as seen in Fig. 5.1B, the wild-type plants under osmotic stress had a lower temperature overall compared to the *Arabidopsis* mutant lines. A significant difference ($P \leq 0.05$) could be seen in the fresh mass of wild type plants, δ^{-1}-*VPE* and γ^{-1}-*VPE* mutants between NS plants and OS plants but no significant difference could be seen in the dry mass (Fig. 5.2 A and B).

Cellular damage and dehydration tolerance of all mutants were measured by examining the electrolyte leakage of plants. Significant differences were seen in mutant lines when NS plants were compared to OS plants (Fig. 5.3), indicating that seven days osmotic stress did affect plant performance. However, due to a lack in material, no comparison could be made in the amount of electrolyte leakage between the different mutants.

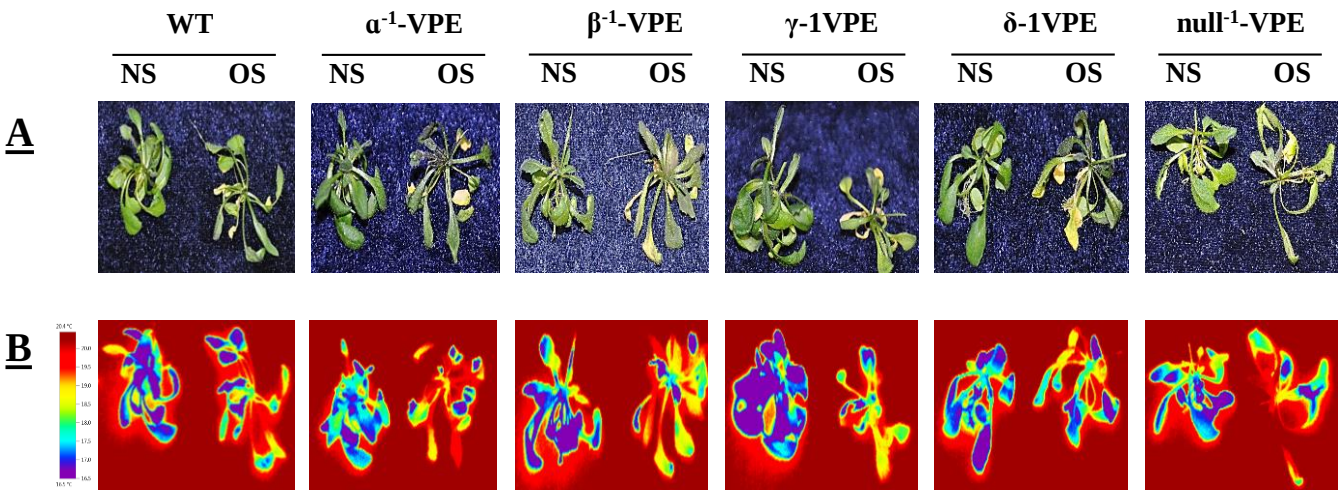


Fig. 5.1: A) *Arabidopsis* mutant's growth after seven days of osmotic stress with their corresponding B) thermal image. NS- Non-Stressed, OS- Osmotic stressed, WT- Wild-type.

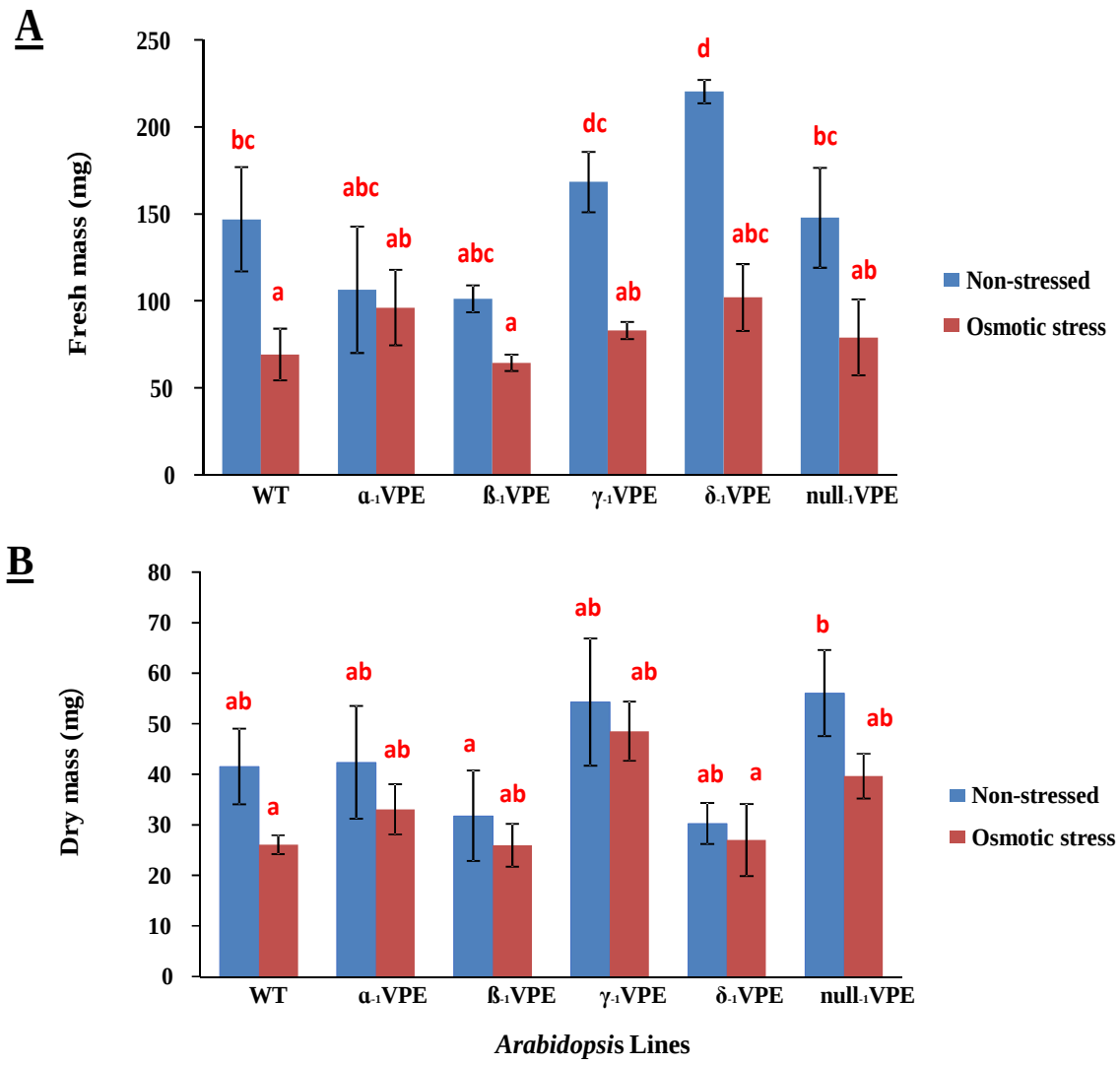


Fig. 5.2: A) *Arabidopsis* mutant's fresh mass and B) dry mass after seven days of osmotic stress.

± SE plants were derived from five biological replicates.

[a-z] Significant differences between all NS and OS samples.

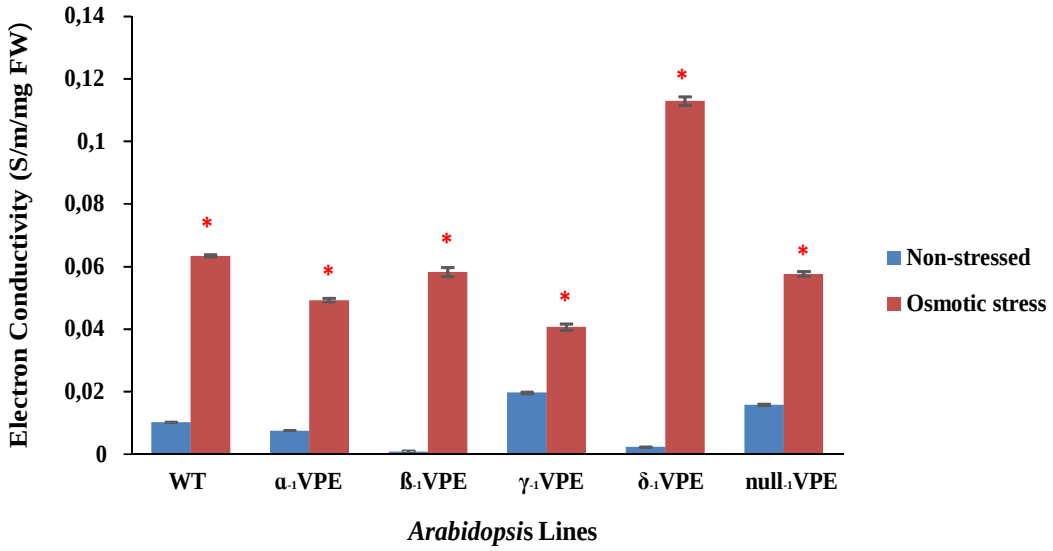


Fig. 5.3: The ion leakage of *Arabidopsis* mutants after seven days of osmotic stress.

± SE plants were derived from three technical replicates from one individual plant.

* Significant differences between all NS and OS samples.

5.3.2. VPE mutant characterization

Cysteine protease activity (C1) measurements were carried out on *Arabidopsis* wild type and mutant lines to determine if VPE proteases activate C1 cysteine proteases. First, the total protein content was determined in all the lines (Fig. 5.4 A). Under low osmotic potential it was seen, due to seven days growth on a PEG-medium, that *Arabidopsis* α^{-1}*-VPE*-mutant plants had significantly ($P \leq 0.05$) more total protein than OS WT *Arabidopsis* plants. This was in contrast to the other mutant lines which did not show a significant diffrence ($P > 0.05$) in total protein content exept or the β^{-1} *-VPE*-mutant which had the least amount ($P \leq 0.05$) of total protein after exposure of osmotic stress conditions.

Fluoresence based kinetic activity assays were used (Fig. 5.4 B and Appendix D, Fig. D3) with a commercial Cathepsin-L substrate. Two mutant lines (α^{-1}*-VPE* and γ^{-1}*-VPE*) had a

considerable greater percentage decrease in C1 cysteine protease activity than WT plants grown under identical conditions (Fig. 5.4 B). It was also seen that both *β*$^{-1}$*-VPE* and *δ*$^{-1}$*-VPE* showed an percentage increase in C1 cysteine proteases activity from NS to OS samples indicating that these types of C13 proteases were not needed for C1 cysteine proteases activation.

