Identification and functional evaluation of a drought-induced “late embryogenesis abundant” gene from cowpea plants

by

Inge Gazendam
97011909

Submitted in partial fulfilment of the requirements for the degree
PhD Biotechnology

In the Faculty of Natural and Agricultural Sciences
Department of Plant Science
University of Pretoria
Pretoria

February 2012

Under the supervision of Prof. DK Berger and
co-supervision of Dr. D Oelofse

© University of Pretoria
# Table of Contents

<table>
<thead>
<tr>
<th>Declaration</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ix</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>x</td>
</tr>
<tr>
<td>Index of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>Index of Tables</td>
<td>xiv</td>
</tr>
</tbody>
</table>

## CHAPTER 1

### Literature review

Cowpea, plant responses to abiotic stresses and the involvement of late embryogenesis abundant (LEA) proteins

1.1 Cowpea

1.1.1 Characteristics and uses of the legume cowpea

1.1.2 Cowpea as a source of drought tolerance genes

1.1.3 Problems experienced when growing cowpea

1.1.4 Drought tolerance mechanisms of cowpea

1.1.5 Abscisic acid is an important role player in the drought tolerance of cowpea

1.1.6 Drought screening methods for breeding drought tolerant cowpea cultivars

1.1.7 Studies to identify cowpea drought tolerance genes

1.1.8 Genomic resources of cowpea

1.1.9 Identification of differentially expressed cowpea transcripts

1.2 Plant responses to abiotic stress

1.2.1 Morphological and physiological responses of plants to drought

1.2.2 Molecular responses of plants to abiotic stress

1.2.3 Salt and drought stress signaling

1.2.4 Drought stress responses

1.2.5 Salt stress responses

1.2.6 Cold stress responses

1.2.7 The involvement of ABA in abiotic stress

1.2.8 Stress- and ABA-regulated gene expression

1.2.9 Transcription factors and DNA recognition sequences
1.3 Late embryogenesis abundant (LEA) proteins
  1.3.1 Characteristics of LEA proteins
  1.3.2 Classification of LEA proteins
  1.3.3 Atypical LEA proteins
  1.3.4 Three-dimensional structure of LEA proteins
  1.3.5 Expression profile of LEA genes
  1.3.6 Subcellular localisation of LEA proteins
  1.3.7 LEA gene families
  1.3.8 Functional role of LEA proteins

1.4 *Arabidopsis thaliana* as a model experimental plant

<table>
<thead>
<tr>
<th>CHAPTER 2</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction of a cowpea drought expression SSH cDNA library, and screening for differentially expressed drought genes using microarray technology</strong></td>
<td></td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>35</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>2.3 Materials and Methods</td>
<td>37</td>
</tr>
<tr>
<td>2.4 Results</td>
<td>45</td>
</tr>
<tr>
<td>2.5 Discussion</td>
<td>69</td>
</tr>
<tr>
<td>2.6 Acknowledgements</td>
<td>75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 3</th>
<th>76</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expression patterns and diversity study of the cowpea LEA5 gene</strong></td>
<td></td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>76</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>76</td>
</tr>
<tr>
<td>3.3 Materials and Methods</td>
<td>77</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>81</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>100</td>
</tr>
<tr>
<td>3.6 Acknowledgements</td>
<td>107</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 4</th>
<th>109</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional evaluation of the cowpea LEA5 gene with regard to drought tolerance</strong></td>
<td></td>
</tr>
<tr>
<td>4.1 Abstract</td>
<td>109</td>
</tr>
<tr>
<td>4.2 Introduction</td>
<td>109</td>
</tr>
</tbody>
</table>
4.3 Materials and Methods
4.4 Results
4.5 Discussion
4.6 Acknowledgements

Concluding remarks
References
Summary
Appendix 1 Oligonucleotide primers used for the amplification of cowpea and A. thaliana genes
Appendix 2 SSHscreen R script, Targets file and SpotTypes file used for ER3 analysis of cowpea SSH libraries
Appendix 3 Nucleotide and amino acid sequences of a LEA5 gene identified from cowpea
Appendix 4 Plasmid maps and cloning strategy for transforming A. thaliana with a cowpea LEA5 gene
Declaration

I, Inge Gazendam, declare that the thesis, which I hereby submit for the degree PhD Biotechnology at the University of Pretoria, is my own work except where acknowledged. This work has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: ……………………………

DATE: ……………………………
Preface

Increased climate variability and change, including global warming, are serious threats to agricultural productivity and poverty eradication worldwide. South Africa is classified as a water-stressed country (Department of Water Affairs and Forestry (DWAF), 2004), experiencing increases in land demand as a result of population growth, degradation of watersheds caused by land-use change, and the siltation of river basins (Bennie and Hensley, 2001). Farming is the most important source of income and sustenance for approximately three quarters of the population of sub-Saharan Africa. On-going dry spells influence water supply by affecting rainfall and evapo-transpiration (Alpert and Oliver, 2002). Farming depends mainly on the quality of the rainy season, and rain-fed subsistence agriculture is predominant in South Africa. Drought therefore disrupts rural and national economies because of declined productivity. Most of the premium quality farmland is already used for agriculture, which means that further area expansion is bound to occur on marginal land that is unlikely to sustain high yields through conventional approaches. It is believed that agricultural biotechnology can make substantial contributions towards increasing food production by resource poor farmers by improving crop production in water-stressed areas.

Breeding for drought tolerant crop species is, therefore, highly important in South Africa. Tolerance to water stress is, however, a quantitative trait, governed by multiple genes (Bohnert et al., 1995). This tolerance manifests at different levels and different stages in the development of the plant, and involves highly complex osmo-regulatory functions. A combination of strategies is, therefore, needed to improve a crop’s drought tolerance. The discovery of candidate drought-tolerance genes is important, followed by the introduction into crops by gene transfer, or to assist in the molecular selection during breeding (marker-assisted selection). The overall aim of this study was to discover drought-responsive genes, which may be involved in a plant’s drought tolerance, and can therefore be used in future studies to improve the drought tolerance of crops.

In order to understand the gene networks that underlie plant stress tolerance, it is necessary to identify and characterise the genes that respond to the particular stress. A number of previous studies have attempted to identify stress-responsive genes and elucidate the tolerance mechanisms in plants (Van den Berg et al., 2004; Kawasaki et al., 2001; Guo et al., 2009; Collett et al., 2004; Seki et al., 2001 & 2002). Gene discovery methods may include expressed sequence tag (EST) sequencing of cloned cDNAs (Ralph et al., 2008). This is often replaced these days by next generation sequencing technologies of cDNA libraries.
Current disadvantages of this technology is its cost, it is labour intensive, and is effective only if an annotated genome sequence is available.

A recent method of creating enriched cDNA libraries, called suppression subtractive hybridisation (SSH), is a popular approach for gene discovery in non-model organisms that do not have genome sequence information (Crampton et al., 2009; Van den Berg et al., 2007). It is particularly useful for laboratories focused on a particular research question without access to resources to conduct whole transcriptome sequencing using next generation technologies. The advantages of SSH lie in the enrichment of genes that are differentially expressed between treatments, as well as the recovery of rare transcripts (Diatchenko et al., 1996; Hillmann et al., 2009). SSH was combined with microarrays, another high-throughput gene discovery technology, to improve the selection of differentially expressed genes for sequencing. A custom microarray slide was developed for a neglected legume crop, cowpea, which was subsequently screened for the identification of drought tolerance genes.

Cowpea can be grown under the harshest growing conditions, and in the poorest soils, and is, therefore, an important crop for subsistence and small-holder farmers (Quass, 1995). Due to the drought tolerance that most cowpea cultivars exhibit, cowpea was selected as a source to identify genes that are able to enhance the drought tolerance of other crops of agricultural importance. Major crops grown in South Africa and producing the highest income include maize, wheat and potato (Department of Agriculture, Forestry and Fisheries (DAFF), 2011). They are also important staple food crops for millions of people around the world. Drought is frequently implicated as the single most important factor limiting their yield. They are therefore important crops for drought tolerance improvement studies. The biotechnology company Monsanto has a “more crop per drop” mandate, and is trying to create maize lines that are higher yielding when grown in the dryer states of the United States of America. Syngenta is also using conventional breeding and transgenic approaches to create drought-tolerant maize [http://www.nytimes.com/2008/10/23/business/23drought.html].

Dr. BB Singh at the International Institute of Tropical Agriculture (IITA) has extensively researched the drought tolerance of cowpea (Dingkuhn et al., 2006). He kindly supplied the two cowpea lines, IT96D-602 and Tvu7778, the one drought tolerant and the other susceptible, respectively, to be used in this study.

The first aim of this study was, therefore, to identify genes in a tolerant cowpea line that respond to drought stress. In order to achieve this aim, SSH was used to create a cowpea
drought expression cDNA library from two cowpea lines differing in drought tolerance. Microarray studies were performed to verify the SSH library, and quantitative PCR (qPCR) performed on selected genes to verify the microarray results.

A cowpea late embryogenesis abundant protein gene (*LEA5*) has been identified in this study as one of the genes that were up-regulated in drought stressed cowpea leaves. This gene belongs to an atypical class of LEA proteins, which has not been widely studied in literature. Much information is available, however, on the involvement of typical LEA proteins in the abiotic stress protection of plants. The second aim was to characterise this *LEA5* gene's structure and expression patterns in cowpea during abiotic stress. Its similarity to other plant genes and the sequence diversity within cowpea for this gene was also studied through comparisons with available gene and protein sequences.

The third aim was to functionally evaluate the role of a LEA5 protein in drought tolerance of plants. This was achieved through generating a transgenic model plant containing the cowpea *LEA5* gene and testing whether it was able to improve its tolerance to drought. *Arabidopsis thaliana* was chosen for its short lifecycle, ease of transformation and availability of established screening protocols (Clough and Bent, 1998). It was, therefore, used as proof-of-concept to improve the drought tolerance of crops of importance to resource poor farmers.

The first hypothesis of the study was that drought stress of cowpea plants would result in differential expression of genes involved in the protection against drought stress. The non-model legume, cowpea, was chosen for the construction of a cDNA library enriched for drought-responsive genes, due to its inherent drought tolerance. The second hypothesis states that the cowpea *LEA5* protein gene plays an important role in the drought tolerance of plants.

This research therefore addresses the following research questions:
1. Can genes responding to drought be identified from cowpea using a SSH cDNA library combined with microarrays?
2. Which genes are responsive during drought stress of cowpea?
3. Which abiotic stress conditions up-regulate the expression of the cowpea *LEA5* gene?
4. What is the sequence diversity of *LEA5* genes among cowpea cultivars?
5. What role does the LEA5 protein play in a plant’s tolerance to drought stress?
6. Can cowpea LEA5 improve *A. thaliana*’s tolerance to drought?
Chapter 1 of the thesis is comprised of a review of the literature on the legume cowpea, plant mechanisms in response to abiotic stresses and LEA proteins. 

Chapter 2 describes the generation of a cowpea SSH drought cDNA library and screening using microarrays and software tools, specifically SSHscreen and SSHdb. This work has been published in Plant Methods with myself as a joint first author (Coetzer et al., 2010).

Chapter 3 characterises the chosen cowpea LEA5 gene further, through bioinformatic analysis and expression studies.

Chapter 4 covers the over-expression of the cowpea LEA5 gene in A. thaliana, the molecular characterisation of the transgenic lines and the effect this gene has on tolerance to drought.

The Concluding remarks section contains a brief discussion of the results obtained in this study and the findings of the research questions. The outcomes of the hypotheses are also addressed.
Acknowledgements

My sincere gratitude and appreciation to:

My supervisor Prof. DK Berger for guidance in this project

My co-supervisor Dr. D Oelofse for support of the project

All my colleagues at the ARC-Roodeplaat Vegetable and Ornamental Plant institute

Funding from the ARC through its Parliamentary grant

Plant Science Department and FABI at the University of Pretoria

Past and present members of the Molecular Plant-Microbe Interactions laboratory

Dr. Bridget Crampton for guidance in SSH and microarray experiments

Mr. Nicky Olivier for assistance with microarray experiments and data analysis

Ms. Nanette Coetzer for SSHscreen, SSHdb and co-authoring of a scientific publication