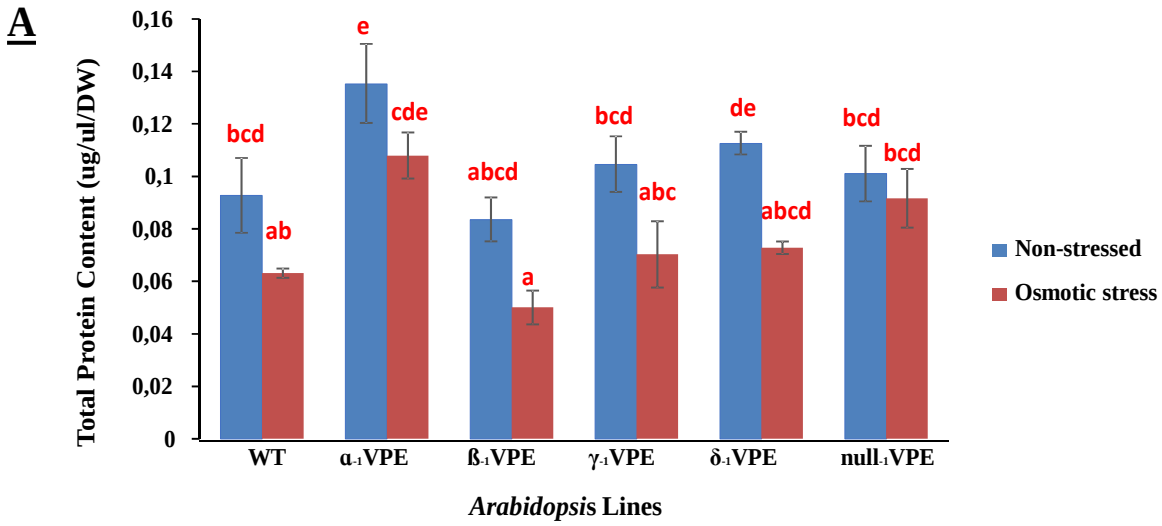


Fig. 5.4: A) Protein content of *Arabidopsis* mutant plants under non-stressed and osmotic stressed conditions, B) Cathepsin-L like activity presented as percentage decrease of E-64 inhibitable cathepsin-L like activity measured as fluorescence units.

± SE plants were derived from six individual plants.

[a-z] Significant differences between all NS and OS samples.

5.3.3 Classification of soybean C13 cysteine proteases

Stress-induced senescence and production of C13 soybean cysteine proteases' homologues to *Arabidopsis* was induced by low osmotic potential conditions in mutant *Arabidopsis* plants. The three C13 cysteine proteases that were induced by osmotic stress in soybean nodules were found on NCBI-BLAST. *Glyma.17G230700* can be considered a *α-VPE* with an E-value of 0 and its protein sequence had 100 % coverage when compared to the *Arabidopsis α-VPE*. This proteases are also very closely related to the *γ-VPE* and has a 95% coverage and also an E-value of 0. *Glyma.14G092800* is also a *α-VPE* with a 98 % coverage and an E-value of 0. The third C13 cysteine proteases (*Glyma.05G055700*) is a *β-VPE* with an E-value of zero and 93 % coverage. Alignments can be found in Appendix D, Fig. D.4.

5.4 DISCUSSION

Primary abiotic stresses such as drought, salt and chilling stress are united by the fact that plant performance is affected by secondary osmotic stress (Verslues *et al*., 2006). Drought stress can be defined as a decrease in the available soil water and quantified by a decrease in water potential (Boyer, 1982). Polyethylene glycol (PEG) can be used *in vitro,* to induce a low water potential as a time, cost and space effective method compared to lengthy field and pot trials. The osmotic stress caused by a lowering of the osmotic potential can also be controlled over different plants and mutant lines so that the water potential are constant (Verslues *et al*., 2006). A decrease in soil water content withdraws water from both the cell wall and the protoplast resulting in cytorrhysis, shrinking both the cell wall and protoplast (Oertli, 1985). Other low molecular solutes such as mannitol, also used to induce osmotic stress caused plasmolysis due to it penetrating the pores of the plants and decreasing the volume of the protoplast while the volume of the cell wall does not change. Because this is different to soil drying, plasmolysis should be avoided (Munns, 2002). To effectively initiate an osmotic stress response, a cytorrhytic treatment should be used which could be caused with a high molecular weight PEG molecule (Verslues *et al*., 2006). Different studies to mimic osmotic stress have been done by using PEG-8000 (Duan *et al*., 2010; Shen *et al*., 2014). Although Lawlor (1970) suggested that PEG-8000 could also be taken up by plants, Verslues *et al*. (2006) showed that maize roots recover to a normal steady-state growth after a period of 48 h. The method used for PEG-infused agar plates allows for low constant transpiration rate and root damage is avoided due to roots not being submerged into PEG-solution. Stress-avoidance is not possible with these conditions as plants equilibrate with the water potential over time. PEG was therefore used to stimulate osmotic stress and induce senescence.

In *Arabidopsis*, VPEs are involved in several processes in the plants such as the maturation of seed storage proteins, leaf senescence, and plant cell death (Hara-Nishimura *et al*. 2005; Hatsugai *et al*. 2006; Hara-Nishimura and Hatsugai, 2011). Using *Arabidopsis* VPE mutants a functional analysis was done to investigate if C13 cysteine proteases induce C1 cysteine protease activity. Previously a VPE gene were shown to activate a cysteine proteases gene in *V. mungo* seeds and cotyledons (Okamoto and Minamikawa, 1995) but the effect the VPE's has on the overall activity of C1 cysteine proteases is unknown. In this study it was shown that *Arabidopsis* α^{-1}*-VPE* mutant plants had a higher protein content associated with lower cysteine protease activity during PEG induced osmotic stress. The γ^{-1}*-VPE* mutant also had lower C1 cysteine proteases activity although it did not have a higher protein content. Lower cysteine protease activity in the α^{-1}*-VPE* and γ^{-1}*-VPE Arabidopsis* mutant was possibly caused by less α and *γ-VPE*-dependant C1 cysteine protease maturation. Less cysteine protease activity likely resulted in less cellular proteolysis and consequently more protein and plant biomass formation. Previously a *γ-VPE Arabidopsis* gene was seen to be involved in drought tolerance and also decreased stomatal opening (Albertini *et al.*, 2014).

Both β^{-1} and δ^{-1}*-VPE Arabidopsis* mutants showed higher cysteine protease activity after drought stress, indicating that these non-vegetative VPE's (Muntz *et al*., 2002) are not involved in C1 cysteine protein maturation in *Arabidopsis* rosettes. Results from a study investigating the processing of pro2S albumins by VPEs in seeds further indicated that vegetative VPEs, *α-VPE* and *γ-VPE,* are not directly required for precursor processing in the presence of *β-VPE,* but partly compensate for any lack in *β-VPE* activity (Shimada *et al.*, 2003).

Even though care was taken to do PEG-mediated osmotic stress as effectively as possible, and although this method is similar to the dry vermiculite system used for crop plants (Sharp *et al*.,

2004), possible problems could occur. The acute loss of water can cause different stress related and ABA-regulated genes to be induced as a rapid stress response (Verslues *et al*., 2006) during the first few hours after treatment with PEG. Therefore measurements were taken after a recovery period of seven days. Due to the nature of the experiment, changes in gene expression due to any abiotic stress that may be caused by the chosen method, would still induce VPE expression due to it being a senescence related proteases (Müntz and Shutov, 2002). Therefore, whether VPEs active C1 cysteine proteases will still be able to be verified using the proposed method and PEG-8000.

As previously mentioned, *Glyma.17G230700* and *Glyma.14G092800, a-VPEs* were highly expressed in our study under drought and during senescence, whereas the *β-VPE Glyma.05G055700* was expressed under natural senescence and drought but only highly expressed under drought. The *a-VPE Glyma.17G230700* is quite distinct from other VPEs. This VPE lacks a GmNAC81/GmNAC30 binding site transducing a cell death signal (Mendes *et al*., 2013) and the VPE *Glyma.17G230700's* function is possibly in controlling protein formation and cysteine protease activity. Localized down-regulation of *α* and *γ-VPE* genes in soybean could possibly result in cultivars which are more tolerant to drought by delaying the initial activation of nodule C1 cysteine proteases. Comparing legume species with non-legume species are not advised due to the complex symbiotic nitrogen fixation process (Benedito *et al*., 2008) but the primary function of a gene is still relevant throughout different plant species. *Arabidopsis* was used as a model plant as the function of a gene was investigated. Whether this gene will still have the exact same function in soybean will need to be investigated in the future by producing gene mutations for soybean and root nodules.

CHAPTER 6

General conclusion and future recommendations

Grain yield of soybean, a commercially important legume crop in South Africa, is affected by prolonged drought stress conditions. As climatic predictions indicate that more frequent and intense weather conditions, with prolonged periods of drought, will be experienced (Zinta *et al*., 2014), drought tolerant soybean cultivars with extended SNF (symbiotic nitrogen fixation) ability is needed to improve food security. Delaying the initiation of nodule senescence thereby protecting nodule tissue during short intermitted drought spells could prolong active nitrogen fixation and aid in the functional recovery of nodules post-stress conditions.

In this PhD study, the drought-induced soybean nodule transcriptome were analysed to look at gene expression profiles focussing on individual components of the cysteine protease–cystatin system to establish gene expression profiles. Several of these components, such as C1 and C13 cysteine proteases as well as cystatins has been characterized and their involvement in drought-induced senescence is evident.

As a first outcome of this study, a drought trial was designed and conducted to induce premature senescence by growing soybean at different levels of drought stress so that induced senescence could be initiated. It was seen that soybean growth and vegetative development was inhibited and the formation of new nodules were repressed due to drought stress. SNF activity in nodules decreased as drought stress intensified.

Secondly, a gene expression profile analysis of drought stressed root nodules was then conducted and expression profiles of C1, C13 cysteine proteases and cystatins identified specific role players in drought-induced nodule senescence. All C1 and C13 cysteine proteases in nodules were identified using homology searches. This was, to my knowledge, the first RNA-Seq analysis of soybean root nodules under drought stress conditions. This study

delivered new knowledge on what types of drought specific cysteine proteases and cystatins are expressed in root nodules and at what level of drought stress these genes are expressed. A recent study on the expression of cysteine proteases and cystatins in developing root nodules (Van Wyk *et al*., 2014) were used as a comparison to identify drought specific genes. Eight C1 cysteine proteases, three C13 cysteine proteases and four cystatins were seen to be induced by drought stress. However, little research is available on the exact function of these proteases and their inhibitors. Future investigations should also include *in vivo* interaction of the cysteine protease-cystatin system, to see how they interact with each other. Reduced cysteine protease and increased cystatin activity could be a strategy to improve nodule life span. However, it needs to be investigated if this strategy will cause phenotypic and physiological changes in soybean plants as seen with ectopically expressing the cystatin, *OCI*, in tobacco which caused stunted growth (Van der Vyver *et al*., 2003). Nodule specific promoters for localized expression could be a solution to phenotypic and physiological changes in soybean plant growth.

Another outcome of this study, was to establish if drought-induced cysteine proteases and cystatins' expression can recover to pre-drought expression levels. The expression of seven analysed C1 cysteine proteases, three C13 cysteine proteases and four cystatins induced during drought stress, were down-regulated to pre-drought stress levels after rehydration. This confirmed these genes ability to be used as drought molecular markers in root nodules. Although gene expression recovered to pre-drought expression levels, nodule tissue did not. Regulating cysteine protease expression at an earlier stage in nodules life cycle could possibly delay nodule senescence and the identified cysteine proteases and cystatins are good candidate genes for creating transgenic soybean plants.

A functional analysis of C13 cysteine proteases was done as a final objective using *Arabidopsis* VPE mutant plants. *Arabidopsis*, the model plant for transgene characterization, has the advantage that gene function data and different gene mutation mutants have been developed. This study showed that C13 cysteine proteases affect the activity of C1 cysteine proteases. A first insight was given that α- and *γ-VPE*s are able to regulate the activity of C1 cysteine proteases. As a *a-VPE* was also seen to be induced by drought in soybean nodules, it would be of interest as a future study to silence *a-VPE* nodule genes to see if this improves the nitrogen fixation capability.