Ms. Adri Veale and Minique de Castro for assistance with Arabidopsis experiments

My family and husband, Albert, for their emotional support
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABA</td>
<td>Abscisic acid</td>
</tr>
<tr>
<td>bp</td>
<td>Basepair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C\textsubscript{q}</td>
<td>Quantification cycle</td>
</tr>
<tr>
<td>DIG</td>
<td>Digoxigenin</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>EF1A</td>
<td>Elongation factor 1-alpha</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tags</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobasepair</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>LEA</td>
<td>Late embryogenesis abundant</td>
</tr>
<tr>
<td>MDA</td>
<td>Malondialdehyde</td>
</tr>
<tr>
<td>MPa</td>
<td>Megapascal</td>
</tr>
<tr>
<td>MS</td>
<td>Murashige and Skoog</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PVP</td>
<td>Polyvinyl pyrrolidone</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative PCR</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse transcriptase-quantitative PCR</td>
</tr>
<tr>
<td>RWC</td>
<td>Relative water content</td>
</tr>
<tr>
<td>SSH</td>
<td>Suppression subtractive hybridisation</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris hydroxy methyl aminoethane</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-galactoside</td>
</tr>
</tbody>
</table>
Index of Figures

| Figure 1.1 | Cowpea plants (A) and seeds (B) | Page 2 |
| Figure 1.2 | Array of the distinctive amino acid motifs in the LEA protein groups | Page 22 |
| Figure 1.3 | Unrooted dendrogram of all Arabidopsis LEA genes | Page 26 |
| Figure 2.1 | Formaldehyde agarose (FA) gel electrophoresis of cowpea RNA isolated using Tri-reagent. | Page 46 |
| Figure 2.2 | RT-PCR and PCR of the actin gene using DNase treated and untreated cowpea Tvu7778 RNA as templates. | Page 47 |
| Figure 2.3 | Secondary PCR amplification of subtracted and unsubtracted cDNA fragments produced from drought stressed IT96D-602 and control Tvu7778 cowpea lines that had been subtracted with suppression subtractive hybridisation. | Page 48 |
| Figure 2.4 | Example of Microarray pseudocolour images following hybridisation. | Page 49 |
| Figure 2.5 | Box plots of M-values of control spots before and after normalisation for all 12 slides used for SSHscreen analysis. | Page 51 |
| Figure 2.6 | MA plots after normalisation of the forward and reverse cowpea SSH libraries. | Page 52 |
| Figure 2.7 | ER3 versus inverse ER2 plot produced by SSHscreen for the cowpea forward (A) and reverse (B) libraries. | Page 54 |
| Figure 2.8 | ER3 versus inverse ER2 plot for sequenced clones to illustrate that redundant partners cluster together. | Page 66 |
| Figure 2.9 | Regulation (A) and relative abundance (B) of selected cowpea genes (qPCR verification). | Page 68 |
| Figure 3.1 | Nucleotide sequence of SSH library clone 07F09-F of the cowpea drought expression library, and the deduced amino acid sequence of VuLEA5. | Page 81 |
| Figure 3.2 | Sequence alignment of the cowpea genomic LEA5 gene of two lines, IT96D-602 and Tvu7778. | Page 83 |
Figure 3.3 Significant match of the VuLEA5-1 protein sequence to the LEA_3 Pfam family.
Figure 3.4 Hydropathicity plot of the amino acid sequence of the VuLEA5-1 protein.
Figure 3.5 Predicted secondary structure of the VuLEA5-1 protein.
Figure 3.6 Amino acid sequence alignment of the cowpea VuLEA5-1 protein and its closest orthologs in soybean, peanut, *Populus*, citrus, castor bean, cotton and *Arabidopsis*.
Figure 3.7 Amino acid sequence alignment of the cowpea VuLEA5-1 protein and its four LEA5 protein orthologs in *A. thaliana*.
Figure 3.8 Phylogram of group 5 LEA proteins including VuLEA5-1 from *Vigna unguiculata* (IT96D-602).
Figure 3.9 Phylogenetic tree of cowpea ESTs containing the *VuLEA5* gene, based on the coding nucleotide sequences.
Figure 3.10 Alignment of translated amino acid sequences of representative *VuLEA5* EST nucleotide sequences from different lines of cowpea.
Figure 3.11 Relative water content (RWC) of cowpea leaves of two lines during a drought trial.
Figure 3.12 DNase treatment removes contaminating gDNA from RNA isolated from cowpea using the Plant RNeasy kit (Qiagen).
Figure 3.13 Gene expression of *VuLEA5* in control and drought stressed samples of two cowpea lines.
Figure 3.14 Cowpea LEA5 (*VuLEA5*) and *NCED1* expression in IT96D-602 leaves after desiccation and rehydration.
Figure 3.15 Cowpea LEA5 (*VuLEA5*) expression in IT96D-602 leaves after 5 h and 24 h treatment with ABA, NaCl or cold.
Figure 3.16 Cowpea *NCED1* expression in IT96D-602 leaves after 5 h and 24 h treatment with ABA, NaCl or cold.
Figure 4.1 PCR colony screen of *E. coli* JM109 transformants of pMDC32-*VuLEA5-1*.
Figure 4.2 PCR screening of T1 *A. thaliana* lines transformed with the *VuLEA5-1* gene.
Figure 4.3 Difference in shoot lengths of *A. thaliana* seedlings selected for hygromycin resistance.
Figure 4.4 PCR amplification of the *VuLEA5-1* transgene and Cap-binding protein 20 gene in T4 generation transgenic and untransformed *A.
Figure 4.5  Restriction enzyme digested genomic DNA of *VuLEA5-1* transgenic *A. thaliana* lines after overnight electrophoresis on an agarose gel.

Figure 4.6  Genomic Southern blot analysis of the *VuLEA5-1* gene in transgenic *A. thaliana* lines.

Figure 4.7  Transcript accumulation patterns of the *VuLEA5-1* transgene in transgenic *A. thaliana* lines.

Figure 4.8  Silver stained gel of heat-soluble proteins prepared from untransformed *A. thaliana* and *VuLEA5-1* transgenic lines.

Figure 4.9  Western-blot analysis of *VuLEA5-1* in untransformed *A. thaliana* and *VuLEA5-1* transgenic lines.

Figure 4.10  Transgenic cowpea LEA5 (VuLEA5-1) expression in T4 homozygous transgenic *A. thaliana* lines.

Figure 4.11  Drought stress and recovery of *A. thaliana* plants transformed with the *VuLEA5-1* gene (line 12).

Figure 4.12  Root elongation of *A. thaliana* seedlings after one week on control (MS) and different osmotic stress *in vitro* mediums.

Figure 4.13  Comparison of osmotic stress response of *VuLEA5-1* transgenic *A. thaliana* lines and wild type Col-0, as induced by PEG-8000 infused plates with a $\psi_W$ of -0.5 MPa or -1.2 MPa.

Figure 4.14  Comparison of osmotic stress response of *VuLEA5-1* transgenic *A. thaliana* lines and wild type Col-0, as induced by PEG-8000 infused plates with a $\psi_W$ of -0.7 MPa.

Figure 4.15  Comparison of osmotic stress response of *VuLEA5-1* transgenic *A. thaliana* lines and wild type Col-0, as induced by 200 mM mannitol or PEG-8000 infused plates with a $\psi_W$ of -0.5 MPa.

Figure 4.16  Hypocotyl elongation of *VuLEA5-1* transgenic *A. thaliana* lines and wild type Col-0 on osmotic stress medium induced by PEG-8000 infused plates with a $\psi_W$ of -0.5 MPa.

Figure 4.17  Fresh weight of *A. thaliana* seedlings after 14 days osmotic stress.

Figure 4.18  Fresh weight of *A. thaliana* seedlings after 13 days recovery from osmotic stress imposed by PEG-8000 (-0.7 MPa) for 11 days.

Figure 4.19  Amino acid sequence alignment of MtPM25 [ABB16353] and *VuLEA5-1*.

Figure 4.20  Amino acid sequence alignment of MtPM25 [ABB16353] and Atrab28 [At3g22490; NP_188888].
# Index of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Top tables produced by SSHscreen for the forward and reverse cowpea libraries.</td>
<td>56</td>
</tr>
<tr>
<td>2.2</td>
<td>Cowpea drought responsive genes annotated in SSHdb (after sequencing of selected clones).</td>
<td>59</td>
</tr>
<tr>
<td>3.1</td>
<td>Amplification efficiency of primer pairs.</td>
<td>97</td>
</tr>
<tr>
<td>3.2</td>
<td>Putative functions of LEA5 proteins closely related to VuLEA5.</td>
<td>108</td>
</tr>
<tr>
<td>4.1</td>
<td>Number of surviving <em>A. thaliana</em> plants after rewatering 10-day drought-stressed plants.</td>
<td>131</td>
</tr>
</tbody>
</table>
Chapter 1

Literature review
Cowpea, plant responses to abiotic stresses and the involvement of late embryogenesis abundant (LEA) proteins

1.1 Cowpea

1.1.1 Characteristics and uses of the legume cowpea
Cowpea (Vigna unguiculata (L.) Walp) is an important legume crop in the semi-arid tropics, including South Africa (Figure 1.1 A). It is of African origin, and domestication occurred in the savanna regions of Africa north of the Equator. Cowpeas are planted on 12-14 million hectares across the world. The world cowpea production is estimated to be over 4.5 million tons dry cowpea grain annually (Singh et al., 2002), with 70% produced on farms in the dry savannah of tropical west and central Africa. In this region it is an important food grain for over 200 million people. Nigeria is the largest producer and consumer of cowpea. Cowpea is also an important crop in marginal areas of eastern and southern Africa, such as Sudan, Somalia, Mozambique and southern Zimbabwe. Other production areas in the world include low elevation areas in South America, parts of India, and southern regions of North America (Timko and Singh, 2008).

Cross-hybridisation occurs frequently between cultivated and wild cowpea, resulting in a continuous variation, especially in west Africa. Cowpea has a diverse plant architecture. Determinate varieties grow upright and are used for grain production, where it facilitates easier grain harvesting. Spreading types are used for both their grain and fodder, and are frequently inter-planted with cereals.

Cowpea seeds (Figure 1.1 B) vary in size, colour and uniformity (being a solid colour or speckled). The germinated plant forms its first two true leaves opposite each other directly on the stem, with the remaining leaves alternating, having petioles and being trifoliate. The optimal growing temperature is 28°C, with all stages of development requiring at least 18°C. Flowers are carried on long peduncles, are primarily self-pollinating and forms two or three pods containing the seeds (Timko and Singh, 2008).
Cowpea is a multifunctional crop, since it provides food and forage to humans and animals, while providing a sustainable income for farmers and grain traders (Timko and Singh, 2008; Singh et al., 2002). Both the seeds and young leaves of cowpeas contain more than 25% protein (on a dry weight basis). Cowpeas are therefore important in the diets of humans, where it is used as a protein substitute for meat products (Singh et al., 2003). The grain is consumed after boiling, and combined with porridge made from cereals and root crops. Cowpea grain cakes are prepared in Nigeria from mashed and fried seed. In eastern and southern Africa, cowpea leaves are added to sauces, served with porridge, or boiled similar to spinach. The stems and foliage are also used as fodder for livestock. The green haulms are harvested, rolled into bundles and stored on rooftops to dry. This serves as a feed supplement in the dry season, but is also a source of income, since the dry haulms fetch 50% or more of the grain price (on a dry weight basis).

1.1.2 Cowpea as a source of drought tolerance genes

Since many lines are drought tolerant, cowpea can be grown under the harshest growing conditions and in the poorest soils. Some cowpea varieties are grown close to the Sahara desert and can tolerate drought and heat (Timko and Singh, 2008). Others can tolerate acid soils, poor soil fertility and shading by other crops. It is therefore popular to subsistence and small-holder farmers (Quass, 1995). Cowpea was therefore selected as a source of drought tolerance genes, with the long term goal of improving other crops of agricultural importance.

Resource poor farmers are typically faced with a dryer climate with no or limited access to irrigation. Drought tolerant cowpeas have, therefore, been generated through mutation breeding to be better adapted to local conditions. Dr. BB Singh at the International Institute of Tropical Agriculture (IITA) has extensively researched the drought tolerance of cowpea
Cowpea is used especially as an intercrop for cereals in west Africa which has low and erratic rainfall, and soils which are sandy and of low fertility. Despite these adverse conditions, about 9 million hectares of the world total of 14 million hectares occurs in west Africa (Dingkuhn et al., 2006). Cowpea also improves the soil fertility by fixing nitrogen.

1.1.3 Problems experienced when growing cowpea
Cowpea growth and yield are constrained by a variety of biotic and abiotic factors. Cowpea typically yields between 100 and 400 kg/ha, without inputs and under intercropping conditions, which is five times lower than its biological potential (Dingkuhn et al., 2006). Causes include shading by tall intercrop species, such as sorghum, a prostrate architecture that is weed-competitive but low-yielding, and susceptibility to pests. The most common pests are the *Maruca* pod borers, the root parasitic weed *Striga* and nematodes. Improving cowpea yields in Africa is especially dependent on better insect resistance. Even though cowpea is highly drought tolerant, water availability is the most significant abiotic constraint to growth and yield (Timko and Singh, 2008).

1.1.4 Drought tolerance mechanisms of cowpea
Cowpea exhibits two types of drought tolerance, frequently called type 1 and type 2 drought tolerance (Singh and Matsui, 2002; Agbicodo et al., 2009).

Type 1 is called drought avoidance, where plants are able to maintain high tissue water potential even at low soil-moisture levels. Turgor is maintained by maximising water uptake, through e.g. deeper root development (De Ronde and Spreeth, 2007), and reducing water loss, by closing of stomata, reducing epidermal conductance, radiation absorption and evapo-transpiration surface. Cowpea responds to drought by reducing its leaf assimilation rate, transpiration rate and stomatal conductance. This is associated with stomatal closure and the reduction of leaf water potential (Anyia et al., 2004). Another drought avoidance mechanism of cowpea is to orient its leaflets parallel to the sun’s rays, in order to be cooler and transpire less. Plants exhibiting type 1 drought avoidance stop growing and all plant parts wilt and eventually die after prolonged drought stress.

Type 2 drought tolerant plants can withstand water deficit at low tissue water potential. It maintains turgor by osmotic adjustment. Type 2 drought tolerant lines remain green for longer during drought stress and there is a continued slow growth of the trifoliate leaves. The unifoliate leaves may senesce and drop off, but the growing tips remain turgid and survive much longer. Type 2 is a more effective mechanism of drought tolerance, since it keeps plants alive for longer, and gives them a better chance of recovery after the drought ends.
These properties make the type 2 drought tolerance response the most attractive source of genes to confer tolerance to other crops. Incorporating genes leading to drought avoidance will be detrimental to crop yield, since the plants halt their growth and photosynthesis is reduced due to closed stomata. True tolerance mechanisms protect growing cells from water stress injury, through the production of protective solutes and proteins, and detoxification of reactive oxygen species (ROS).

Inheritance of both type 1 and type 2 drought tolerance in cowpea are shown to be dominant using crossings, with a single gene responsible in each, but independent from each other (Singh and Matsui, 2002). The contribution of the roots is excluded in their experiment using wooden-box screening. Field conditions induce various different mechanisms of drought tolerance and made it appear to be a complex trait. The wooden-box screening for dehydration tolerance in shoots is primarily due to stomatal behaviour or osmotic adjustment, which their study found to be under a single major gene’s control.

A third response employed by cowpea, termed drought escape, is where cultivars are early maturing, but they have very poor performance when exposed to intermittent drought during the vegetative or reproductive stages (Hall et al., 2002; Agbicodo et al., 2009). A delayed-leaf-senescence trait of cowpea allowed an early flowering cultivar to recover after an early drought, by producing a second flush of pods (Hall et al., 2002).

1.1.5 Abscisic acid is an important role player in the drought tolerance of cowpea

The plant hormone abscisic acid (ABA) is suggested to be important in the drought tolerance of cowpea (Iuchi et al., 1996). The ABA content in cowpea is increased 160-times by dehydration. The level of accumulation is also much higher compared to A. thaliana, which is about 2% of dehydrated cowpea. ABA pre-treatment is shown to protect membranes from osmotic stress when leaf disks of cowpea are subjected to dehydration conditions by polyethylene glycol (PEG) with a molecular weight of 8000 g/mol, an osmotic stress agent (Campos and Pham-Thi, 1997). The mechanism of action is through a hardening effect of membranes, by preventing the dehydration-induced membrane lipid degradative processes. ABA can affect cellular events through both transcriptional and translational processes. It can also induce antioxidant enzymes that protect membranes against oxidative stress.

1.1.6 Drought screening methods for breeding drought tolerant cowpea cultivars

Because drought tolerance is a quantitatively inherited trait, an integrated screening approach is the most promising for phenotyping cowpea for drought tolerance. There are two
approaches for screening and breeding for drought tolerance in plants (Agbicodo et al., 2009):

1. **Empirical/ performance approach**: Yield summarises the expression of all the traits affecting productivity under stress. The disadvantages of this method are that it is slow, needs several locations and years of yield trials, and there is substantial variation caused by the environment.

2. **Analytical/ physiological approach**: A specific physiological or morphological trait is identified that will contribute to yield and growth under drought stress.

A list of the most suitable methods for screening large numbers of cowpea lines for drought tolerance has been proposed (Agbicodo et al., 2009). It includes the measuring of chlorophyll fluorescence, stomatal conductance, ABA, free proline, wooden box screening at seedling stage (Singh and Matsui, 2002; De Ronde and Spreeth, 2007), delayed leaf senescence (Hall et al., 2002), and relative water content (RWC) in some cases. The wooden box screening method is used to visually identify shoot drought tolerance, by eliminating the effect of roots. Ranking cultivars on their drought tolerance using this method are found to give the same results as during field and pot screening experiments. Screening cowpea at the seedling stage therefore appears to be a reliable method for identifying drought-tolerant varieties. The root-box pin-board method is used to study the two-dimensional root architecture of individual plants, with deep and dense roots characteristic of drought-tolerant lines. Combining these two methods allow the breeding and selection of highly drought-tolerant plants. Several other methodologies to estimate drought tolerance exist, but they are expensive, time consuming and therefore unpractical for screening large segregating populations (Singh and Matsui, 2002).

### 1.1.7 Studies to identify cowpea drought tolerance genes

Numerous studies have been performed to identify and characterise drought genes in cowpea and other crops. A research group in Paris, under guidance of Thu Pham-Thi and Zuily-Fodil, has characterised a number of cowpea genes responsive during water stress. Some of them are involved in fatty acid metabolism (phospholipases D and C, phosphatidate phosphatases (PAPs) and galactolipid acyl hydrolase) (El Maarouf et al., 1999, 2001; Marcel et al., 2000; Matos et al., 2001). This is to be expected, since drought causes the degradation of membrane lipids in cowpea (Campos and Pham-Thi, 1997; Agbicodo et al., 2009). Galactosylglycerolipids are found to accumulate in membranes, and the expression of their synthase enzymes induced, during drought. They are thought to be important for plant tolerance to water deficit and recovery after rehydration (Torres-Franklin et al., 2007). Other identified genes are involved during redox processes (glutathione reductase, ascorbate
peroxidase) (Contour-Ansel et al., 2006 and D'Arcy-Lameta et al., 2006). Water deficit is known to cause the generation of ROS that causes cell damage, and these two are key detoxification enzymes. Genes have also been identified that are involved in protein metabolism (aspartic protease, cysteine protease and multicystatin (VuC1)) (Cruz de Carvalho et al., 2001; Khanna-Chopra et al., 1999; Diop et al., 2004).

Various other novel drought-inducible genes have been identified in cowpea by Iuchi et al. (1996), termed CPRD (cowpea clones responsive to dehydration). The application of exogenous ABA to a drought tolerant cowpea induces the drought-inducible genes. One of them, CPRD22, codes for a group 2 LEA protein, also called a dehydrin. A 35 kDa dehydrin protein with an almost identical N-terminal amino acid sequence has been purified from dry cowpea seeds (Ismail et al., 1999a). It is involved in the chilling tolerance of cowpea seedlings (Ismail et al., 1999b). Iuchi et al. (1996) also describes two signal transduction pathways between water stress and CPRD genes. The ABA biosynthesis enzyme 9-cis-epoxy carotenoid dioxygenase (NCED1) has also been cloned from cowpea (Iuchi et al., 2000).

This study describes the identification of cowpea drought responsive genes by applying SSH and microarrays. Preliminary sequencing results have been published and included the stress-related genes glutathione-S-transferase (GST) and pathogenesis related protein-1 (PR1) (Gazendam and Oelofse, 2007).

1.1.8 Genomic resources of cowpea

In recent years, there has been extensive interest in the genomics of cowpea. With a genome size of 620 Mb (Arumuganathan and Earle, 1991), it is small among the legumes, and because it is so drought tolerant it is an attractive source of drought-tolerant genes. The Vigna genus has a diploid genome with 22 chromosomes, and is primarily selfing [http://www.rbgkew.org.uk/cval/homepage.html]. The Cowpea Genomics Initiative (CGI) is responsible for sequencing the gene-rich, hypomethylated portion of the cowpea genome [http://cowpeagenomics.med.virginia.edu] (Timko et al., 2008). Thousands of expressed sequence tags (EST’s) are being deposited on the public gene database GenBank by the Generation Challenge Program (GCP) [http://www.generationcp.org]. Various bacterial artificial chromosome (BAC) libraries are also being generated from various cowpea cultivars. Comparative genomics with the fully sequenced model legume, Medicago truncatula, hope to transfer genomic data to this phylogenetically related crop (Varshney et al., 2009; Choi et al., 2004; Zhu et al., 2005).
Conventional breeding strategies aim to stack desirable agronomic traits through marker-assisted selection and other molecular breeding techniques. Attempts to combine single gene traits and quantitatively inherited characteristics are underway at various research institutes. These include tolerance to abiotic stresses (drought, salinity, and heat), photoperiod sensitivity, plant growth type, and seed quality with resistances to the numerous bacterial, fungal, and viral diseases, as well as insect, invertebrate (nematode) and herbivorous pests (Timko and Singh, 2008). Cowpea genomics hope to find an application in the breeding for drought and pest tolerance, by increasing the speed of delivery of well-adapted varieties with desired stacked traits.

Genetic transformation of cowpea, regeneration of transformants or the stable inheritance of the transgene in subsequent progeny generations proves to be highly challenging. Efforts have been unsuccessful until TJ Higgins from CSIRO in Australia managed this feat by developing a successful transformation system (Popelka et al., 2006). New traits can, therefore, be introduced into cowpea with precision, with various insect-resistance traits, such as *Bacillus thuringiensis* (Bt) toxin, being first in line.

1.1.9 Identification of differentially expressed cowpea transcripts

In order to understand the gene networks that underlie plant drought stress, it is necessary to identify and characterise the genes that respond to drought stress. The first step in this quest, as well as in many other biological investigations, is to construct a cDNA library enriched for differentially expressed gene transcripts. Several techniques have been developed to isolate differentially expressed genes, but the SSH technique was employed during this study to identify and isolate the genes responding to drought in cowpea. SSH equalises the abundance of cDNAs in the target, increasing the likelihood of low-copy genes (rare transcripts) being incorporated into the library. Thus, SSH is expected to enrich for rare transcripts and filter out abundant transcripts, thereby limiting the redundancy of sequences in a cloned cDNA library (Diatchenko et al., 1996). Successful projects that followed a similar approach include the characterisation of the defence pathways in pearl millet and *Fusarium* wilt in banana (Van den Berg et al., 2004; Crampton et al., 2009). No application of SSH has yet been described for cowpea. Novel drought-inducible genes have been identified from cowpea through a differential screening approach different from SSH (Iuchi et al., 1996). Representational difference analysis (RDA) is another technique that has been used to look at how the cowpea genome changes under drought stress (van der Vyver, 2004).
Microarray studies of the generated library were used to verify the differential expression of individual genes in two populations. Many genes are expected to be influenced by water deficit, and microarray technology is particularly suited for the high throughput screening of large numbers of clones. The procedures as outlined in Berger et al. (2007) were followed to create and screen the library. PCR-amplified fragments of cDNAs are arrayed onto glass slides at high densities. The slides were hybridised to pairs of fluorescently labelled cDNA probes, produced by labelling RNA from two sources with a different colour. The gene expression in these two sources could then be directly compared on a large scale.

Microarrays have been applied to various plants to characterise the gene expression changes that occur under abiotic stress conditions. The salt stress response of two rice cultivars differing in salt tolerance differed in the rate of induction of stress-responsive transcripts (Kawasaki et al., 2001). A similar comparison between cultivars of barley under drought stress identified induced and constitutive transcripts specific to drought-tolerant genotypes, as well as general drought-responsive genes in barley (Guo et al., 2009).

A microarray spotted with a small number of genes of the desiccation tolerant plant *Xerophyta humilis* has been employed to identify cDNAs up-regulated by dehydration (Collett et al., 2004). Many of them correspond to various known dehydration stress-responsive genes. Various LEA genes have been identified among these cDNAs, suggesting that these proteins are important to provide mechanical and antioxidant protection of cells during water loss.

Microarrays have been used to characterise drought and cold-induced gene expression of *A. thaliana*. Full-length cDNAs of 1300 and 7000 genes have been used to construct cDNA microarrays (Seki et al., 2001; Seki et al., 2002). Cold- and drought-inducible gene expression overlapped frequently, and many of these genes have also been classified as dehydration response element binding (DREB1A) target genes.

All the examples of microarrays mentioned here, except for barley, involve the amplification of inserts from cloned libraries and spotting onto microarray slides. Modern technology has facilitated the production of pre-fabricated slides on which probes are synthesised chemically, such as the Affymetrix Barley GeneChip or *Arabidopsis* ATH1 Genome Array (http://www.affymetrix.com).
1.2 Plant responses to abiotic stress

During the evolution of plants, biotic and abiotic environmental changes are important forces that shape plant adaptations. Adverse abiotic environmental factors manifest through drought stress, salt stress and low temperatures, and of these, drought is the most important problem facing agriculture, since it prevents plants from being optimally productive (Zhu, 2002). A plant’s survival during times of drought or when grown in arid lands is dependent on its desiccation tolerance until soil moisture levels improve (Bernacchia and Furini, 2004). Desiccation is an extreme form of drought. Applications in developing drought tolerant crops can be generated by studying the desiccation tolerance mechanisms employed by plants.

1.2.1 Morphological and physiological responses of plants to drought

The first response of plants to drought is a rapid inhibition of shoot growth (Neumann, 2008). This adaptation is to reduce the leaf area that causes water loss through transpiration. Stomatal closure follows, which has a direct effect on plant photosynthesis due to decreased diffusion of CO₂ through stomata and mesophyll cells and a reduction of transpiration (Chaves et al., 2009). A secondary effect on photosynthesis is due to oxidative stress. Acclimation responses to restrict water expenditure include leaf shedding and growth inhibition, so that plants can maintain their water status (Chaves et al., 2009).

Plants also employ the accumulation of osmotic compounds as another strategy. Sugars have been found to accumulate in seeds at the onset of desiccation, and also in the vegetative tissues of resurrection plants (Bernacchia and Furini, 2004). They are involved in osmotic adjustment and to stabilise membrane structures and proteins under high concentrations, by forming a glass with the mechanical properties of a solid. Osmotic adjustment allows water uptake by the plant and the maintenance of cell turgor under water-stress conditions (Chaves et al., 2009).

Basal root growth is also inhibited by drought, primarily to preserve the availability of water to the apical meristem to ensure its survival (Neumann, 2008). When water levels are restored, the plant then has a better ability to resume growth. In deep soils, however, it is advantageous for the roots to maintain growth towards the water remaining in the deeper layers of drying soil.

Plants respond to periods of drought by shortening their vegetative growth period. This adaptation is observed in desert plants adapted to terminal drought environments (Neumann, 2008). Flowering and seed set is completed before the dry period starts.
The rest of this literature review will focus on the molecular responses that plants have to drought and other abiotic stresses.

1.2.2 Molecular responses of plants to abiotic stress
The abiotic stresses of drought, salt and cold all have the same effect on plants: a reduced water potential ($\psi_W$) inside the plant cell caused by a deficit of cellular water (Verslues et al., 2006). The molecular responses of plants to these abiotic stresses are, therefore, closely related and some of the mechanisms overlap (Zhu, 2002). Water deficit responses in plants are complex and affect almost all physiological and metabolic processes, and involve the expression and regulation of a large number of genes. Many adaptation mechanisms have been reported, some of them are plant specific, while others are common to many species (Bohnert et al., 1995).

Plant responses to stress include upstream signal transduction events (e.g. transcription factors and protein kinases), adaptive responses (e.g. ion transporters, aquaporins, protein degradation, membrane modification and protective proteins) and downstream responses (e.g. detoxification, accumulation of solutes or senescence).

Gene expression changes are responsible for all of the many adaptive responses. Osmolyte accumulation, metabolite changes, ion homeostasis, physical adaptations (e.g. membrane modification, cytoplasm stabilisation) and protein turnover changes are mediated through the activation of signaling pathways and modulation of transcription, e.g. transcription factors and mRNA stability (Bohnert et al., 1995).