Although this study focussed on cysteine proteases and cystatins, the establishment of a RNA-Seq database for genes expressed during drought stress provides a catalogue of genes that significantly increased in expression at the different levels of drought stress. These genes included *LEA* and defensin like genes which have been associated with drought tolerance (Savitri *et al*., 2013; Graham *et al*., 2007). This database can be used in the future to identify other possible genes due to their uniqueness in expression for marker assisted breeding. The database can also be used to compare gene expression of different abiotic stress conditions.

The hypothesis that, C1 and C13 cysteine proteases and cystatins are involved in premature nodule senescence caused by drought stress was accepted. It was further hypothesised that C13 cysteine proteases could activate C1 cysteine proteases which leads to the premature senescence of root nodules and this hypothesis was also accepted.

The aim of this study, to advance our knowledge of the expression of induced cysteine proteases and cystatins were achieved. Our understanding of how drought stress and rehydration affects soybean root nodule growth, development and gene expression was

improved. This allowed the identification of possible drought markers which could be used to improve root nodules life span and functionality in drought conditions.

LITERATURE CITED

Ahmed SU, Rojo E, Kovaleva V,Venkataraman S, Dombrowski JE, Matsuoka K, Raikhel NV (2000) The plant vacuolar sorting receptor AtELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. Journal of Cell Biology. 149: 1335–1344.

Albertini A, Simeoni F, Glabiati M, Bauer H, Tonelli C, Cominelli E (2014) Involvement of the vacuolar processing enzyme γVPE in response of *Arabidopsis thaliana* to water stress. Biologia Plantarum. 3: 531-538.

Alesandrini F, Frendo P, Puppo A, Hérouart D (2003) Isolation of a molecular marker of soybean nodule senescence. Plant Physiology and Biochemistry. 41: 727-732.

Altschul SF., Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biolology. 215:403-410.

Anjum S, Xie X, Wang L (2011) Morphological, physiological and biochemical responses of plants to drought stress. African Journal of Agricultural Research. 6: 2026-2032.

Appleby CA (1984) Leghemoglobin and Rhizobium respiration. Annual Review of Plant Physiology. 35: 443-478.

Beers EP, Woffenden BJ, Zhao C (2000) Plant proteolytic enzymes: possible roles during programmed cell death. Plant Molecular Biology. 44: 399-415.

Belenghi B, Acconcia F, Trovato M, Perazzolli M, Bocedi A, Polticelli F, Ascenzi P, Delladonne M (2003) AtCYS1, a cystatin from Arabidopsis thaliana, suppresses hypersensitive cell death. 270: 2593–2604.

Benchabane M, Schluter U, Vorster J, Goulet MC, Michaud D (2010) Plant cystatins. Biochimie. 92: 1657-1666.

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, *et al* (2008) A gene expression atlas of the model legume *Medicago truncatula.* The Plant Journal. 55: 504-513.

Benjamin JG, Nielsen DC (2006) Water deficit effects on root distribution of soybean, field pea and chickpea. Field Crops Research. 97: 248-253.

BFAP Baseline – Agricultural Outlook 2014-2023 (2014) SAGIS, Accessed 13 December, <www.sagis.org.z/BFAP_2014.pdf>

Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics. 26: 1783-1785.

Bottari A, Capocchi A, Galleschi L, Jopova A, Saviozzi F (1996) Asparaginyl endopeptidase during maturation and germination of durum wheat. Physiologia Plantarum. 97: 475-480.

Boyer JS (1982) Plant productivity and environment. Science. 218: 443–448.

Bradford MM (1976) A rapid sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry. 72: 248–254.

Brockwell J, Gault R (1976) Effects of irrigation water temperature on growth of some legume species in glasshouses. Animal Production Science. 16: 500-505.

Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Mare C, Tondellia A, Stanca M (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crops Research. 105:1-14.

Chaves MM, Pereira JS, Maroco JP (2003) Understanding plant responses to drought-from genes to the whole plant. Functional Biology. 30:239-264.

Choudhary SP, Tran LS (2011) Phytosterols: perspectives in human nutrition and clinical therapy. Current Medical Chemistry. 18: 4557–4567.

Christoff AP, Turchetto-Zolet AC, Margis R (2014) Uncovering legumain genes in rice. Plant Science. 215: 100– 109.

Chu M, Liu K, Wu H, Yeh K, Cheng Y (2011) Crystal structure of tarocystatin–papain complex: implications for the inhibition property of group-2 phytocystatins. Planta. 234: 243-254.

Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research. 38: 1767-1771.

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska,S, Kopka J, Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. The Plant Journal. 39: 487-512.

Cutforth HW, McGinn SM, McPhee KE, Miller PR (2007) Adaptation of pulse crops to the changing climate of the northern Great Plains. Agronomy Journal. 99: 1684–1699.

Dall E, Brandstetter H (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation, Proceedings of the Natural Academy of Science (USA). 110: 10940–10945.

Daryanto S, Wang L, Jacinthe P (2015) Global synthesis of drought effects on food legume production. PloS one 10: e0127401.

De Silva M, Purcell LC, King CA (1996) Soybean petiole ureide response to water deficits and decreased transpiration. Crop Science. 36: 611-616.

Demirevska K, Simova-Stoilova L, Fedina I, Georgieva K, Kunert K, (2010) Response of Oryzacystatin I Transformed Tobacco Plants to Drought, Heat and Light Stress. Journal of Agronomy and Crop Science. 196:90-99.

Denancé N, Szurek B, Noël LD (2014) Emerging functions of nodulin-like proteins in non-nodulating plant species Plant Cell Physiology. 55: 469-474.

Deu E, Verdoes M, Bogyo M (2012) New approaches for dissecting proteases function to improve probe development and drug discovery. Nature Structural and Molecular Biology. 19: 6-19.

Diop NN, Kidric M, Repellin A, Gareil M, d'Arcy-Lameta A, Pham Thi AT, Zuily-Fodil Y (2004) A multicystatin is induced by drought-stress in cowpea (*Vigna unguiculata (L.) Walp*) leaves. FEBS Letters. 577: 545–550.

Dovring F (1974) Soybeans. Scientific American. 230: 14-15.

Downie JA (2005) Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules. Current Biology. 15: 196-198.

Du Pont L, Alloing G, Pierre O, Msehli SE, Hopkins J, Herouart D, Frendo P (2012) The legume root nodule: from symbiotic nitrogen fixation to senescence. DOI: 10.5772/34438 · Source: InTech.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) AgriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research. 38: 64-70.

Duan Y, Zhang W, Li B, Wang Y, Li K, Han C, Zhang, Y, Li, X (2010) An endoplasmic reticulum response pathway mediates programmed cell death of root tip induced by water stress in *Arabidopsis.* New Phytologist. 3: 681-695.

Dubey VK, Pande M, SingH BK, Jagannadham MV (2007) Papain-like proteases: applications of their inhibitors. African Journal of Biotecnology. 6: 1077-1082.

Dunse KM, Kaasc Q, Guarinoa RF, Bartond PA, Craikc DJ, Andersona MA (2010) Molecular basis for the resistance of an insect chymotrypsin to a potato type II proteinase inhibitor. Proceeding of the National Academy of Science (USA). 107: 15016–15021.

Erickson RO, Michelini FJ (1957) The plastochron index. American Journal of Botany. 44: 297-305.

Eriksson SK, Harryson P (2011) Dehydrins: molecular biology, structure and function. Plant Desiccation Tolerance. Ecological Studies. 215: 289-305.

Escuredo PR, Minchin FR, Gogorcena Y, Iturbe-Ormaetxe I, Klucas RV, Becana M. (1996). Involvement of Activated Oxygen in Nitrate-Induced Senescence of Pea Root Nodules. Plant Physiology. 110: 1187-1195.

Espinosa-Victoria D, Graham PH, Vance C (2000) Host variation in traits associated with crown nodule senescence in soybean. Crop Science. 40: 103-109.

Esteban-Garcia B, Garricardenas JA, Alonso DL, Garcia-Maroto F (2010) A distinct superfamily of papain-like cysteine proteinases regulated by senescence and stresses in Glycine max. Journal of Plant Physiology. 167: 1101–1108.

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

Farooq, Gogoi N, Barthakur S, Baroowa B, Bharadwaj N, Alghamdi SS, Siddique KHM (2016) Drought stress in grain legumes during reproduction and grain filling. Journal of Agronomy and Crop Science. doi.org/10.1111/jac.12169.

Fedorova E, Brown S (2007) Cytochemistry of proteolytic activity and pH status of vacuoles in *Medicago truncatula* root nodules. Russian Journal of Plant Physiology. 54: 25-31.

Fellows RJ, Patterson R P, Raper Jr CD, Harris D (1987) Nodule activity and allocation of photosynthate of soybean during recovery from water stress. Plant Physiology. 84: 456–460.

Fenta B, Driscoll S, Kunert K, Foyer C (2012) Characterization of Drought-Tolerance Traits in Nodulated Soya Beans: The Importance of Maintaining Photosynthesis and Shoot Biomass Under Drought-Induced Limitations on Nitrogen Metabolism. Journal of Agronomy and Crop Science. 198: 92-103.

Fenta BA, Beebe SE, Kunert KJ, Burridge JD, Barlow KM, Lynch JP, Foyer CH (2014) Field phenotyping of soybean roots for drought stress tolerance. Agronomy. 4: 418-435.

Fenta BA, Garcia BM, Foyer CH, Kunert KJ, Du Plessis M, Schluter U (2011) Identification and application of phenotypic and molecular markers for abiotic stress tolerance in soybean. In Soybean - Genetics and Novel Techniques for Yield Enhancement, ed. Dora Krezhova (New York, NY: InTech).

Fernandez-Luquen F, Dendooven L, Munive A, Corlay-Chee L, Serrano-Covarrubias L.M, Espinosa-Victoria D (2008) Micro-morphology of common bean (*Phaseolus vulgaris L.)* nodules undergoing senescence. Acta Physiologiae Plantarum. 30: 545–552.

Findley SD, Cannon S, Varala K, Du J, Ma J, Hudson ME, Birchler JA, Stacey G (2010) A fluorescence in situ hybridization system for karyotyping soybean. Genetics. 185: 727-744.

Fitzgerald MS, McKnight TD, Shippen DE (1996) Characterization and developmental patterns of telomerase expression in plants. Proceedings of the National Academy of Sciences. 93: 14422-14427.

Food and Agriculture Organization of The United Nations Statistics (2014) Accessed 13 December 2016, <www.fao.org/faostat/en/#search/soybean>.

Foyer CH, Lam H-M, Nguyen HT, Siddique KHM, Varshney R, Colmer TD, Cowling W, Bramley H, Mori TA, Hodgson JM, Cooper JW, Miller AJ, Kunert K, Vorster J, *et al*. (2016) Neglecting legumes has compromised global food and nutritional security. Nature Plants. Article number: 16112 (2016) doi:10.1038/nplants.2016.112.

Franchini JC, Debias H, Sacoman A, Nepomuceno AL, Farias JRB (2009) Manejo do Solo para Reduca ˜o das Perdas de Produtividade pela Seca. Londrina:Embrapa Soja.

Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Vale G (2005) Marker assisted selection in crop plants, plant cell. Tissue Organ Culture. 82: 317-345.

Ge Y, Cai Y, Bonneau L, Rotari V, Danon A, McKenzie E, McLellan H, Mach L, Gallois P (2016) Inhibition of cathepsin B by caspase-3 inhibitors blocks programmed cell death in *Arabidopsis.* Cell Death & Differentiation. 23: 1493–1501.

Gepstein S, Sabehi G, Carp M, Hajouj T, Nesher M, Yariv I, Dor C (2003) Large-scale identification of leaf senescence-associated genes. The Plant Journal. 36: 629–642.

Geurts R, Fedorova E, Bisseling T (2005) Nod factor signaling genes and their function in the early stages of *Rhizobium* infection. Current Opinion in Plant Biology. 8: 346–52.

Gijzen M, Kuflu K, Qutob D, Chernys JT (2001) A class I chitinase from soybean seed coat. Journal of Experimental Botany. 52: 2283-2289.

Gil-Quintana E, Lyon D, Staudinger C, Wienkoop S, González EM (2015) Medicago truncatula and Glycine max: different drought tolerance and similar local response of the root nodule proteome. Journal of proteome research. 14: 5240-5251.

Gogorcena Y, Gordon AJ, Escuredo, PR, Minchin FR, Witty JF, Moran JF, Becana M (1997) N2 fixation, carbon metabolism and oxidative damage in nodules of dark stressed Common Bean plants. Plant Physiology. 113: 1193-1201.

Graham M, Silverstein KAT, Van den Bosch KA **(**2007) Defensin-like Genes: Genomic Perspectives on a Diverse Superfamily in Plants. Crop Science. 48: 3–11.

Grudkowska M, Zagdanska B (2004) Multifunctional role of plant cysteine proteinases. Acta Biochimica Polonica. 51: 609-624.

Guerin V, Trinchant JC, Rigaud J (1990) Nitrogen Fixation C(2)H(2) Reduction by Broad Bean (Vicia faba L.) Nodules and Bacteroids under Water-Restricted Conditions. Plant Physiology. 92: 595-601.

Ha CV, Watanabe Y, Tran UT, Le DT, Tanaka M, Nguyen KH, Seki M, Nguyen DV, Tran L (2015) Comparative analysis of root transcriptomes from two contrasting drought-responsive Williams 82 and DT2008 soybean cultivars under normal and dehydration conditions. Frontiers in Plant Science. 6: 551.

Handa K, Yong Son S (1974) On the expression of plant age of soybeans by means of plastochron index. Proceedings of Crop Science Society of Japan. 43: 8–28.

Hara-Nishimura I, Hatsugai N (2011) The role of vacuole in plant cell death. Cell Death Differentiation. 18: 1298-1304.

Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M., Nishimura M (2005) Vacuolar processing enzyme: an executor of plant cell death. Current Opinion in Plant Biology. 8: 404–408.

Hatsugai N, Kuroyanagi M, Nishimura M, Hara-Nishimura I (2006) A cellular suicide strategy of plants: vacuole-mediated cell death. Apoptosis. 6: 905–911.

Hsien WLY (2015) Utilization of vegetable oil as bio-lubricant and additive. In Towards Green Lubrication in Machining. Springer Briefs in Green Chemistry for Sustainability. 7–17.

Huang YM, Xiao BZ, Xiong LZ (2007) Characterization of a stress responsive proteinase inhibitor gene with positive effect in improving drought resistance in rice. Planta. 226: 73–85.

Ivessa EN, Kitzmüller C, de Virgilio M (1999) Endoplasmic-reticulum-associated protein degradation inside and outside of the endoplasmic reticulum. Protoplasma. 207: 16-23.

Jones LR, Zhang L, Sanborn K, Jorgensen AO, Kelley J (1995) Purification, Primary Structure, and Immunological Characterization of the 26-kDa Calsequestrin Binding Protein (Junctin) from Cardiac Junctional Sarcoplasmic Reticulum. Journal of Biological Chemistry. 270: 30787-30796.

Jones SI, Vodkin LO (2013) Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS One. 8: e59270.

Jongdee B, Fukai S, Cooper M (2002) Leaf water potential and osmotic adjustment as physiological traits to improve drought tolerance in rice. Field Crops Research. 76:153-163.

Jury WA, Vaux HJ (2007) The emerging global water crisis: managing scarcity and conflict between water users. Advances in Agronomy. 95: 1–76.

Katoh K, Misawa K, Kuma K, Miyataa T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acid Research. 30: 3059–3066.

Kardailsky IV, Brewin NJ (1996) Expression of cysteine protease genes in pea nodule development and senescence. Molecular Plant Microbe Interactions. 9: 689-695.

Kawai-Yamada M, Ohori Y, Uchimiya H (2004) Dissection of Arabidopsis Bax inhinbitor-1 suppressing Bax-, hydrogen peroxide-, and salicylic acid- induced cell death. Plant Cell. 16: 21-32.

Keyser HH, Li F (1992) Potential for increasing biological nitrogen fixation in soybean. Plant and Soil. 141: 119-135.

Keyster M, Adams R, Klein A, Ludidi N (2013) Nitric oxide (NO) regulates the expression of single-domain cystatins in Glycine max (soybean). Plant Omics Journal. 6:183-192.

Khanna-Chopra R, Srivalli B, Ahlawat Y (1999) Drought induces many forms of cysteine proteases not observed during natural senescence. Biochemical and Biophysical Research Communications. 255: 324-327.

Kidrič M, Kos J, Sabotič J (2013) Proteases and their endogenous inhibitors in the plant response to abiotic stress. Botanica Serbica. 38: 139-158.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 14: 36-46.

Kim KH, Kang YJ, Kim DH, Yoon MY, Moon JK, Kim MY, Van K, Lee SH (2011) RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and -susceptible alleles. DNA Results. 18: 483-497.

King CA, Purcell LC (2005) Inhibition of N2 fixation in soybean is associated with elevated ureides and amino acids. Plant Physiology. 137: 1389-1396.

Ku Y, Wen C, Lam H, Li M, Wan-, Liu X, Yung Y (2013) Drought stress and tolerance in soybean. In A Comprehensive Survey of International Soybean Research - Genetics,

Physiology, Agronomy and Nitrogen Relationships, ed. J. E. Board (New York, NY: InTech). 209–237.

Kucharik CJ, Ramankutty N (2005) Trends and variability in U.S. corn yields over the 20th century. Earth International. 9: 1–29.

Kunert K J, Van Wyk S G, Cullis CA, Vorster BJ, Foyer CH (2015) Potential use of phytocystatins in crop improvement, with a particular focus on legumes. Journal of Experimental Botany. 66: 3559–3570.

Kunert KJ, Voster BJ, Feant BA, Kibido T, Dionisio G, Foyer CH (2016) Drought stress response in soybean root and nodules. Frontiers in Plant Science doi: 10.3389/fpls.2016.01015.

Kuroyanagi M, Nishimura M, Hara-Nishimura I (2002) Activation of Arabidopsis vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. Plant and Cell Physiology. 43: 143–151.

Kuroyanagi, M, Yamada K, Hatsugai N, Kondo M, Nishimura M, Hara-Nishimura I. (2005) Vacuolar processing enzyme is essential for mycotoxin-induced cell death in *Arabidopsis thaliana*. Journal of Biological Chemistry. 280: 32914-32920.

Ladrera R, Marino D, Larrainzar E, Gonzalez EM, Arrese-Igor C (2007) Reduced carbon availability to bacteroids and elevated ureides in nodules, but not in shoots, are involved in the nitrogen fixation response to early drought in soybean. Plant Physiology. 145: 539-546.

Larrainzar E, Wienkoop S, Scherling C, Kempa S, Ladrera R, Arrese-Igor C, González EM (2009) Carbon metabolism and bacteroid functioning are involved in the regulation of nitrogen fixation in *Medicago truncatula* under drought and recovery. Molecular Plant-Microbe Interactions. 22: 1565–76.

Lawlor DW (1970) Absorption of polyethylene glycol by plants and their effects on plant growth. New Phytologist. 69: 501–513.

Lay F, Anderson M (2005) Defensins-components of the innate immune system in plants. Current Protein and Peptide Science. 6: 85-101.

Lee H, Hur CG, On CJ, Kim BH, Park SY, An CS (2004) Analysis of the root nodule-enhanced transcriptome in soybean. Molecules and Cells. 18: 53-62.

Li Y, Zhou L, Li Y, Chen D, Tan X, Lei L, Zhou J (2008) A nodule-specific plant cysteine proteinase, AsNODF32, is involved in nodule senescence and nitrogen fixation activity of the green manure legume *Astragalus sinicus*. New Phytology. 180: 185-192.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. The Plant Journal. 63: 86-99.

Lohar DP, Sharopova N, Endre G, Penuela S, Samac D, Town C, Silverstein KA, VandenBosch KA (2006) Transcript analysis of early nodulation events in Medicago truncatula. Plant Physiology. 140: 221–234.

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: An overview. Archives Biochemistry Biophysics. 444: 139-158.

Manavalan LP, Guttikonda SK, Tran LS, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiology. 50: 1260-1276.

Marino D, Frendo P, Ladrera R, Zabalza A, Puppo A, Arrese-Igor C, Gonzalez EM (2007) Nitrogen fixation control under drought stress. Localized or systemic? Plant Physiology. 143: 1968-1974.

Márquez-Garcia B, Shaw D, Cooper JW, Karpinska B, Quain MD, Makgopa EM, Kunert K, Foyer CH (2015) Redox markers for drought-induced nodule senescence, a process occurring after drought-induced senescence of the lowest leaves in soybean (Glycine max). Annuals of Botany. 116: 497-510.

Martinez M, Cambra I, Gonzalez-Melendi P, Santamaria ME, Diaz I (2012) C1A cysteine-proteases and their inhibitors in plants. Physiologia Plantarum. 145: 85–94.

Martinez M, Diaz-Mendoza M, Carrillo L, Diaz I (2007) Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. FEBS Letters. 581: 2914–2918.

Massonneau A, Condamine P, Wisniewski JP, Zivy M, Rogowsky, PM (2005) Maize cystatins respond to developmental cues, cold stress and drought. Biochimica et Biophysica Acta—Gene Structure and Expression. 1729: 186–199.

Matamoros MA, Baird LM, Escuredo PR, Dalton DA, Minchin FR, IturbeOrmaetxe I, Rubio MC, Moran JF, Gordon AJ, Becana M (1999) Stress induced legume root nodule senescence. Physiological, biochemical and structural alterations. Plant Physiology. 121: 97-112.

Mendes GC, Reis PAB, Calil IP, Carvalho HH, Arogão FJL, Fontes EPB (2013) GmNAC30 and GmNAC81 integrate the endoplasmic reticulum stress- and osmotic stress-induced cell death responses through a vacuolar processing enzyme. Plant Biology. 110: 19627–19632.

Merbach W, Schilling G (1980) Aufnahme, Transport und Verwertung von 15N-markiertem Mineralstickstoff durch Weißlupinen im Gefäßversuch. Arch Acker-Pflanzenbau Bodenkd. 24: 39–46.

Muneer S, Ahmad J, Bashir H (2012) Proteomics of nitrogen fixing nodules under various environmental stresses. Plant Omics Journal. 5: 167-176.

Munns, R. (2002) Comparative physiology of salt and water stress. Plant Cell Environment. 25: 239–250.

Muntz K and Shutov AD (2002) Legumains and their function in plants. Trends Plant Science. 7: 340–344

Muntz K, Blattner FR, Shutov AD (2002) Legumains – a family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants. Journal of Plant Physiology. 159: 1281–1293.

Nuruzzaman M, Sharoni AM, Kikuchi S (2013) Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Frontiers in Microbiology doi: 10.3389/fmicb.2013.00248.