Good comprehension of the pathways requires knowledge of the precise inputs and outputs of the pathways (Zhu, 2002). Not all changes that are seen upon drought treatment are due to the drought input signal, and not all of the outputs are necessarily adaptive responses. They may be consequences of the stress injury. These changes remain important targets for genetic suppression to improve salt and drought tolerance of crop plants.

1.2.3 Salt and drought stress signaling
Adaptive or presumed adaptive responses can be grouped into three groups: (1) homeostasis (ionic and osmotic); (2) stress damage control and repair (detoxification); and (3) growth control (Zhu, 2002). Salt and drought stress signaling can similarly be grouped into three functional categories or signaling pathways. They are (1) Ionic and osmotic homeostasis signaling pathways; (2) detoxification (damage control and repair) response pathways; and (3) growth regulation pathways. These signaling pathways regulate each
other negatively. When homeostasis is achieved, stress injury is reduced and less detoxification responses are necessary. Homeostasis and detoxification signaling lead to stress tolerance, and the growth inhibition effect is reduced.

1.2.3.1 Ionic and osmotic stress signaling contributes to re-establishing cellular homeostasis under stress conditions. An example of an ionic signaling pathway includes the SOS (Salt Overly Sensitive) pathway and protein kinases complexes which control the expression and activity of ion transporters (Na⁺, K⁺ or H⁺). The input is an excess of intra/extracellular Na⁺ ions. The input for osmotic signaling is a turgor change in the cell. Osmotic stress activates several protein kinases which mediates osmotic homeostasis and detoxification responses. Phospholipid systems, e.g. Phospholipase C (PLC), are activated to generate messenger molecules such as inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol (DAG) from phosphotidylinositol 4,5-bisphosphate (PIP₂). IP₃ releases Ca²⁺ from internal stores, and increases in cytoplasmic Ca²⁺ is important in inducing the expression of osmotic stress-responsive genes. Osmotic stress also regulates ABA biosynthesis at multiple stages. The output of osmotic signaling is unknown, but may include osmotic stress-activated protein kinases that mediate signaling, gene expression or activation of osmolyte biosynthesis enzymes, and transport systems for water and osmolytes (Zhu, 2002).

1.2.3.2 Detoxification signaling include most of the other changes that are induced by salt or drought stress. The function of detoxification signaling is to control, prevent and repair stress damages. The input is a product of stress injury, such as ROS or protein denaturation. Examples of these changes include phospholipid hydrolysis, gene expression changes (such as LEA or dehydrin, molecular chaperones and proteinases) and enzymes involved in ROS injury processes (Zhu, 2002).

1.2.3.3 Plant growth regulation (cell division and expansion)
Water deficit inhibits plant growth, by slowing cell division and reducing expansion processes of existing cells. This is effected by reduced cyclin-dependent kinase (CDK) activity, through suppressed transcription of cyclins and CDKs, and induction of CDK inhibitors. The input is probably a product of stress injury or any primary or intermediary signal of the homeostasis and detoxification pathways (Zhu, 2002).

The gene expression responses of A. thaliana during three methods of creating cellular water deficit, namely filter paper desiccation, mannitol treatment and progressive soil-water deficit, have been compared (Bray, 2004). Three repressed genes are found in common, and they
are all related to cell expansion. Repression of such genes by water deficit, therefore, results in decreased growth.

1.2.4 Drought stress responses

Drought is defined as a period of below normal precipitation that causes plant productivity in a natural or agricultural system to be limited. The stress is imposed through a decrease in the availability of soil water, quantified by the $\psi_W$. A plant will struggle to take up enough water at a decreased $\psi_W$. It responds by avoiding further water loss, switching on mechanisms to continue water uptake at a reduced $\psi_W$ or to tolerate reduced tissue water content (Verslues et al., 2006).

Plant responses to decreased water availability (low $\psi_W$) can, therefore, be categorised as either low $\psi_W$ stress and dehydration avoidance or dehydration tolerance.

During low $\psi_W$ stress avoidance, a plant avoids the decrease of its tissue $\psi_W$ and water content by balancing its water uptake and loss. In other words, it keeps the stress outside the plant tissue. The short term strategy employed is to close its stomata to reduce water loss. Water uptake is increased through increasing its root/shoot ratio, which is a longer term strategy. When water stress is mild or of short duration, avoidance strategies are beneficial to crop productivity. In the long run, however, the loss in photosynthesis due to closed stomata, or resources shifted into root growth, may have a negative impact on the plant’s productivity (Verslues et al., 2006).

If the plant is not able to withstand the low $\psi_W$ in the soil, the tissue $\psi_W$ is forced to equilibrate with the low soil $\psi_W$ by either losing water or through adjustments to decrease $\psi_W$ without any water loss. These adjustments are called dehydration avoidance and are mechanisms to avoid water loss. The mechanisms consist mainly of the accumulation of solutes (also called osmotic adjustment) or cell wall hardening. It enables the plant to maintain a high water content despite a low tissue $\psi_W$ (Verslues et al., 2006).

Osmotic adjustment lowers the leaf $\psi_W$ to below the external $\psi_W$. The resulting leaf to soil $\psi_W$ gradient is the driving force for water uptake by a plant (Bray, 1993). Proline is an example of a compatible solute that is accumulated in response to low $\psi_W$, salt and freezing, these being abiotic stresses that affect the water status of the plant. Compatible solutes can accumulate to high levels in a plant cell without interfering with its metabolism. The production of compatible solutes may be resource intensive, and may not be sufficient to increase water uptake in severely dry soil.
When cell walls are hardened, the cell volume does not change much when a small loss of water is experienced, but the turgor and $\psi_w$ drops steeply and thereby prevent further water loss. A disadvantage of low turgor is that the cell cannot expand, so this response to avoid water loss is only applied by non-growing tissues.

When the low-$\psi_w$ stress becomes too severe, it becomes impossible for the plant to avoid dehydration. Dehydration tolerance mechanisms are activated when water is lost from the plant cell, and the cell strives to avoid cellular damage caused by this. Tolerance mechanisms include the production of protective solutes and proteins, metabolic changes and ROS detoxification. Protective proteins, such as dehydrins, other LEA proteins and compatible solutes, are produced to protect proteins and membrane structures from the damage caused by dehydration. During tolerance to dehydration, and other abiotic and biotic stresses, the levels of ROS are controlled, or the damage caused by them is prevented.

Desiccation-tolerant plants display dehydration tolerance. They can recover from a fully air-dried state, and are, therefore, also called resurrection plants (Bernacchia and Furini, 2004; Collett et al., 2004). They are metabolically inactive when they are fully dehydrated. At the molecular level, this state resembles seed dormancy, and correlates to the freezing tolerance and the dehydration responses encountered in less tolerant species. Crop plants cannot recover from complete desiccation, but try to tolerate milder water loss while remaining metabolically active.

The mechanisms of dehydration avoidance and tolerance overlap in certain cases, such as compatible solute and dehydrin accumulation. The molecules may function to bind water and increase the cell water content, thereby avoiding dehydration, in addition to protecting cellular structures (dehydration tolerance). The low $\psi_w$ responses are controlled by a complex regulatory network. ABA is a key regulatory factor in many abiotic stress responses. It has been found to be active in regulating stomatal conductance and root growth (low $\psi_w$ avoidance), inducing compatible solute accumulation (dehydration avoidance) and synthesis of dehydrins and LEA proteins (dehydration tolerance).

Approaches to improve drought tolerance include the transfer of genes involved in drought avoidance or tolerance. These genes can be signaling components or downstream effector genes. Successes in A. thaliana to improve drought tolerance have rarely been transferred to crop plants. The proposed reason is that the selection of lines with an enhanced survival rate under lethal drying conditions does not translate to improved productivity (growth, biomass or seed yield) in crops when grown under moderate drought conditions (Skirycz et al., 2011).
When a mild stress assay was performed on several transgenic *A. thaliana* lines containing a selection of 25 previously identified “stress tolerance genes” (loss or gain of function), drought reduced all the genotypes’ growth equally. Severe drought often results in plants activating water-saving mechanisms, such as closing the stomata. A primary adaptation of plants under stress is also to reduce their growth, so that they can save their resources for reproduction. Both these mechanisms lead to unnecessary yield loss under mild stress conditions, which do not threaten the survival of the plant. A strategy to increase plant growth under stress, by limiting growth reduction, is proposed as a solution. These mechanisms need to be studied further.

### 1.2.5 Salt stress responses

Rapid responses to salt stress resemble responses to low $\psi_w$, but longer-term responses are salt-specific due to the ionic imbalances and toxic effects that excess Na$^+$ ions exhibit in a plant cell. Salt is carried to the shoot in plants that transpire, so blocking uptake and transport to the shoots is a specific tolerance mechanism. Maintaining ion homeostasis through activating salt export or compartmentalising it in the vacuole is important in avoiding salt injury. Several *salt overly sensitive* (SOS) mutants have been identified by Zhu *et al.* (2002) that are hypersensitive to salt, but not to non-ionic osmotic stress. Potassium nutrition plays an important role in plant salt tolerance (Zhu *et al.*, 1998). The SOS pathway regulates ion transport processes and genes specifically involved in tolerance to salt stress. The sensor for cytoplasmic Na$^+$ ions is presumably located on the long C-terminal tail of SOS1, a plasma membrane Na$^+$/H$^+$ antiporter (Shi *et al.*, 2000). Over-expressing the SOS genes involved in the salt regulation pathway in *A. thaliana* results in plants with an increased salt tolerance (Yang *et al.*, 2009). Approaches to engineer salt tolerance in crop plants are generally to keep their cytosolic Na$^+$ levels low. This can be achieved by increasing outflow by over-expressing plasma membrane Na$^+$ exporters and tonoplast importers to increase vacuole sequestration, or by restricting inflow by lowering plasma membrane Na$^+$ importers (Zhang *et al.*, 2004).

Root growth is inhibited in medium containing NaCl due to the direct toxicity of Na$^+$. A root bending assay is used to quickly identify seedlings that are able to continue growth after transfer to salt-containing media and inverting the plates so that the roots point upwards (Zhu *et al.*, 1998; Verslues *et al.*, 2006; Yang *et al.*, 2009).

Plant responses to salt and drought are very similar, since both cause an osmotic stress. Mechanisms of these two stress signaling pathways frequently overlap (Zhu, 2002; Shinozaki and Yamaguchi-Shinozaki, 2007). Many overlapping genes have been found to be inducible
by drought, high salinity and/or ABA in *A. thaliana* (Seki *et al.*, 2002). Osmotic signaling pathways can activate the gene expression of LEA and dehydrin-type genes, chaperones and proteinases, all of which form part of the plant’s detoxification pathways that respond to products of stress injury, such as ROS or denatured proteins.

### 1.2.6 Cold stress responses

Cold stress of plants can be divided into two situations. Chilling occurs when the temperature is lower than the plant’s normal growing temperature. Freezing is the formation of extracellular ice crystals that dehydrates the cell. Ice crystals first develop in xylem vessels and intercellular spaces, such as substomatal cavities. It cannot penetrate an intact plasma membrane, but decreases the $\psi_W$ of the extracellular space. To re-establish equilibrium of $\psi_W$ across the membrane, water therefore has to move out of the cells and cell walls. This is the same as what happens in plants in drying soil, known as cytorrhytic dehydration. Consequently, the same mechanisms have been described for freezing tolerance than dehydration tolerance, e.g. accumulation of compatible solutes or proteins, and ROS detoxification. For example, over-expression of a DREB transcription factor (TF), the *DREB1A/CBF3* cDNA, showed increased tolerance to all abiotic stress conditions, namely dehydration, salt and freezing, since a wide range of stress responsive genes are switched on (Kasuga *et al.*, 1999; Liu *et al.*, 1998).

The primary event in cold perception by plants is considered to be the cold-induced membrane rigidification (Mukhopadhyay *et al.*, 2004). Membrane rigidification at low temperature can be prevented experimentally by benzyl alcohol, which is a membrane fluidiser. To simulate membrane rigidification at room temperature, plants can be treated with DMSO which is a membrane rigidifier.

A common factor between plants’ adaptation to cold and drought stresses is that these mechanisms are mostly under transcriptional control. In addition, some signal transduction pathways are dependent on ABA while others are ABA-independent. To engineer crop plants with enhanced cold and drought tolerance require the regulated expression of many genes, perhaps through a TF, such as DREB1 (Zhang *et al.*, 2004).

### 1.2.7 The involvement of ABA in abiotic stress

ABA is an important phytohormone. It is involved in the regulation of physiological and developmental processes, such as stomatal closure and dormancy of seeds and buds. It also plays an important role in the response of plants to water deficiency, as well as cold and salinity stress. The main function of ABA is to regulate the plant’s water balance and osmotic
stress tolerance, as proven by mutants deficient in ABA (Zhu, 2002). The water balance is regulated by controlling the guard cells of the stomata, while its role in cellular dehydration tolerance is achieved through the induction of genes coding for dehydration tolerance proteins (Zhu, 2002). Genes responsive to ABA have the cis-element ABRE (ABA responsive element) in their promoters. Examples include Em proteins and A. thaliana rd29B.

ABA’s role in biotic stresses is mainly to negatively regulate plant defence responses (Bari and Jones, 2009). ABA-deficient mutant plants are more resistant to various biotrophic and some necrotrophic pathogens. Applied ABA increased susceptibility of various plants to bacterial and fungal pathogens (Bari and Jones, 2009). There are, however, also examples where ABA positively regulates defence, e.g. by closing the stomata to prevent bacterial infection. ABA influences plant defence through modulating cell wall metabolism, such as callose deposition, production of ROS intermediates and defence gene expression. The biotic and abiotic stress signaling pathways are, therefore, interconnected through the involvement of ABA.