O'Rourke JA, Bolon Y-T, Bucciarelli B, Vance CP (2014) Legume genomics: understanding biology through DNA and RNA sequencing. Annals of Botany. 113: 1107–1120.

Oertli JJ (1985) The response of plant cells to different forms of moisture stress. Journal of Plant Physiology. 121: 295–300.

Oh CJ, Lee H, Kim HB, An CS (2004) Isolation and characterization of a root nodule-specific cysteine proteinase cDNA from soybean. Journal of Plant Biology. 47: 216-220.

Okamoto T, Minamikawa T (1995) Purification of a processing enzyme (VmPE-1) that is involved in post-translational processing of a plant cysteine endopeptidase (SH-EP). European Journal of Biochemistry. 231: 300-305.

Oldroyd GED, Murray JD, Poole PS, Downie JA (2011). The rules of engagement in the legume-rhizobial symbiosis. Annual Review Genetics. 45: 119–144.

Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Passioura JB (2006) The perils of pot experiments. Functional Plant Biology. 33: 1075-1079.

Pfeiffer NE, Torres CM, Wagner FW (1983) Proteolytic activity in soybean root nodules: activity in host cell cytosol and bacteroids throughout physiological development and senescence. Plant Physiology. 71: 797-802.

Pierre O, Hopkins J, Combier M, Baldacci F, Engler G, Brouquisse R, Hérouart D, Boncompagni E (2014) Involvement of papain and legumain proteinase in the senescence process of Medicago truncatula nodules. New Phytologist. 202: 849–863.

Pimratch S, Jogloy S, Vorasoot N, Toomsan B, Patanothai A, Holbrook C (2008) Relationship between biomass production and nitrogen fixation under drought-stress conditions in Peanut genotypes with different levels of drought resistance. Journal of Agronomy and Crop Science. 194: 15-25.

Pinheiro C, Kehr J, Ricardo CP (2005) Effect of water stress on lupin stem protein analysed by two-dimensional gel electrophoresis. Planta. 221: 716–728.

Pladys D, Vance CP (1993) Proteolysis during development and senescence of effective and plant gene-controlled ineffective alfalfa nodules. Plant Physiology. 103: 379-384.

Preisig O, Anthamatten D, Hennecke H (1993) Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. Proceeding of the National Academy of Science. 90: 3309–3313.

Prins A, van Heerden PDR, Olmos E, Kunert KJ, Foyer CH. (2008) Cysteine proteinases regulate chloroplast protein content and composition in tobacco leaves: a model for dynamic

interactions with ribulose-1,5- bisphosphate carboxylase/oxygenase (Rubisco) vesicular bodies. Journal of Experimental Botany. 59: 1935–1950.

Puppo A, Groten K, Bastian F, Carzaniga R, Soussi M, Lucas MM, De Felipe MR, Harrison J, Vanacker H, Foyer CH (2005) Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. New Phytology. 165: 683-701.

Purcell L, Sinclair TR (1995) Nodule gas exchange and water potential response to rapid imposition of water deficit. Plant Cell Environment. 18: 179-187.

Quain MD, Makgopa ME, Cooper JW, Kunert KJ, Foyer CH (2015) Ectopic phytocystatin expression increases nodule numbers and influences the responses of soybean (*Glycine Max*) to nitrogen deficiency. Phytochemistry. 112: 179–87.

Rees DC, Akif Tezcan F, Haynes CA, Walton MY, Andrade S, Einsle O, Howard JB (2005) Structural basis of biological nitrogen fixation, vol. 363.

Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stüber K, Colby T, Overkleeft HS, Bogyo M, Van der Hoorn RAL (2012) Sub classification and Biochemical Analysis of Plant Papain-Like Cysteine Proteases Displays Subfamily-Specific Characteristics. Plant Physiology. 158: 1583-1599.

Roberts IN, Caputo C, Criado MV, Funk C (2012) Senescence-associated proteases in plants. Physiology Plantarum. 145: 130-139.

Salas CE, Gomes MT, Hernandez M, Lopes MT (2008) Plant cysteine proteinases: evaluation of the pharmacological activity. Phytochemistry. 69: 2263-2269.

Salvesen G, and Nagase H, (1989) Inhibition of proteolytic enzymes. Proteolytic enzymes: A practical approach. 83-104.

Savitri E, Basuki N, Aini N, Arumingtyas E (2013) Identification and characterization drought tolerance of gene *LEA*-D11 soybean (*Glycine max L. Merr*) based on PCR-sequencing. American Journal of Molecular Biology. 3: 32-37.

Schaller A (2004). A cut above the rest: the regulatory function of plant proteases. Planta. 220: 183-197.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. Nature. 463: 178-183.

Scott W, Aldrich SR (1983) Modern Soybean Production. S & A Publication. Champaign. IL. 209-219

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. The Plant Journal. 31: 279-292.

Serraj R, Sinclair TR (1996) Inhibition of nitrogenase activity and nodule oxygen permeability by water deficit. Journal of Experimental Botany. 47: 1067–1073.

Serraj R, Sinclair TR, Purcell LC (1999) Symbiotic N2 fixation response to drought. Journal of Experimental Botany. 50: 143-155.

Serraj R, Vadez V, Sinclair T (2001) Feedback regulation of symbiotic Nitrogen fixation under drought stress. Agronomy. 21: 621-626.

Severin AJ, Woody JL, Bolon Y, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE (2010) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biology 10: 160-176.

Sha A, Li M, Yang P (2016) Identification of phosphorus deficiency responsive proteins in a high phosphorus acquisition soybean (Glycine max) cultivar through proteomic analysis. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 5: 427–434.

Sharp RE, Poroyko V, Hejlek LG, Spollen WG, Springer GK, Bohnert HJ, Nguyen HT (2004) Root growth maintenance during water deficits: physiology to functional genomics. Journal of Experimental Botany. 55: 2343–2351.

Shen J, Xing T, Yuan H, Liu z, Jin Z, Zang L, Pei Y (2014) Hydrogen sulphide improves drought tolerance in *Arabidopsis thaliana* by MicroRNA expressions. PLoS ONE 8(10): e77047. doi:10.1371/journal.pone.007704.

Sheokand S, Dahiya P, Vincent JL, Brewin NJ (2005) Modified expression of cysteine protease affects seed germination, vegetative growth and nodule development in transgenic lines of *Medicago truncatula*. Plant Science. 169: 966–975.

Shimada T, Yamada K, Kataoka M, et al. (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in Arabidopsis thaliana. Journal of Biological Chemistry. 278: 32292-32299.

Siddique M, Hamid A, Islam M (2000) Drought stress effects on water relations of wheat. Botanical Bulletin of Academia Sinica. 41: 35-39.

Simova-Stoilova L, Vaseva I, Grigorova B, Demirevska K, Feller U (2010) Proteolytic activity and cysteine protease expression in wheat leaves under severe soil drought and recovery. Plant Physiology and Biochemistry. 48: 200-206.

Sinclair TR, and Serraj R (1995) Dinitrogen fixation sensitivity to drought among grain legume species. Nature. 378-344.

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

Song L, Prince S, Valliyodan B, Joshi T, dos Santos, Joao V Maldonado, Wang J, Lin L, Wan J, Wang Y, Xu D (2016) Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. BMC Genomics. 17: 57-65.

Sprent JI (1972) The effects of water stress on nitrogen-fixing root nodules. New Phytologist. 71: 603-611.

Stougaard J (2000) Regulators and regulation of legume root nodule development. Plant Physiology. 124: 531-540.

Swaraj K, Bishnoi N (1996) Physiological and biochemical basis of nodule senescence in legumes: a review. Plant physiology and Biochemistry. 23: 105-116.

Tajima T, Yamaguchi A, Matsushima S, Satoh M, Hayasaka S, Yoshimatsu K, Shioi Y (2011) Biochemical and molecular characterization of senescence-related cysteine protease–cystatin complex from spinach leaf. Physiologia Plantarum. 141: 97–116.

Than ME, Helm M, Simpson DJ, Lottspeich F, Huber R, Gietl C (2004) The 2.0 A crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. Journal of Molecular Biology. 336: 1103–1116.

Timmers ACJ, Soupène E, Auriac MC, de Billy F, Vasse J, Boistard P, Truchet G (2000) Saprophytic intracellular rhizobia in alfalfa nodules. Molecular Plant-Microbe Interactions. 13: 1204-1213.

Todd CD, Tipton PA, Blevins DG, Piedras P, Pineda M, Polacco JC (2006) Update on ureide degradation in legumes. Journal of Experimental Botany. 57: 5-12.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology. 28: 511-515.

Turk V, Turk B, Turk D (2001) Lysosomal cysteine proteases: facts and opportunities. EMBO Journal. 20: 4629–4633.

Turner G, Gibson AH (1980) Measurement of nitrogen fixation by indirect means. In Methods for Evaluating Biological Nitrogen Fixation. Edited by F. J. Bergersen. Chichester: Wiley: 111 -138.

Vadez V and Sinclair TR (2000) Ureide degradation pathways in intact soybean leaves. Journal of Experimental Botany. 51: 1459–1465.

Valenzuela-Vazquez M, Escobedo-Mendoza A, Almanza-Sandoval JL, Ríos-Torres A (1997) Pressure bomb. Final report: New Mexico State University

Van de Velde W, Guerra JCP, De Keyser A, De Rycke R, Rombauts S, Maunoury N, Mergaert P, Kondorosi E, Holsters M, Goormachtig S (2006) Aging in legume symbiosis. A molecular view on nodule senescence in *Medicago truncatula*. Plant Physiology. 141: 711-720.

Van Der Hoorn RAL (2008) Plant proteases: from phenotypes to plant mechanisms. Annual Review of Plant Biology. 59: 191-223.

van der Vyver C, Schneidereit J, Driscoll S, Turner J, Kunert K, Foyer CH (2003) Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances chilling tolerance. Plant Biotechnology Journal. 1: 101–112.

Van Heerden PD, Strasser RJ, Krüger GH (2004) Reduction of dark chilling stress in N2-fixing soybean by nitrate as indicated by chlorophyll a fluorescence kinetics. Physiology Plantarum. 121: 239-249.

Van Heerden P, De Beer M, Mellet D, Maphike H, Foit W (2007) Growth media effects on shoot physiology, nodule numbers and symbiotic nitrogen fixation in soybean. South African Journal Botany. 73: 600-605.

Van Heerden PDR, Kiddle G, Pellny TK, Mokwala PW, Jordaan A, Strauss AJ, de Beer M, Schluter U, Kunert KJ, Foyer CH (2008) Regulation of respiration and the oxygen diffusion barrier in soybean protect symbiotic nitrogen fixation from chilling induced inhibition and shoots from premature senescence. Plant Physiology. 148: 316-327.

Van Wyk SG, Du Plessis M, Cullis CA, Kunert KJ, Vorster BJ (2014) Cysteine protease and cystatin expression and activity during soybean nodule development and senescence. BMC Plant Biology. 14: 294-014-0294-3.

Verslues PE and Bray EA (2004) LWR1 and LWR2 are required for osmoregulation and osmotic adjustment in *Arabidopsis*. Plant Physiology. 136: 2831–2842.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu J-K (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. The Plant Journal. 45: 523-539.

Vorster BJ, Schlüter U, Du Plessis M, Van Wyk S, Makgopa ME, Ncube I, Quain MD, Kunert K, Foyer CH (2013) The cysteine protease–cysteine protease inhibitor system explored in soybean nodule development. Agronomy. 3: 550-570.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 10: 57–63.

Wanga X, Zhanga H, Suna G Jina Y, Qiu L (2014) Identification of active VQ motif-containing genes and the expression patterns under low nitrogen treatment in soybean. Gene. 543: 237–243.

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

Yamada K, Shimada T, Nishimura M, Hara-Nishimura I (2005) A VPE family supporting various vacuolar functions in plants. Plant Physiology. 123: 369-375.

Young EG, Conway CF (1942) On the estimation of allantoin by the Rimini-Schryver reaction, Journal of Biological Chemistry. 142: 839-853.

Yuan S, Li R, Wang L, Chen H, Zhang C, Chen L, Hao Q, Shan Z, Zhang X, Chen Z, et al. (2016) Search for Nodulation and Nodule Development-Related Cystatin Genes in the Genome of Soybean (Glycine max). Frontiers in Plant Science 7:1595-1620.

Zahran HH (1999) *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiology and Molecular Biology Reviews. 63: 968-989.

Zinta G, AbdElgawad H, Domagalska MA, Vergauwen L, Knapen D, Nijs I, Janssens IA, Beemster GT, Asard H (2014) Physiological, biochemical, and genome-wide transcriptional

analysis reveals that elevated CO2 mitigates the impact of combined heat wave and drought stress in *Arabidopsis thaliana* at multiple organizational levels. Global Change Biology. 20: 3670-3685.

APPENDIXES

APPENDIX A

Table A.1 and A.2: A1) Fresh mass and A2) dry mass of plant organs in grams (g), used for moisture content determination.

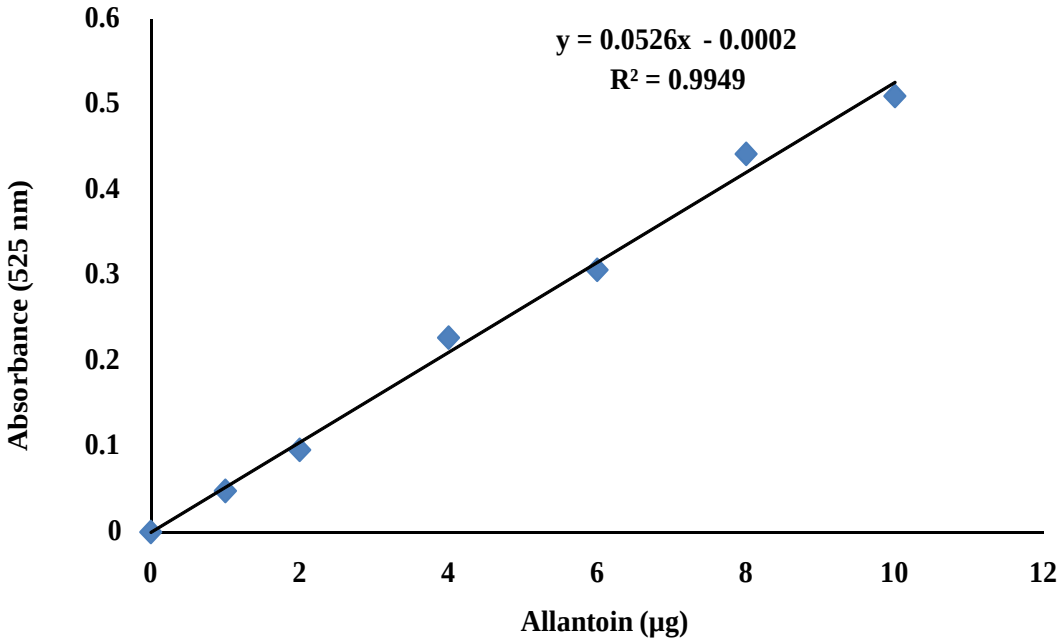
A1

	Water deficit treatments					
Organs	**DNS (60 %)**	**60 %**	**DNS (40 %)**	**40 %**	**DNS (30 %)**	**30 %**
Young leaves	1.94 ± 0.30	1.59 ± 0.10	1.89 ± 0.20	2.35 ± 0.30	1.99 ± 0.20	0.93 ± 0.20
Old leaves	3.84 ± 0.50	2.99 ± 0.20	9.02 ± 1.60	3.36 ± 0.23	18.94 ± 1.50	2.33 ± 0.39
Root	7.74 ± 0.80	9.48 ± 1.10	13.71 ± 1.80	11.26 ± 0.90	26.45 ± 5.60	6.65 ± 0.70
Shoots	4.04 ± 0.73	2.85 ± 0.28	7.86 ± 1.40	3.32 ± 0.20	13.82 ± 1.50	2.40 ± 0.30
Crown nodules (mg)	318.57 ± 53.9	221.25±44.7	536.67 ± 88.5	178.57±30.12	611.25±104.6	111.43±25.12

A2

	Water deficit treatments					
Organs	**DNS (60 %)**	**60 %**	**DNS (40 %)**	**40 %**	**DNS (30 %)**	**30 %**
Young leaves	0.31 ± 0.06	0.29 ± 0.03	0.33 ± 0.04	0.49 ± 0.07	0.37 ± 0.03	0.38 ± 0.10
Old leaves	0.53 ± 0.05	0.59 ± 0.12	1.60 ± 0.33	0.66 ± 0.04	3.12 ± 0.33	0.62 ± 0.05
Root	0.71 ± 0.07	2.05 ± 0.33	1.97 ± 0.35	4.08 ± 0.85	3.37 ± 0.45	3.28 ± 0.22
Shoots	0.72 ± 0.15	0.59 ± 0.07	1.47 ± 0.28	0.81 ± 0.06	2.56 ± 0.30	0.81 ± 0.13
Crown nodules (mg)	56.39 ± 15.07	51.09±10.66	98.80±16.77	46.77 ± 3.88	113.50 ±17.02	53.12 ± 9.49

Fig. A.2: Standard curve of the absorbance (525 nm) of 0-8 µg of Allantion used for the calculation of ureides content.



APPENDIX B

Table B.1: Tophat 2 settings and parameters used during RNA-Seq data analysis.

Input parameter	Value
Is this library mate-paired?	Paired
RNA-Seq FASTQ file, forward reads	13: FASTQ Trimmer 1/1
RNA-Seq FASTQ file, reverse reads	14: FASTQ Trimmer 1/2
Mean Inner Distance between Mate Pairs	12
Std. Dev for Distance between Mate Pairs	20
Report discordant pair alignments?	Yes
Use a built in reference genome or own from your history	history
Select the reference genome	95: Gmax_275_v2.0.fa
TopHat settings to use	Full
Max realign edit distance	1000
Max edit distance	2
Library Type	FR Unstranded
Final read mismatches	2
Use bowtie -n mode	No
Anchor length (at least 3)	8
Maximum number of mismatches that can appear in the anchor region of spliced alignment	0
The minimum intron length	70
The maximum intron length	50000
Allow indel search	Yes
Max insertion length.	3
Max deletion length.	3
Maximum number of alignments to be allowed	20
Minimum intron length that may be found during split-segment (default) search	50
Maximum intron length that may be found during split-segment (default) search	50000
Number of mismatches allowed in each segment alignment for reads mapped independently	2
Minimum length of read segments	25
Use Own Junctions	Yes
Use Gene Annotation Model	Yes
Gene Model Annotations	Gmax_275_Wm8.2.a2.v1 .gene_exons.gff3
Use Raw Junctions	No
Only look for supplied junctions	No

Table B.2: Cufflinks settings and parameters used during RNA-Seq data analysis.

Input Parameter	**Value**
SAM or BAM file of aligned RNA-Seq reads	116: Tophat2 D1 accepted_hits
Max Intron Length	30000
Min Isoform Fraction	0.05
Pre MRNA Fraction	0.05
Perform quartile normalization	Yes
Use Reference Annotation	Use reference annotation
Reference Annotation	96:Gmax_275_Wm8.2.a2.v1.gene_exons.gff3
Perform Bias Correction	No
Use multi-read correct	Yes
Use effective length correction	Yes

Table B.3: Cuffdiff settings and parameters used during RNA-Seq data analysis.

Input Parameter	**Value**
Transcripts	96: Gmax_275_Wm8.2.a2.v1.gene_exons.gff3
Name	DNon-Stressed 60%
Add replicate	11: Tophat2 DNon-stressed 60%: accepted_hits
Name	Drought 60%
Add replicate	12: Tophat2 Drought 60%: accepted_hits
Library normalization method	classic-fpkm
Dispersion estimation method	blind
False Discovery Rate	0.05
Min Alignment Count	5
Use multi-read correct	Yes
Perform Bias Correction	No
Include Read Group Datasets	No
Set Additional Parameters?	No

Table B.4: C1-cysteine proteases identified in soybean root nodules through RNA-Seq with a sequence similar to papain.

* Indicates C1-Proteases seen as transcriptionally active sequence.

	FPKM						
	DNS			DS			
C1	DNS (60 %)	DNS (40 %)	DNS (30 %)	60 %	40 %	30 %	
Glyma.06G014700	0.3	0.2	0.7	0.5	1.4	0.4	
Glyma.04G041500	10.4	10.2	4.7	8.3	6.5	1.5	*
Glyma.04G028300	4.8	4.4	4.2	4.5	14.0	19.0	*
Glyma.04G027600	316.0	285.7	303.3	351.8	358.4	265.4	*
Glyma.04G014800	1.1	0.2	1.5	1.2	2.9	0.8	*
Glyma.06G014800	0.6	1.1	0.3	0.5	0.5	0.0	
Glyma.04G190700	0.0	0.0	0.0	0.0	0.0	0.0	
Glyma.05G096800	1.8	1.9	0.7	1.4	0.4	3.4	*
Glyma.03G226300	167.3	179.2	159.8	172.4	204.6	149.6	*
Glyma.06G027700	64.1	55.9	66.4	66.7	89.9	70.4	*
Glyma.08G116400	0.8	0.5	0.5	0.5	0.2	0.0	
Glyma.08G116900	36.5	4.3	3.8	29.8	1.2	0.8	*
Glyma.09G069800	82.8	62.9	142.9	133.1	193.7	144.8	*
Glyma.11G113500	7.2	10.3	11.2	8.1	9.1	11.9	*
Glyma.12G039400	4.2	4.8	4.2	3.4	14.0	23.1	*
Glyma.12G208200	4.2	4.8	4.2	1.8	0.8	1.6	*
Glyma.14G085800	121.8	117.0	122.5	134.1	415.9	685.7	*
Glyma.14G216300	82.6	84.9	248.1	168.5	296.9	551.2	*
Glyma.15G177800	435.9	379.2	482.7	598.6	514.9	271.7	*
Glyma.17G049000	42.6	45.6	41.3	43.8	43.7	29.1	*
Glyma.17G126300	0.4	0.6	0.7	0.5	1.6	0.8	*

Glyma.17G168300	0.7	1.2	0.8	1.0	0.6	1.9	*
Glyma.17G239000	139.2	134.2	265.3	204.2	456.3	588.6	*
Glyma.17G254900	143.5	135.9	327.9	250.6	411.7	690.4	*
Glyma.19G223300	47.1	46.8	84.8	67.4	117.9	66.2	*
Glyma.06G174800	0.0	0.0	0.2	0.0	2.8	0.3	*
Glyma.06G283100	0.0	0.0	0.0	0.0	2.7	0.6	*
Glyma.10G207100	2.5	2.3	1.1	1.5	7.1	35.8	*

Table B.5: C13-cysteine proteases identified in soybean root nodules through RNA-Seq with a sequence similar to caspase.