Osmotic or drought stress triggers the endogenous production of ABA in plants (Zhu, 2002). For example, the endogenous ABA levels of 10 hour dehydrated cowpea plants are 160-fold higher than in unstressed plants (Iuchi et al., 1996). ABA activates the metabolic signaling that leads to stress adaptation, after a stress has been perceived by the plant. ABA is thought to be the signal produced by roots growing in drying soil, which causes the leaf stomata to close, so that cell turgor can remain high. The mechanisms of increasing ABA levels are through activated synthesis, inhibition of ABA degradation or release of sequestered ABA stored in organelles. The rate-limiting step in the biosynthesis pathway is the cleavage of 9-cis-epoxycarotenoid, catalysed by the enzyme NCED (9-cis-epoxy carotenoid dioxygenase). Transcription and translation is required for ABA to be synthesised during stress (Bray, 1993). The signaling between osmotic stress perception and the induction of ABA biosynthesis genes are unknown. Signaling is proposed to occur through calcium signaling or protein phosphorylation cascades. The candidate sensor/ receptor for drought/ salt/ ABA or cold signaling in A. thaliana is probably a histidine kinase, AtHK1, also called the osmosensor (Urao et al., 1999). This gene’s transcription is induced by osmotic stress, and it is able to complement an osmosensing-defective yeast mutant. It senses and transmits the stress signal to a downstream MAPK cascade.

An increase in ABA content due to desiccation induces transcription of response genes (luchi et al., 1996). The exogenous application of ABA also induce proteins known to be induced by
drought, therefore, ABA has a role in inducing proteins in response to drought stress (Bernacchia and Furini, 2004).

The importance of ABA during drought stress is demonstrated through the up-regulation of the NCED enzyme in cowpea during drought stress (Iuchi et al., 2000). The enzyme accumulates in cowpea within 2 hours of drought stress, primarily in the leaves and stems, before any ABA accumulation is detected. This enzyme, therefore, plays a key role in ABA biosynthesis in cowpea under drought stress. This enzyme has been transformed into creeping bentgrass, and is able to also increase the ABA levels of the transgenic plants (Aswath et al., 2005).

Not all osmotic stress responses are mediated by ABA. Work performed with A. thaliana mutants deficient in or insensitive to ABA elucidated that there are different water stress signal-transduction pathways. Osmotic stress responsive genes can be induced by ABA (ABA-dependent pathway), induced independent of ABA, or be only partially ABA dependent (Shinozaki and Yamaguchi-Shinozaki, 2007). Bray (1993) categorises genes into these three types: ABA-responsive, ABA-requiring or non-ABA responsive genes. An example of ABA-independent regulation (also called ABA irresponsible) is the presence of the DRE cis-element in downstream stress response genes’ promoters, such as rd29A of A. thaliana. They are induced by drought, salt and cold without the direct involvement or any observed increased levels of ABA. ABA alone cannot activate the DRE element (Seki et al., 2001).

1.2.8 Stress- and ABA-regulated gene expression
A large number of plant genes have been identified that are regulated by drought, salt, ABA or cold stress (Seki et al., 2001 & 2002; Collett et al., 2004; Bray, 2004; Shinozaki and Yamaguchi-Shinozaki, 2007; Guo et al., 2009). They may be involved in the initial stress response or to facilitate plant stress tolerance. Their products may function in ionic homeostasis, such as ion transporters; osmotic homeostasis, such as aquaporins and osmolyte biosynthesis genes; and detoxification or damage prevention and repair. Examples of such genes include the LEA or dehydrin proteins, detoxification enzymes and chaperones. Osmolyte biosynthesis enzymes have a dual role, since osmolytes may also function to prevent or repair damage caused in a cell by water deficit.

1.2.9 Transcription factors and DNA recognition sequences
The regulation of stress-responsive genes may be early or delayed in response to the stress. The model for drought gene expression, after the detection of both ABA dependent or independent signals, is to modify constitutively expressed TFs post-translationally, e.g.
through phosphorylation (Zhu, 2002). These TFs switch on the expression of early response genes. These genes typically encode transcriptional activators, also called element binding proteins because they bind to cis-elements in the promoters of downstream delayed-response genes. Their expression is very quick after stress and sometimes only transiently. Examples of early response genes in salt, drought, cold or ABA regulation include CBF/DREB, RD22BP and AtMyb. The downstream delayed-response genes constitute the majority of stress-responsive genes. They are activated more slowly than the TFs and have a more sustained expression pattern. They are also called stress tolerance effector genes, and contain specific cis-elements in their promoters depending on the type of stress, which is recognised by the early response element binding proteins.

Cis-elements are DNA sequences in promoters of genes that are recognised and bound by TFs. They determine the abiotic-stress responsiveness of a gene. DRE (dehydration-responsive elements) confers dehydration, salt and cold specificity in an ABA-independent manner, and are recognised by the DREB or CBF/DREB early response binding proteins. They are of the AP2 TF type (Zhu, 2002). The DRE core sequence is (TACCGACAT) (Seki et al., 2001). Transgenic plants in which the DREB1A/CBF3 cDNA has been over-expressed shows increased tolerance to all abiotic stress conditions, namely dehydration, salt and freezing, since a wide range of stress responsive DREB1A target genes are switched on (Kasuga et al., 1999; Liu et al., 1998).

ABRE (ABA-responsive elements) confer ABA-specific inducibility to genes, e.g. Em and rd29B, and are bound by e.g. EmBP1, osZIP-1a, TRAB1 and AREB1 (ABA-responsive element binding protein). These TFs are of the bZIP type (Zhu, 2002). The ABRE core sequence is (PyACGTG (T/G)C) (Seki et al., 2001).

Examples of genes containing the DRE or ABRE cis-elements are the Arabidopsis desiccation responsive genes rd29A and rd29B, respectively. These two genes are tandemly located in a 8 kb region of the A. thaliana genome (Yamaguchi-Shinozaki and Shinozaki, 1993). They encode hydrophilic proteins and have three introns each at the same positions. The rd29A gene promoter contains the 9 bp DRE element (TACCGACAT), and is quickly induced (within 20 min) after desiccation. This element is shown to be important for the regulation of gene expression in response to drought, high salt and cold stresses, in an ABA-independent manner. rd29B accumulates later in response to ABA (3 h), and contains the cis-element ABRE (Yamaguchi-Shinozaki and Shinozaki, 1994).
A third type of abiotic stress-responsive *cis*-elements has the MYCRS or MYBRS elements (CANNTG and NyAACPyPu, respectively). They confer dehydration and ABA inducibility, and are bound by TFs of the myc/myb type, such as RD22BP and AtMyb2 (Abe *et al.*, 1997, 2003). The following example illustrates the role of the MYB *cis*-element during ABA regulation. The *Arabidopsis* gene family of MYB TFs contains 196 members. They are classified into four subgroups on the arrangement of their MYB domains (1R-, R2R3-, 3R- and 4R-MYB), which forms helix-turn-helix DNA binding motifs. The *Arabidopsis* MYB60 TF, in the R2R3-MYB group, is regulated by different levels of ABA, and modulates root growth and stomatal behaviour during drought stress of plants (Oh *et al.*, 2011).

Some DREB1A target genes also contain the ABRE sequence in their promoter regions (Seki *et al.*, 2001). Many drought- and cold-inducible genes are, therefore, controlled by both ABA-dependent and ABA-independent pathways. Novel *cis*-acting elements may also exist in genes regulated by cold and drought, that do not contain any of the *cis*-elements in their promoters as described above (DRE or ABRE sequences).

### 1.3 Late embryogenesis abundant (LEA) proteins

LEA proteins were initially discovered in mature embryos of plant seeds that underwent desiccation. This early observation is responsible for the protein’s name (Baker *et al.*, 1988). Their role were thought here to protect the embryo from desiccation in the seed. A relationship between the expression of LEA genes, ABA and seed dormancy was suggested early on (Dure *et al.*, 1989). LEA proteins have subsequently been described in various plants and plant tissues, and a few other phylogenetically distant organisms studied to date (e.g. bacteria, anhydrobiotic nematodes, rotifers (Goyal *et al.*, 2003; Tunnacliffe and Wise, 2007)). Several LEA genes are expressed in vegetative tissues, making their classification as specific to late embryogenesis no longer true (Bies-Ethève *et al.*, 2008).

#### 1.3.1 Characteristics of LEA proteins

LEA proteins are typically small (10-30 kDa), with a high content of hydrophilic amino acids, therefore highly water soluble. Some amino acids, such as glutamine, glycine and alanine, are over-represented, while cysteine, phenylalanine, isoleucine, leucine and tryptophan are significantly under-represented (Wise, 2003). There is a bias in amino acid arrangements, with certain amino acid sequence motifs found tandemly repeated. The periodicity of the repeats enables the proteins to arrange into amphiphilic helices (Dure *et al.*, 1989). Because LEA proteins are ‘natively unfolded’ or ‘intrinsically disordered’ proteins, they do not denature at high temperatures and remain soluble (Russouw *et al.*, 1997). Heat treatment is, therefore, a common step in LEA protein purification protocols.
LEA proteins are able to bind water due to their charged amino acids. Due to their amorphous random coils they are able to solvate structural surfaces of macromolecules through their hydroxyl groups (Baker et al., 1988; Wise and Tunnaciffe, 2004). They provide a water-rich micro-environment to important enzymes, thereby maintaining proteins’ integrities and preventing them from being inactivated while there is a shortage of water (Battaglia et al., 2008). LEA proteins are, therefore, frequently implicated as important during abiotic stress tolerance in plants, having a protective role on cellular structures and acting as chaperones to protect proteins under water deficit stress.

1.3.2 Classification of LEA proteins

LEA proteins are commonly grouped into at least six different groups, based on their nucleotide and amino acid sequences, and specific arrangements of amino acid motifs (Dure et al., 1989; Battaglia et al., 2008; Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008). Bioinformatic tools, namely peptide profile analysis (Protein or Oligonucleotide Probability Profile, POPP), have also been used to identify super-families of LEA proteins containing over- or underrepresented amino acids in protein sequences (Wise, 2003; Wise and Tunnaciffe, 2004). These groupings correspond closely to the traditional groups.

The three main groups (groups 1 to 3) are considered the true LEA proteins. These LEA proteins frequently contain tandemly repeated amino acid motifs, typically charged and hydrophilic, of specific lengths, that is predicted to form amphipathic α-helices under dehydrated conditions. The hydrophobic faces may bind to each other to stabilise each other through non-polar bonds, while the solvent-exposed surface contain a high concentration of charged residues (Baker et al., 1988). The amphipathic α-helices are also thought to interact with the cell membrane, to protect it during dehydration (Battaglia et al., 2008).

Group 1 LEA proteins have a high glycine content, are therefore highly charged and hydrophilic, and exist as random coils in solution. They are characterised by a 20-mer hydrophilic motif that are tandemly repeated up to four times in plants. Plant group 1 LEA proteins also contain conserved N- and C-terminal motifs (Figure 1.2, Battaglia et al., 2008). An example includes the wheat Em protein.

LEA proteins from group 2 are also called dehydrins. They have similar properties to group 1, but are characterised by unique arrangements of three motifs: an N-terminal Y motif, a lysine-rich K motif, with both present in multiple copies, and a tract of serine residues called the S motif (Figure 1.2, Battaglia et al., 2008). Five different subgroups are classified based on the presence and arrangements of the KYS motifs. This group is found in all types of
plants studied to date, including nonvascular plants, such as mosses, seedless and seed plants. Examples include the cowpea and wheat dehydrins (Ismail et al., 1999a and b; Brini et al., 2007). No examples of group 2 proteins have been found in organisms outside the plant kingdom. Since they have a more complex domain structure, they may have evolved uniquely in plants (Tunnacliffe and Wise, 2007).

Group 3 LEA proteins contain an 11 amino acid repeating motif. Due to variability in the 11-mer, they are classified into two subgroups. Subgroup 3A is more conserved, with always the same motifs at the N- and C-termini. Subgroup 3B is more heterogeneous, but with a highly conserved 5th motif unique to the 3B subgroup (Figure 1.2, Battaglia et al., 2008). Examples include the barley HVA1 and wheat group 3 gene (Xu et al., 1996; Ried and Walker-Simmons, 1993). This group is found in plants as well as non-plant organisms, such as prokaryotes and nematodes, where they accumulate in the desiccated state (Goyal et al., 2003).

Group 4 LEA proteins are conserved at the N-terminus, where 80 residues are predicted to form amphipathic α-helices (Battaglia et al., 2008). Four other conserved motifs are found in many group 4 LEA proteins (Figure 1.2, Battaglia et al., 2008). They are also subgrouped into two groups, on the basis of the presence/absence of the 4th and 5th motifs, with the cotton D-113 gene an example of the 4B subgroup.

Group 6 LEA proteins are small in size (7-14 kDa) and highly conserved, with four motifs distinguishing this group (Figure 1.2, Battaglia et al., 2008). An example includes the bean PvLEA18 protein (Colmenero-Flores et al., 1999).

Group 7 contains the ASR proteins, hydrophilins that share physiochemical properties and expression patterns of LEA proteins. They contain five highly conserved regions, containing stretches of Histidine residues. The C-terminal region contains a nuclear localization signal (Figure 1.2, Battaglia et al., 2008).