* Indicates C13-Proteases seen as transcriptionally active sequence.

	FPKM						
	DNS			DS			
C13	DNS (60 %)	DNS (40 %)	DNS (30 %)	60 %	40 %	30 %	
Glyma.14G092800	8.2	6.0	39.9	24.9	45.6	78.0	*
Glyma.06G217200	7.2	8.1	10.3	9.0	11.5	12.1	*
Glyma.06G050700	79.7	69.5	94.7	121.0	138.0	141.3	*
Glyma.04G049900	70.1	61.8	106.4	100.3	141.8	172.9	*
Glyma.17G230700	28.1	17.4	226.9	97.1	309.1	367.8	*
Glyma.17G137800	51.4	56.4	307.0	154.6	185.6	351.3	*
Glyma.05G055700	8.2	11.9	41.5	28.1	63.3	221.4	*
Glyma.04G156000	6.8	7.7	7.7	7.3	9.2	6.5	*
Glyma.16G066600	0	0	0	0	0	0	

Table B.6: Expression of different protease families of same age DNon-stressed and DS nodules.

Protease	DNS (60 %)	DS 60 %	DNS (40 %)	DS 40 %	DNS (30 %)	DS 30 %	Soybean transcriptome
Aspartic proteases	7.23%	6.48%	7.21%	7.24%	6.59%	7.14%	6.91%
Cysteine proteases	**17.73%**	**18.21%**	**17.40%**	**18.10%**	**17.50%**	**18.75%**	**22.12%**
Metallo-proteases	37.69%	38.53%	39.37%	38.28%	38.99%	39.46%	31.23%
Serine proteases	29.60%	28.90%	28.47%	28.62%	29.29%	26.96%	34.63%
Threonineproteases	7.75%	7.88%	7.56%	7.76%	7.63%	7.68%	5.11%

DNS=DNon-stressed nodules of same age than DS nodules
Expression determined as FPKM (transcript abundances in fragments per kilo base of exon per million fragments mapped)
Soybean genome: Proteases (%) of entire soybean genome (14 tissue types) under non-stressed conditions (Severin *et al*., 2010)

Table B.7: Cysteine proteases expressed in DNon-stressed and drought stressed crown nodules.

Cysteine protease	DNS (60 %)	DS 60 %	DNS (40 %)	DS 40 %	DNS (30 %)	DS 30 %	Soybean transcriptome
C1 - papain-like	17.48%	17.31%	18.18%	17.14%	**16.83%**	**19.05%**	**46.15%**
C2 - calpain-like	1.94%	1.92%	2.02%	1.90%	1.98%	1.90%	0.90%
C12 - ubiquitinyl hydrolase-L1	4.85%	4.81%	4.04%	3.81%	3.96%	3.81%	2.26%
C13 – legumain-like	7.77%	7.69%	8.08%	7.62%	**7.92%**	**7.62%**	**6.79%**
C15 - pyroglutamyl peptidase	4.85%	4.81%	4.04%	3.81%	4.95%	3.81%	2.26%
C19 - ubiquitin-specific peptidase 14	1.94%	1.92%	2.02%	1.90%	1.98%	1.90%	0.90%
C26 - gamma-glutamyl hydrolase	1.94%	1.92%	1.01%	0.95%	1.98%	0.95%	0.90%
C44 - phosphoribosyl transferase	4.85%	4.81%	5.05%	6.67%	4.95%	6.67%	3.17%
C48 - Ulp1 peptidase (SUMO-like)	9.71%	9.62%	10.10%	11.43%	11.88%	10.48	7.24%
C50 - separase	0.00%	0.00%	0.00%	0.95%	0.00%	0.95%	0.90%
C54 - autophagin-1	1.94%	1.92%	2.02%	1.90%	1.98%	1.90%	0.90%
C56 - PfpI peptidase	3.88%	4.81%	4.04%	4.76%	4.95%	4.76%	3.17%
C65 - otubain-1	1.94%	1.92%	2.02%	1.90%	1.98%	1.90%	0.90%
C78 - UfSP1 peptidase	1.94%	1.92%	2.02%	1.90%	1.98%	2.86%	2.26%
C83 -Gamma-glutamylcysteine dipeptidyltranspeptidase	1.94%	1.92%	2.02%	1.90%	1.98%	1.90%	1.36%
C85 - OTLD1 deubiquitinylating enzyme	17.48%	16.35%	16.16%	15.24%	14.85%	15.24%	7.69%
C86 - ataxin	3.88%	3.85%	4.04%	3.81%	3.96%	3.81%	1.81%
C97 - DeSI-1 peptidase	11.65%	12.50%	13.13%	12.38%	11.88%	10.48%	10.41%

DNS=DNon-stressed nodules of same age than DS nodules.

Expression determined as FPKM (transcript abundances in fragments per kilo base of exon per million fragments mapped).

Soybean genome: Proteases (%) of entire soybean genome (14 tissue types) under DNon-stressed conditions (Severin *et al*., 2010).

Table B.8: Cystatins identified in soybean root nodules through RNA sequencing with a sequence similar to oryzacytatin-I.

	FPKM						
	DNS			DS			
Cystatins	DNS (60 %)	DNS (40 %)	DNS (30 %)	60 %	40 %	30 %	
Glyma.05G149800	59.8	55.1	45.5	53.8	132.7	358.0	*
Glyma.07G266000	2.3	4.5	2.8	2.6	2.8	3.4	*
Glyma.13G071800	89.6	99.1	39.9	67.6	82.2	59.6	*
Glyma.13G189500	50.2	49.9	42.8	44.2	115.1	216.9	*
Glyma.14G038200	7.4	8.7	15.6	18.2	24.6	19.3	*
Glyma.14G038300	0.1	0.2	0.3	1.2	0.0	0.0	
Glyma.14G038500	0.1	0.2	0.0	1.1	0.0	0.0	
Glyma.15G115300	136.2	168.1	331.6	200.2	135.7	143.0	*
Glyma.15G227500	45.5	52.6	52.2	52.7	82.8	115.7	*
Glyma.18G003700	0.0	0.0	0.0	0.0	10.1	8.4	*
Glyma.18G103700	0.0	0.3	0.4	0.1	3.8	0.2	*
Glyma.20G045500	118.4	97.7	98.8	83.5	98.2	123.3	*

* Indicates cystatins seen as transcriptionally active sequence.

Table B.9: Expression of different protease inhibitors families in DNon-stressed and drought stressed nodules.

Protease Inhibitors	DNS (60 %)	DS 60 %	DNS (40 %)	DS 40 %	DNS (30 %)	DS 30 %	Soybean transcriptome
I3 - Kunitz trypsin inhibitors	50%	46.43%	45%	46.43%	48.15%	45.16%	30%
I4 - Serpin serine protease inhibitors	0%	0%	0%	0%	0%	0%	2.22%
I12 - Bowman-Birk serine inhibitors	8.33%	7.14%	8.33%	7.14%	7.41%	6.45%	12.22%
I13 - Glutamyl peptidase inhibitor	4.17%	10.71%	4.17%	7.14%	3.70%	6.45%	8.89%
I25 - Phytocystatins	29.17%	25%	33.33%	28.57%	25.93%	29.03%	22.22%
I51 - Phosphatidylethanolamine-binding protein	4.17%	7.14%	4.17%	7.14%	11.11%	9.68%	23.33%
I75 - Putative serine esterase inhibitor	4.17%	3.57%	4.17%	3.57%	3.70%	3.23%	1.11%

DNS=DNon-stressed nodules of same age than DS nodules.
Expression determined as FPKM (transcript abundances in fragments per kilo base of exon per million fragments mapped).
Soybean genome: Proteases (%) of entire soybean genome (14 tissue types) under DNnon-stressed conditions (Severin *et al*., 2010).

Table B.10: Primer sets used to amplify target genes.

Description	Phytozome ID		5'/3'
C1	*Glyma.14G085800*	F R	TCTTCGCGGTTTCATCGGC RTACTAGCCACTGCTCGTAC
	Glyma.04G028300	F R	GAGGACAACTATGCCTTTGTCG CGAGACGAGTGGTTTTGAA
	Glyma.10G207100	F R	AGCTGTCACCGAAGTTAAGT GGCTCCTTCTATAGATCCTGTG
	Glyma.06G174800	F R	ACCAACCTGTTTCTGTAGCC CCAGTAACTAGTCCCATCAACAG
	Glyma.12G039400	F R	GCGCTTCTCCGATCTTACGC GTTCGCGTCCAAAGGCAACC
	Glyma.06G283100	F R	ACCATCCACAGTGGATTGGAG CTTCAGTTGCTGCAACAGCG
	Glyma.17G230700	F R	TTGATGATTGGCACTGCCTT CCTCATATGTTTCATGCCATACT
C13	*Glyma.14G092800*	F R	TTGATGACTGGCACTGCCTG CCTCATGTGTTTCATCCCGTAT
	Glyma.05G055700	F R	CTGAGGATGATGATGGCGCAG AGCATCGACCGGTTCAGC
	Glyma.05G149800	F R	TGATAAGGTCACCGGCGGTG TCAGGGATCTTCGACCTTCCC
Cystatins	*Glyma.13G189500*	F R	TAGGGCACAGGAACAGGTT CTTCTCACCAGCCTCAATAGCC
	Glyma.18G103700	F R	CATTTTGCTGTTGATGAGCAT TCCTTGGCCTCTAGAGTGATAT
	Glyma.18G003700	F R	CGATTCTGGTGACCCTTCTCT CTCACTTCAGGGATCTCCGT
40s	*Glyma.08G325000*	F R	GCATTATGGCGTTGAGGTTG CGGTTCTGCTTTCGCTTTTC
ELF	*Glyma.02G276600*	F R	GTGGTACGATGCTGTCTCTTC CCACTGAATCTTACCCCTTGAG

APPENDIX C

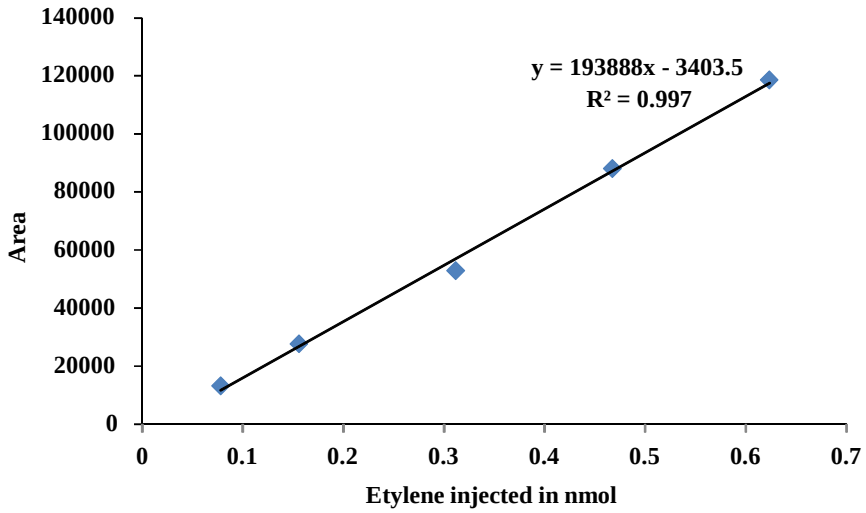


Fig. C.1 Standard curve area of the chromatogram of 0.1-.0.7 nmol of ethylene injected, used for the calculation of the nitrogenase activity of crown nodules.