Figure 1.2 does not contain a representation of the distinctive motifs of group 5 LEA proteins, because they do not have any (Battaglia et al., 2008). This group will be discussed in the next section.
Figure 1.2 Array of the distinctive amino acid motifs in the LEA protein groups (from Battaglia et al., 2008). Each block contains a schematic representation of the arrangement of the motifs that distinguish each group of LEA proteins and their corresponding subgroups. Although similar colors and numbers indicate the different motifs for each group, they do not imply any sequence relation among the motifs in the different blocks. The range of protein sizes in each group is indicated at the top of each block, in number of amino acid (aa) residues.
1.3.3 Atypical LEA proteins

All LEA proteins with a significant higher proportion of hydrophobic residues than typical LEA proteins are grouped into Group 5 (Battaglia et al., 2008). This group, therefore, includes non-homologous proteins, and is classified further according to sequence similarities to the first proteins discovered in cotton, called LEA12, LEA5 and LEA14. Their cDNAs are labelled D-34, D-73 and D-95, respectively (Baker et al., 1988; Dure, 1993; Galau et al., 1993). The LEA5 and LEA14 genes are highly induced in water-stressed cotton leaves (Galau et al., 1993). These three atypical genes all probably arose from a common ancestral group 5 gene sequence (Battaglia et al., 2008). Tunnacliffe and Wise (2007) suggest that this atypical protein group should be considered to be outside of the LEA protein family.

Due to variation in protein sizes and sequence differences, it is proposed that the group 5 LEA proteins (Battaglia et al., 2008) are split into their three groups based on their Pfam domain content, as suggested by Hundertmark and Hincha (2008) during their classification of A. thaliana LEA genes (Figure 1.3). The LEA_2 domain is found in LEA proteins corresponding to the cotton D-95 type, and genes are frequently named LEA14. Examples include the A. thaliana LEA14 (Singh et al., 2005) and cotton LEA14 (Galau et al., 1993). LEA5 proteins contain the LEA_3 domain and are of the D-73 type. Genes of this type have been cloned from citrus (Naot et al., 1995), soybean (Burns et al., 1996), cotton (Galau et al., 1993), Lotus japonicus (Haaning et al., 2008), A. thaliana (Gosti et al., 1995; Miller et al., 1999), Craterostigma plantagineum (Piatkowski et al., 1990) and various other plants. The third class is called seed maturation proteins (SMP), is of the D-34 type and is frequently called LEA12. Examples include MtPM25 from M. truncatula (Boudet et al., 2006) and Atrab28 from A. thaliana (Arenas-Mena et al., 1999; Borrell et al., 2002).

A number of extensive reviews have been written on the typical LEA proteins, but they focus mostly on the groups other than group 5 (Tunnacliffe and Wise, 2007; Battaglia et al., 2008; Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008). Very little is said about the properties of group 5 proteins, except that they are atypical, have significantly more hydrophobic residues and with no sequence conservancy. Alignment of known LEA5-like proteins from different plant species, however, indicates conservancy of the N-terminal region, probably coding for a hydrophobic signal peptide for transport to the endoplasmic reticulum, a small alanine rich domain (RRGYAAAPAAV) and a 32-amino acid domain at the C-terminal (Jain and Minocha, 2000). They are predicted to have a globular conformation and are, therefore, not soluble after boiling (Singh et al., 2005).
1.3.4 Three-dimensional structure of LEA proteins

The charged amino acids in the amino acid repeats found in typical LEA proteins have a high water binding capacity. In aqueous solutions the typical LEA proteins are, therefore, highly unstructured. During dehydration, however, they have been found to change into folded structures (Battaglia et al., 2008; Goyal et al., 2003). The folded structures may protect the cell from collapse by stabilising membranes and acting as space fillers at low water activity. An example is from the anhydrobiotic nematode *Aphelenchus avenae*. Its group 3 LEA-like protein (*Aav-LEA1*) is natively unfolded in solution, but increases in α-helix content upon drying (Goyal et al., 2003). During dehydration LEA proteins also interact with oligosaccharides, strengthening the average hydrogen bond strength in the glassy matrix that is formed in the cell.

A group 5 LEA protein that had its secondary structure determined, is from the D-34-like SMP group. It is MtPM25 from *M. truncatula* (Boudet et al., 2006). It has been found to form α-helices and β-sheets upon drying. Similarly, the *Lotus japonicus* D-73 type LEA5 protein is found to be intrinsically disordered, characteristic of chaperone proteins, but it undergoes a large conformational shift during dehydration to become largely α-helical (Haaning et al., 2008).

In contrast to the unstructured LEA proteins in solution, a member of group 5 of the D-95 type, *A. thaliana* LEA14, exhibits a stable three-dimensional structure in solution consisting of one α-helix and seven β-strands that form two antiparallel β-sheets (Singh et al., 2005).

1.3.5 Expression profile of LEA genes

LEA gene expression is regulated by various abiotic stresses, such as drought, salinity, cold and high light, as well as the plant stress hormone ABA (Hundertmark and Hincha, 2008). An example where LEA protein gene expression has been found to be induced by ABA and water stress is in bean (*Phaseolus vulgaris*), where six cDNAs corresponding to two types of LEA proteins have been discovered (Colmenero-Flores et al., 1997). A LEA protein of group 3 is also up-regulated in leaves of *P. vulgaris* under progressive drought stress (Kavar et al., 2008). Iuchi et al. (1996) identified a cowpea group 2 LEA (CPRD22) that is up-regulated by desiccation, exogenous ABA application and high salt.

Various cis-elements have been discovered in LEA gene promoters, leading to the stimulus-responsive expression of their downstream genes. Many LEA genes from *Arabidopsis* contain ABA response (ABRE) and/or low temperature response (LTRE) elements in their promoters, and are inducible by ABA, cold or drought (Hundertmark and Hincha, 2008).
1.3.6 Subcellular localisation of LEA proteins

LEA proteins are localised to the cytoplasm and nuclear regions of the plant cell. Some groups are also targeted to chloroplasts and mitochondria, through the presence of signal peptides (Tunnacliffe and Wise, 2007; Hundertmark and Hincha, 2008). They are not predicted to be transmembrane proteins.

1.3.7 LEA gene families

Plant genomes invariably contain multiple copies of LEA genes. For example, barley has a dehydrin (LEA group 2) gene family of at least 13 members (Choi et al., 1999). Peanut has at least 19 LEA-encoding genes that can be classified into eight groups (Su et al., 2011). The A. thaliana genome contains fifty-one LEA genes (Figure 1.3). Several of them arose from whole-genome and direct tandem duplications (Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008).

The Arabidopsis genome resulted from an ancient genome duplication event, followed by gene loss from the tetraploid genome to result in the current diploid genome. Subsequent sequence diversification took place, but a high proportion (43%) of the duplicated LEA genes have been retained, at a higher than average percentage than all the rest of the genes in the genome (26%). Random mutation and loss of function of redundant duplicated genes usually happens after a polyploidy event. This has been avoided by LEA genes through functional diversification, at the level of coding sequence, subcellular localisation or expression patterns. Conservation of the modifications indicates that they are beneficial to the organism. Bies-Ethève et al. (2008) obtained a complete overview of the regulation of all the LEA genes in A. thaliana through expression studies with semi-quantitative and quantitative RT-PCR, as well as microarray data available through Genevestigator (www.genevestigator.ethz.ch/at/index). Diverse expression patterns of the different gene family members are observed in maturing seed and other plant organs. No single group of LEA genes has a homogeneous pattern of expression. This suggests that it confers an evolutionary advantage to an organism to express at least one copy of each gene family in each tissue at any developmental stage (Bies-Ethève et al., 2008).

LEA proteins from the different groups do not show any sequence similarity, therefore do not seem to share a common ancestor. But LEA proteins do share an important physiochemical characteristic, related to their common goal to protect cellular functions under dehydration. This characteristic, therefore, is proposed to have developed independently in the different protein families, or within different organisms (Battaglia et al., 2008).
1.3.8 Functional role of LEA proteins

LEA proteins are not enzymes, since they are present in high molar concentrations in late embryogenesis tissues and have biased amino acid arrangements (Baker et al., 1988). Because they are soluble and unstructured they are not implicated with any structural role. Because many LEA proteins are expressed during desiccation, and they have been shown to undergo structural changes to a more folded state, their role in protecting the structural integrity of proteins and membranes in desiccating plant cells is suggested (Baker et al., 1988). Since LEA proteins are found in all cellular compartments and plant tissues, it can be inferred that their function is required in all cellular locations during water stress (Tunnacliffe and Wise, 2007).

The highly hydrophilic and randomly coiled LEA proteins may substitute the water molecules required to solvate cytosol components. Molecular shielding is the term given to the ability of LEA proteins to prevent protein aggregation due to their steric hindrance (Wise and
Tunnacliffe, 2004; Tunnacliffe and Wise, 2007).

The amphiphilic helices formed upon desiccation may also project charged amino acid residues that can bind ions that increase in concentration during desiccation of the cytosol. Metal ions are involved in the production of ROS, therefore, LEA proteins are also described as having anti-oxidant functions. Histidine and glycine is over-represented in group 2 LEA proteins, making them highly flexible with histidine residues making multiple contacts with metal ions (Tunnacliffe and Wise, 2007). An example of this function of dehydrins (group 2) is the protection of membranes under cold stress, through scavenging free radicals and thereby inhibiting lipid peroxidation (Hara et al., 2003 & 2004).

The same role is also predicted for group 3 LEA proteins. It is suggested by Baker et al. (1988) that amphiphilic helices within a LEA protein may undergo intramolecular binding of the surfaces to form a four helix barrel. The hydrophobic faces of the helices in group 3 proteins may interact to form a homodimer, while the outside charged surface may sequester ions that accumulate during water deficit (Dure, 1993). Group 3 LEA proteins are also predicted to form filaments during desiccation that mechanically strengthens the cell (Wise and Tunnacliffe, 2004).

1.3.8.1 In vitro protection assays
The roles of LEA proteins have been investigated in the literature by means of in vitro protection assays of enzymes that are sensitive to denaturation due to dehydration or freezing. The enzyme lactate dehydrogenase (LDH) is protected from aggregation due to desiccation and freezing (Goyal et al., 2005) by adding increasing amounts of group 1 or 3 LEA proteins (from wheat and an anhydrobiotic nematode, respectively). The remaining LDH enzyme activity after treatment is assayed by adding a substrate solution containing NADH and pyruvate, and measuring the change in absorbance at 340 nm. The LEA proteins act as molecular chaperones in synergy with the non-reducing disaccharide trehalose, which is known to accumulate during dehydration of plants and nematodes. A. thaliana Cor15am, a LEA protein expressed during cold acclimation, is also able to prevent protein aggregation during freezing (Nakayama et al., 2008). The citrus dehydrin CuCOR19 is also able to protect catalase and LDH against freezing inactivation (Hara et al., 2001). Enzyme activities are stabilised and preserved by the wheat dehydrin DHN-5 from adverse effects induced by heating (Brini et al., 2010). The same stabilising effect is obtained on proteins in vitro by the intrinsically disordered protein from the legume Lotus japonicus. It prevents misfolding and acts much like a chaperone protein (Haaning et al., 2008).
1.3.8.2 Transgenic studies

Over-expression of plant LEA proteins in yeast have conferred increased tolerance to osmotic, freezing or salt stress (Mowla et al., 2006; Imai et al., 1996). Over-expression of a soybean group 3 LEA protein, PM2, in the bacteria *E. coli* results in increased tolerance to temperature extremes and high-salinity stresses (Liu Y et al., 2010). It is suggested that group 3 LEA proteins are effective against a salt stress rather than osmotic stress (Tunnacliffe and Wise, 2007).

Enhanced tolerance to abiotic stresses has also been observed in transgenic plants over-expressing certain LEA genes (Table 1.1). In a successful attempt to transfer the protective role of a LEA protein, creeping bentgrass has been transformed with the barley *hva1* gene. These transgenics shows better survival under water-stressed conditions, compared to non-transgenic control plants (Fu et al., 2007). This barley gene codes for a group 3 LEA protein member. The same gene has also been used to transform a variety of other crop plants, such as rice, wheat and mulberry (Xu et al., 1996; Bahieldin et al., 2005). *Arabidopsis* transformed with the wheat dehydrin *DHN-5* (group 2) has improved tolerance to high salinity and water deprivation (Brini et al., 2007). Lettuce and kidney bean have been transformed with a *Brassica napus* LEA group 3 gene, and shows enhanced tolerance to salt stress and water deficit stress (Park et al., 2005; Liu et al., 2005). Over-expression of an *A. thaliana* SMP-class LEA protein, *Atrab28*, leads to an increase in germination under control and high salinity conditions, and transgenics survive better in 14 mM LiCl (Arenas-Mena et al., 1999; Borrell et al., 2002). Tobacco has been transformed with two group 4 LEA genes from the resurrection plant *Boea hygrometrica*. Membrane and protein stability are protected during dehydration stress (Liu et al., 2009). Over-expression of *Brassica napus* group 4 LEA genes in transgenic *A. thaliana* lead to an increase in stress tolerance, as indicated by cold, dehydration and salt tolerance assays (Dalal et al., 2009).

Expressing LEA genes do not always result in improved stress tolerance. LEA genes from the resurrection plant *Craterostigma plantagineum* has not conferred higher stress tolerance to transgenic tobacco (Iturriaga et al., 1992). One of them, pcC27-45, is related to the hydrophobic group 5 D-95 type (Piatkowski et al., 1990). There are no effects of *Craterostigma* LEA proteins on desiccation-related physiological functions of transgenic tobacco plants. Reasons proposed by the authors include the unsuitability of the measured parameters, the requirement of simultaneous expression of several drought-related proteins to have an effect, or the cooperation with other substances, such as osmolytes, since the introduced drought-related protein is necessary, but not sufficient on its own, for
osmoprotection (Iturriaga et al., 1992). One LEA gene may not be sufficient, or it did not function in the heterologous plant as in their native context (Bray, 1993).

A group 2a LEA protein, RAB18, has been over-expressed in A. thaliana with no improvement in freezing or drought tolerance (Lång and Palva, 1992). When pairs of group 2 LEA genes have been over-expressed in the same plant, freeze tolerance assays shows improvement, but only when they are of the group 2b type. No differences are, however, observed in drought tolerance (Puhakainen et al., 2004).

A loss-of-function approach has been followed by (Olvera-Carrillo et al., 2010) to identify the participation of group 4 LEA proteins of Arabidopsis in its adaptation to water-limiting stress conditions. Since there are multiple LEA proteins in organisms, functional overlap may obscure their role (Tunnacliffe and Wise, 2007).

A recent literature survey revealed that only one group 5 member of the D-73 type has been tested for its role in adaptation to stress conditions by transforming it into another plant (Wang et al., 2006; Zhao et al., 2011). A LEA5 gene has been isolated from Tamarix androssowii, a shrub or tree able to tolerate saline soil. Transgenic tobacco and blueberry have been produced that shows enhanced drought and cold tolerance, respectively.
Table 1.1 Studies of transgenic plants containing LEA genes. Examples of LEA genes and their effect on the abiotic stress tolerance of transgenic plants are presented.

<table>
<thead>
<tr>
<th>LEA gene name</th>
<th>Group</th>
<th>Source organism</th>
<th>Transformed plant</th>
<th>Result during dehydration stress</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHN-5</td>
<td>2</td>
<td>Wheat</td>
<td>A. thaliana</td>
<td>Improved tolerance to high salinity and water deprivation</td>
<td>Brini et al., 2007</td>
</tr>
<tr>
<td>Dehydrin</td>
<td>2</td>
<td>Citrus unshiu</td>
<td>Tobacco</td>
<td>Enhanced cold tolerance and inhibition of lipid peroxidation</td>
<td>Hara et al., 2003</td>
</tr>
<tr>
<td>RAB18</td>
<td>2a</td>
<td>A. thaliana</td>
<td>A. thaliana</td>
<td>No improvement in freezing or drought tolerance</td>
<td>Lång and Palva, 1992</td>
</tr>
<tr>
<td>Hva1</td>
<td>3</td>
<td>Barley</td>
<td>Creeping bentgrass</td>
<td>Maintain high water content, better survival</td>
<td>Fu et al., 2007</td>
</tr>
<tr>
<td>Hva1</td>
<td>3</td>
<td>Barley</td>
<td>Rice</td>
<td>Tolerance to water deficit and salinity</td>
<td>Xu et al., 1996</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Barley</td>
<td>Wheat</td>
<td>Improved biomass and grain yield under field conditions</td>
<td>Bahieldin et al., 2005</td>
</tr>
<tr>
<td>ME-leaN4</td>
<td>3</td>
<td>Rape (Brassica napus)</td>
<td>Lettuce (Lactuca sativa L.)</td>
<td>Enhanced growth under water deficit and salt stress</td>
<td>Park et al., 2005</td>
</tr>
<tr>
<td>ME-leaN4</td>
<td>3</td>
<td>Rape (Brassica napus)</td>
<td>Kidney bean (Phaseolus vulgaris L.)</td>
<td>Enhanced growth under water deficit and salt stress</td>
<td>Liu et al., 2005</td>
</tr>
</tbody>
</table>
Table 1.1 continued

<table>
<thead>
<tr>
<th>LEA gene name</th>
<th>Group</th>
<th>Source organism</th>
<th>Transformed plant</th>
<th>Result during dehydration stress</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcC6-19</td>
<td>?</td>
<td>Resurrection plant</td>
<td>Tobacco</td>
<td>No difference measured in ion-leakage tests</td>
<td>Iturriaga et al., 1992</td>
</tr>
<tr>
<td>pcC3-06</td>
<td>3 (D-29)</td>
<td><em>Craterostigma plantagineum</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pcC27-45</td>
<td>5 (D-95)</td>
<td>Brassica napus</td>
<td>A. thaliana</td>
<td>Increase in cold, dehydration and salt stress tolerance</td>
<td>Dalal et al., 2009</td>
</tr>
<tr>
<td>BhLEA1</td>
<td>4</td>
<td>Resurrection plant</td>
<td>Tobacco</td>
<td>Slow water loss, membrane and protein stabilised</td>
<td>Liu et al., 2009</td>
</tr>
<tr>
<td>BhLEA2</td>
<td></td>
<td>Boea hygrometrica</td>
<td>Tobacco</td>
<td>Slow water loss, membrane and protein stabilised</td>
<td></td>
</tr>
<tr>
<td>Atrab28</td>
<td>5 (D-34)</td>
<td>A. thaliana</td>
<td>A. thaliana</td>
<td>Increase in germination under control and high salinity conditions, transgenics survive better in 14 mM LiCl</td>
<td>Arenas-Mena et al., 1999; Borrell et al., 2002</td>
</tr>
<tr>
<td>LEA</td>
<td>5 (D-73)</td>
<td>Tamarix androssowii</td>
<td>Tobacco</td>
<td>Enhanced drought tolerance</td>
<td>Wang et al., 2006</td>
</tr>
<tr>
<td>LEA</td>
<td>5 (D-73)</td>
<td>Tamarix androssowii</td>
<td>Northland blueberry</td>
<td>Increased cold tolerance</td>
<td>Zhao et al., 2011</td>
</tr>
<tr>
<td>CaLEA6</td>
<td>5 (D-95)</td>
<td>Capsicum annuum</td>
<td>Tobacco</td>
<td>Enhanced tolerance to dehydration and salinity, but not low temperature</td>
<td>Kim et al., 2005</td>
</tr>
</tbody>
</table>
1.4 *Arabidopsis thaliana* as a model experimental plant

*A. thaliana* is a small diploid angiosperm of the mustard family (*Cruciferae* or *Brassicaceae*) that has become a model system for plant research. Its life cycle, from germination of seeds, formation of the rosette plant, bolting of the main stem, flowering and maturation of the first seeds, is completed in six weeks. This rapid growth cycle makes it very useful for research. Its flowers are 2 mm long, self-pollinated and develop into siliques that contain seeds of 0.5 mm. The rosette plant grows to 2 to 10 cm in diameter while mature plants are 15 to 20 cm in height. Several thousand seeds can be collected from a single plant (Meinke *et al.*, 1998).

The *A. thaliana* genome has been fully sequenced (The Arabidopsis Genome Initiative, 2000) and is only 120 Mb. It is organised into five chromosomes and contains an estimated 26 000 genes. It contains a relatively small amount of interspersed repetitive DNA, making sequencing of its genome a cost-effective method in identifying every gene in a representative flowering plant (Meinke *et al.*, 1998). The original idea behind using *A. thaliana* as a model system was to help in the identification of related genes important in crop plants. Information obtained from *A. thaliana* regarding mechanisms of drought, salt and cold tolerance is being applied to engineer improved commercial crops (Zhang *et al.*, 2004).

Mutants defective in almost every aspect of plant growth and development have been identified and studied by various research groups over the world. Random large-scale sequence-indexed T-DNA insertion lines have been created with gene knockouts so that reverse genetic screens can be applied to deduce the functions of the sequenced genes (Alonso *et al.*, 2003; Parinov and Sundaresan, 2000).

*A. thaliana* is a useful model plant for studying the biological function of plant genes, since it is readily transformed without the need for plant tissue culture. *Agrobacterium*-mediated transformation is performed by simply dipping the flowering plant into a solution of *Agrobacterium* and allowing it to produce seed (Clough and Bent, 1998). The target for *Agrobacterium* transformation is the ovules of young flowers. To achieve efficient transformation, the *Agrobacterium* has to be delivered to the interior of the developing gynoecium before the locules close (Desfeux *et al.*, 2000).

The *Agrobacterium*-treated T₀ plant is not treated with selection agents, such as herbicide or antibiotics. Rather, the progeny seed is harvested and selection is applied to the resultant T₁ seedlings as they germinate. Transformants derived from seed from the same seedpod will contain independent T-DNA integration events. The T₁ transformants are found to be mostly hemizygous and there is no homozygous self-fertilised offspring. Thus, the transformation
event is thought to occur in the germ-line cells after divergence of female and male gametophyte cell lineages, or the T1 embryo is transformed soon after fertilisation.

This chapter gave a brief overview of the abiotic stress responses of plants and how they frequently overlap due to similar effects on the cellular water status. It has also reviewed the literature on late embryogenesis abundant (LEA) proteins, and the studies performed to identify their role in plants’ drought tolerance. Literature on how cowpea has been used in the past as a source of drought tolerance genes was also reviewed. Some methodologies used during this study will be introduced in the following experimental chapters.
Chapter 2

Construction of a cowpea drought expression SSH cDNA library, and screening for differentially expressed drought genes using microarray technology

The majority of this chapter has been published in a peer reviewed journal, *Plant Methods* 2010, 6:10, entitled: SSHscreen and SSHdb, generic software for microarray based gene discovery: application to the stress response in cowpea.

Nanette Coetzer¹, Inge Gazendam¹²³, Dean Oelofse² and Dave K Berger³

**Author Details**

¹ACGT Computational Biology and Bioinformatics Unit, Department of Biochemistry, University of Pretoria, 0002, South Africa,
²Germplasm Development Division, Agricultural Research Council-Vegetable and Ornamental Plant Institute, Private Bag X293, Pretoria, 0001, South Africa and
³Department of Plant Science, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, 0002, South Africa

* Correspondence: dave.berger@fabi.up.ac.za
† Contributed equally

**Authors' contributions**

IG and NC contributed equally to the published article. IG performed the cowpea drought experiments, constructed the SSH library, performed the microarray hybridisations and qPCR experiments. IG performed the microarray and sequence data analysis with BLAST and with the SSHdb software, developed by NC and DKB, and drafted the manuscript. NC developed the SSHscreen and SSHdb software, analysed the microarray and sequence data and drafted the manuscript. DO contributed to conceiving the study, the design of the biological experiments and assisted in writing the manuscript. DKB was instrumental in the design of the study, development of the software, interpretation of data, and drafted the manuscript. All authors have read and approved the final manuscript.
2.1 Abstract
Suppression subtractive hybridisation (SSH) is a popular technique for gene discovery from non-model organisms without an annotated genome sequence, such as cowpea (*Vigna unguiculata* (L.) Walp). We used this method to enrich for genes expressed during drought stress in a drought tolerant cowpea line. Forward and reverse cDNA libraries enriched for cowpea drought response genes were screened on microarrays, and the R software package SSHscreen 2.0.1 was used (i) to normalise the data effectively using spike-in control spot normalisation, and (ii) to select clones for sequencing based on the calculation of enrichment ratios with associated statistics. Enrichment ratio 3 values for each clone showed that 62% of the forward library and 34% of the reverse library clones were significantly differentially expressed by drought stress (adjusted p-value < 0.05). Enrichment ratio 2 calculations showed that > 88% of the clones in both libraries were derived from rare transcripts in the original tester samples, thus supporting the notion that SSH enriches for rare transcripts. A set of 118 clones were chosen for sequencing. SSHdb, a web-accessible database, was used to manage the clone sequences and combine the SSHscreen data with sequence annotations derived from BLAST and Blast2GO. SSHscreen plots are a useful tool for choosing anonymous clones for sequencing, since redundant clones cluster together on the enrichment ratio plots. Various drought-induced cowpea genes were identified in the library, the most interesting encoding a late embryogenesis abundant LEA5 protein, a glutathione S-transferase, a thaumatin, a universal stress protein, and a wound induced protein. A lipid transfer protein and several components of photosynthesis were down-regulated by the drought stress. Reverse transcriptase quantitative PCR confirmed the enrichment ratio values for the selected cowpea genes. SSH, in combination with SSHscreen and SSHdb, is therefore a popular and powerful gene discovery tool in non-model organisms, since it improves the selection of clones for sequencing after screening the library on a small number of microarrays.

2.2 Introduction
A range of techniques are available for gene discovery. Expressed sequence tag (EST) sequencing of cloned cDNAs is a common approach with the advantage that if full-length cDNAs are cloned they can be directly employed for further gene function experiments (Ralph *et al.*, 2008). Cloned cDNAs can be arrayed on high-density microarrays and used for expression profiling (Chen *et al.*, 1998). Next generation sequencing, such as 454 technology™, has been employed for sequencing cDNA libraries (Cheung *et al.*, 2006), and the term RNA-Seq has been dubbed for this approach when applied at deep enough coverage to compare transcript counts between one or more biological states (Wang *et al.*, 2009).
2009). Previous methods, such as serial analysis of gene expression (SAGE), are also based on counting short sequence tags (Velculescu et al., 1995). Although these methods provide exceptional quantitative analysis, they are labour-intensive and currently very costly. Additionally, they are most effective if an annotated genome sequence is available.

Many research laboratories that are investigating non-model crops without genome sequence resources or have research questions that do not require a full genome analysis have the option of applying different "RNA fingerprinting" techniques for gene discovery. Examples of these techniques are differential display RT-PCR (DD-RT-PCR), RNA-fingerprinting by arbitrarily primed PCR (RAP-PCR) and cDNA amplified fragment length polymorphism (cDNA-AFLP) where cDNA sub-populations are amplified and visualised on polyacrylamide gels, whereafter differentially expressed transcripts are isolated from the gel for sequencing (Liang et al., 1992; Welsh et al., 1992; Bachem et al., 1998). These methods have limitations, such as bias based on choice of initial primer sets, problems with reproducibility, generation of false positives, and reliance on time-consuming polyacrylamide gel electrophoresis and gel extraction to obtain sequence information. Another limitation of the above methods is the difficulty to capture low abundance clones.

A third alternative for gene discovery are PCR-based cDNA subtractive hybridisation methods. These methods exclude common cDNA sequences between the two or more samples and, thus, enrich for target sequences of interest, which are subsequently cloned. These methods include representational difference analysis (RDA) and SSH (O’Neill et al., 1997; Diatchenko et al., 1996; Morissette et al., 2008). SSH is considered an effective method to enrich for rare transcripts (Diatchenko et al., 1996). A recent search with the keywords ‘suppression subtractive hybridisation’ in the title of research articles in PubMed produced 1213 hits (data not shown), indicating that SSH remains a popular method for the construction of enriched cDNA libraries. We chose to apply SSH to gene discovery in the non-model crop cowpea (Vigna unguiculata (L.) Walp.), and the paper by Coetzer et al. (2010) describes the two software innovations that facilitate gene discovery using SSH.