APPENDIX D

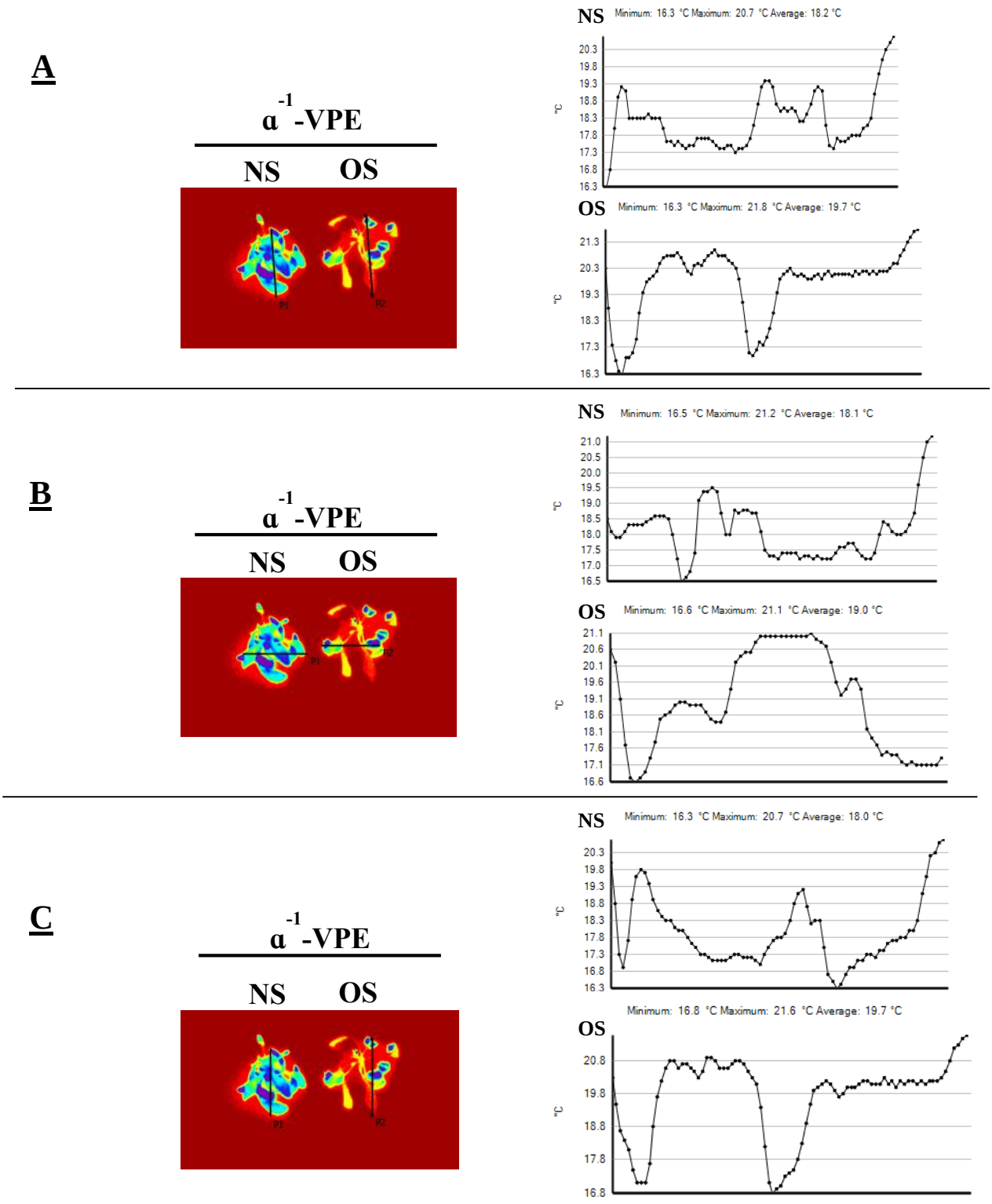


Fig D.2: Overall temperature measured A) random B) horizontal and C) vertical across NS and OS rosettes with corresponding temperature graphs.

Table D.1: Temperature averages over *Arabidopsis* mutants rosettes.

	NS	OS
WT		
Random	17.3 ºC	18.5 ºC
Horizontal	17.8 ºC	18.4 ºC
Vertical	17.6 ºC	18.1 ºC
α^{-1}-VPE		
Random	18.2 ºC	19.7 ºC
Horizontal	18.1 ºC	19.0 ºC
Vertical	18.0 ºC	19.7 ºC
β^{-1}-VPE		
Random	18.1 ºC	19.1 ºC
Horizontal	18.1 ºC	19.8 ºC
Vertical	18.4 ºC	19.7 ºC
δ^{-1}-VPE		
Random	18.1 ºC	19.2 ºC
Horizontal	18.7 ºC	19.5 ºC
Vertical	17.9 ºC	19.1 ºC
γ^{-1}-VPE		
Random	16.9 ºC	18.8 ºC
Horizontal	16.4 ºC	19.2 ºC
Vertical	17.0 ºC	19.6 ºC
null^{-1}-VPE		
Random	18.0 ºC	19.0 ºC
Horizontal	17.9 ºC	19.8 ºC
Vertical	18.1 ºC	19.2 ºC

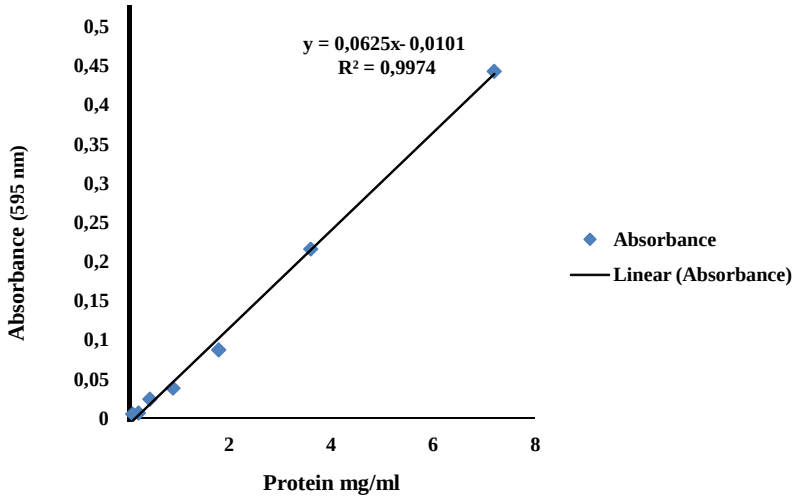


Fig D.2: Standard curve of the absorbance (595 nm) of 0-8 µg of protein standard used for the calculation of protein content.

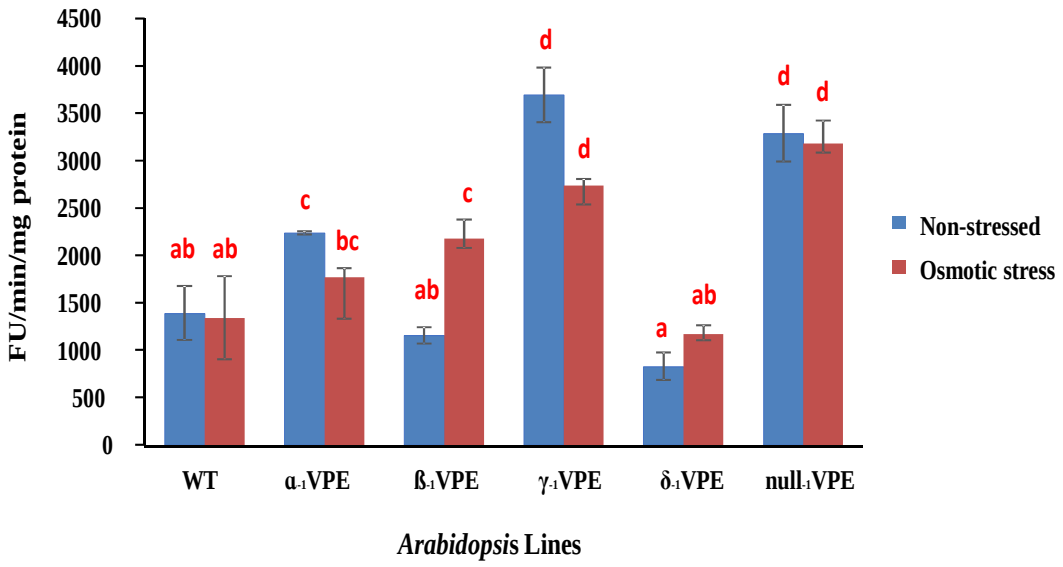


Fig. D.3: Raw data of Cathepsin-L like activity of *Arabidopsis* mutant lines.

[a-z] Significant differences between NS and OS samples.

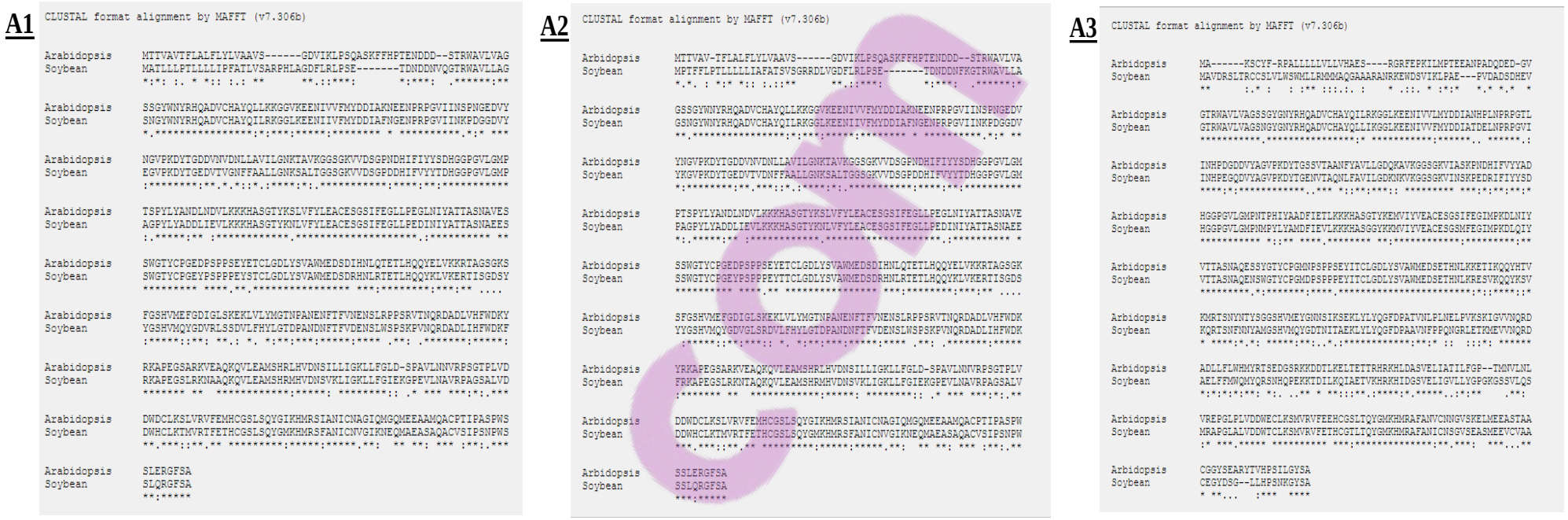


Fig. D4: Sequence alignments of A1) *Glyma.17G230700* and A2) *Glyma.14G092800* to the α-VPE as identified in the *Arabidopsis* genome. A3) *Glyma.05G055700* aligned to β-VPE.