Cowpea is a tropical legume crop with a high protein content, since it is able to fix nitrogen, and is used as a protein substitute for meat products (Singh et al., 2003). The crop is fully utilised by people in Africa as leaves and seeds are consumed, and the plants are used for grazing and the feeding of livestock. Since many lines are drought tolerant, cowpea can be grown under the harshest growing conditions, and in the poorest soils, and is, therefore, an important crop for subsistence and small-holder farmers (Quass, 1995). Breeding efforts to improve yield of cowpea under different production systems is ongoing (Singh et al., 2003),
and lines with differential drought tolerance have been identified (Dingkuhn et al., 2006; Agbicodo et al., 2009; Muchero et al., 2008). Promising QTLs for drought tolerance in cowpea have recently been reported (Muchero et al., 2009).

Cowpea can be classified as an orphan crop, which means that it is important for food security in many developing countries, however, limited research funding has been devoted to it (Varshney et al., 2009). Genomics resources for cowpea are starting to be developed with sequencing of a methyl-filtered genomic library (Timko et al., 2008), as well as an EST dataset (Varshney et al., 2009). The availability of a cowpea breeding line that exhibited drought tolerance in the field prompted us to investigate gene expression in this line in response to drought stress. Based on previous experience of using SSH for gene discovery in other orphan crops, banana and pearl millet (Van den Berg et al., 2004; Crampton et al., 2009), we encountered bottlenecks in the process. Consequently, in this study we used an improved gene discovery pipeline (Coetzer et al., 2010), through the software SSHscreen 2.0.1, an R package, which quantitatively describes each clone in the library in terms of up/down regulation and rarity/abundance in the treated sample. We then validated the enrichment ratio calculations from the microarray screening and SSHscreen 2.0.1 analysis for selected drought-responsive cowpea clones using quantitative PCR (qPCR). SSHscreen facilitated the efficient choice of clones to be sequenced, while a web-based sequence database, SSHdb, facilitated the management and annotation of the SSH cDNA library clones.

2.3 Materials and Methods

2.3.1 Plant materials and treatments

Cowpea (V. unguiculata L. Walp) breeding lines IT96D-602 and Tvu7778 were provided by the Dr. BB Singh of the International Institute of Tropical Agriculture (IITA) (Dingkuhn et al., 2006). Seeds were germinated and plants were grown in a glasshouse under 11 h day length, 28°C and 18°C day and night temperatures, respectively, and watering three times weekly. At six weeks, five replicate plants of each variety were divided into two groups. One group was subjected to drought stress by withholding water, and the other group was kept to the control watering scheme.

2.3.2 RNA extraction

RNA was isolated from cowpea leaves using Tri-reagent (Sigma Aldrich) and Polyvinyl pyrrolidone (PVP) (Ambion’s Plant RNA isolation aid). Contaminating genomic DNA was removed with the Turbo DNA-free kit (Ambion) and the RNA cleaned up with the Plant
RNeasy kit (Qiagen, Hilden, Germany). Formaldehyde agarose gel (1.2%) electrophoresis of RNA was performed using standard protocols (running buffer composition: DEPC-treated dH2O containing 20 mM MOPS, 5 mM Sodium Acetate, 1 mM EDTA, 0.7% Formaldehyde; pH 7.0). Genomic DNA contamination was checked with RT-PCR using the bean Actin primers (Appendix 1) and the C-therm Polymerase One-Step RT-PCR system (Roche). PCR was performed using GoTaq DNA polymerase (Promega). The RT-PCR reaction was incubated at 60°C for 30 min before cycling. The primer annealing temperature for both RT-PCR and PCR was 56°C, and 30 cycles were included. All amplification products were analysed on a 2% agarose/ 0.5x TAE gel.

2.3.3 Construction of cDNA library using SSH

Differential expression analysis by means of SSH (Diatchenko et al., 1996) was employed to prepare a cDNA drought expression library for cowpea. Messenger RNA (mRNA) was isolated from 50 µg pools of stressed IT96D-602 RNA (9 and 12 days without water) (treated) and unstressed Tvu7778 RNA (9 and 12 days) (control) using an Oligotex mRNA purification kit (Qiagen). cDNA was synthesised from each mRNA using the cDNA synthesis system (Roche Diagnostics, Basel, Switzerland) to be used as unsubtracted testers and unsubtracted drivers in SSH (Van den Berg et al., 2004). Subtractive hybridisation was performed on Rsal (Roche Diagnostics) -digested tester and driver cDNA fragments using the PCR-Select cDNA subtraction kit (BD Biosciences Clontech, Palo Alto, CA), as previously described (Van den Berg et al., 2004). A forward subtraction was performed by using the treated sample as tester (and control as driver), and a separate reverse subtraction was performed by using the control cDNA as tester (and treated cDNA as driver). After subtraction the products were amplified by a primary PCR and a nested secondary suppression PCR to generate differentially expressed cDNA fragments (termed ST (subtracted treated) and SC (subtracted control) for the forward and reverse libraries, respectively). Replicate PCR reactions were pooled, size fractionated and cloned into the pGEM-T Easy cloning vector and transformed into Escherichia coli JM109 following the manufacturers' instructions (Promega, Madison, WI). Transformed colonies were selected by blue-white selection on 100 µg/ml ampicillin LB-agar selection media (spread with X-Gal and IPTG) and stored as 25% glycerol stocks at -70°C in sterile 96-well culture plates (Corning, NY). In addition, unsubtracted PCR products from the treated cDNA (drought stressed IT96D-602) (termed UT) and control cDNA (control Tvu7778) (termed UC) were also prepared to be used for SSHscreen analysis as described in (Berger et al., 2007).
2.3.4 Fabrication of SSH library on glass slide array

Inserts of the cowpea drought expression cDNA library were amplified with PCR directly from overnight bacterial cultures in 96-well format (Thermo-Fast, ABGene, Epsom, UK) in 100 µl reactions with 1 U Biotaq DNA polymerase (Bioline) and the SP6 and T7 primers (Appendix 1). The PCR plate was sealed with a silicon mat (Corning). Reactions were incubated in a PTC-100 Thermocycler (MJ Research) at 94°C for 5 min; 30 cycles of (94°C for 30 s, 50°C for 30 s and 72°C for 1 min); and 72°C for 5 min.

The PCR products were purified with Montage PCR purification plates on a vacuum manifold (Microsep) and resuspended in 50 µl sterile distilled water (SDW). The suspensions were transferred to 96-well storage plates, covered with well caps (Nunc, Roskilde, Denmark) and stored at -20°C. The purified PCR products were dried down in a vacuum centrifuge at 45°C, resuspended in 50% dimethyl sulfoxide (DMSO), transferred to 384-well spotting plates and stored at -70°C until microarray spotting.

The control genes *gfp* (717 bp fragment in pGEM-T Easy, positions 1603-2319 of GenBank accession number AF078810), *globin* (human beta-globin; 474 bp fragment in pBluescriptSK, positions 50-523 of NM_000518) and *nptII* (812 bp fragment in pGEM-T Easy, positions 142-953 of V00618) were purchased from the Nottingham Arabidopsis stock centre (NASC, http://arabidopsis.org.uk). They were transformed into *E. coli* JM109 (Promega). An *its* clone in pGEM-T Easy (193 bp fragment from the internal transcribed spacer 2 of the rRNA genes from *Leptographium elegans*) was also used as a control gene. It matches to positions 268-458 of AF343675.1. Plasmids were isolated from cultures using the Qiaspin miniprep plasmid isolation kit (Qiagen). PCR products of the four control genes were prepared using the T7 and SP6 primers (PCR product sizes: *gfp* (893 bp), *nptII* (988 bp), *its* (369 bp)) or the T7 and M13-26-R primers (Appendix 1) for *globin* (677 bp). Montage purified PCR products of twelve 100 µl PCR reactions each were pooled, concentrated and transferred to 12 wells each of a 384-well spotting plate. An equal volume of DMSO was added so that the final concentrations in 50% DMSO ranged from 70-100 ng/µl. Five two-fold serial dilutions were also prepared for each PCR fragment (*gfp*: 180-11.25 ng/µl; *globin*: 100-6.25 ng/µl; *nptII*: 150-9.375 ng/µl; and *its*: 130-8.125 ng/µl), transferred to an additional 10 wells per fragment, an equal volume of DMSO added, and spotted on the glass slides.

Glass slides (Corning GAPS II) were spotted with the cowpea drought expression library (4160 clones in total from the forward and reverse libraries) and controls using the Array Spotter Generation III (Molecular Dynamics, Sunnyvale, CA) at the University of Pretoria, Pretoria, South Africa http://microarray.up.ac.za. Each sample spot was duplicated on the
slide in replicate blocks on either side of the slide, therefore, replicates are not spatially close together. The slides were allowed to dry overnight in the protective atmosphere of the spotter, after which the DNA was cross-linked under ultraviolet (UV) light for 3 min. The slides were stored in a desiccator covered in foil at room temperature.

2.3.5 Screening SSH library on microarrays

SSH cDNA fragments (ST, SC, UT and UC), purified by PCR Minelute cleanup kit (Qiagen), were digested with Rsal (10 U per microgram DNA) in the appropriate buffer overnight at 37°C. The fragments were separated from the adaptor fragments by electrophoresis on a 1.5% low melting point agarose gel (Seaplaque, FMC Bioproducts) in 0.5× TAE and purified from the gel using the Qiaquick gel extraction kit (Qiagen).

The control fragments were excised from their plasmids using restriction digestion to exclude any T7 and SP6 primer binding sites (KpnI/XbaI for globin (product of 548 bp); Ncol/PstI for gfp (768 bp); EcoRI for nptII (830 bp) and its (211 bp)). Restriction fragments were purified with the Qiaquick gel extraction kit (Qiagen). Each target sample of SSH cDNA fragments (200 ng) were spiked with equal amounts of a control fragment pool made up of different quantities of four control fragments (45 ng globin, 45 ng its, 4.5 ng nptII and 0.45 ng gfp) for within-slide normalisation. Spiking with equal amounts of fifteen- or three-fold dilutions of the control fragment pool were tested and also gave sufficient hybridisation for within-slide normalisation (data not shown).

Targets were labelled by direct Cyanine™-dUTP incorporation by Klenow enzyme (Fermentas, Vilnius, Lithuania). Each SSH fragment sample was labelled with both dyes (Cy3 and Cy5) for a dye-swap experiment of each slide. The protocols and data analysis techniques described in (Berger et al., 2007) were followed, with some modifications. DNA to be labelled, in a volume of 12 µl, was denatured at 95°C for 5 min and placed on ice. The following were added to the pairs of denatured DNA samples to yield a total reaction volume of 20 µl: 2 µl of 10× Klenow buffer (Fermentas); 2 µl 10× Hexanucleotide mix (Roche Diagnostics); 2 µl Klenow enzyme (5 U/µl; Fermentas); 2 µl of a dNTP mix containing 1 nmol each of dATP, dCTP and dGTP, 0.74 nmol dTTP and 0.27 nmol of either Cy3-dUTP or Cy5-dUTP (Amersham Biosciences). The labelling reaction was incubated overnight (17-20 h) at 37°C. The labelled DNA was cleaned up from unincorporated dye using the Qiaquick PCR purification kit (Qiagen). Dye incorporation was measured using a NanoDrop ND-1000 UV-Vis Spectrophotometer (Nanodrop Technologies, Wilmington, DE).
Labelled SSH targets were combined in pairs using equal amounts of Cy3 and Cy5 dye incorporation for each target in each pair required for SSHscreen analysis. Each labelled target DNA mix was dried down in a vacuum centrifuge at 45°C and resuspended in 50 µl hybridisation solution (50% formamide, 25% 4× Microarray hybridisation buffer (Amersham Biosciences), 25% SDW). Labelled targets in hybridisation solution were denatured at 95°C for 2 min and placed on ice.

Glass slides arrayed with the SSH cDNA libraries were pre-treated in 1% bovine serum albumin (BSA; Roche Diagnostics), 3.5× SSC (525 mM sodium chloride and 52.5 mM sodium citrate) and 0.2% sodium dodecyl sulphate (SDS) at 60°C for 20 min. After rinsing in SDW at room temperature, the slide was dried by centrifugation in a 50 ml tube at 1000 × g for 4 min at room temperature in a swing-out rotor (Eppendorf 5810R centrifuge). The slide was placed in a locally manufactured hybridisation chamber (HybUP, NB Engineering, Pretoria, South Africa) with 20 µl SDW in the reservoirs on either side. Labelled and denatured target was applied to the slide and gently overlaid with a cover slip. The chamber was sealed and incubated in a water bath at 42°C for 16 h. Slides were washed for 4 min at 42°C with 1× SSC (150 mM NaCl, 15 mM sodium citrate)/0.2% SDS, twice with 0.1× SSC (15 mM NaCl, 1.5 mM sodium citrate)/0.2% SDS and three washes of 0.1× SSC for 1 min at room temperature. After dipping the slide in SDW at room temperature and centrifuged to dry, it was immediately scanned using a GenePixTM 4000B scanner (Axon Instruments, Foster City, CA).

GenePix Pro 5.1 software (Axon Instruments) was used to automatically locate all the spot positions from the scanner-generated TIFF images and associate them with each specific clone in a GenePix Array List (GAL file) (available at NCBI GEO Accession # GSE20273). The GAL file links the information from the arraying process to the analysis, since it provides identification information for each spot printed on the slide. Bad quality spots (irregularly shaped or with hybridisation artifacts; signal/noise ratio < 3) were flagged for exclusion during data analysis and the array of circles were manually adjusted for a better fit. GenePix Pro 5.1 was used to extract the dye intensity data of each spot and save the data for each slide in a GenePix Results file (gpr).

### 2.3.6 SSHscreen software analysis of microarray data

The SSHscreen 2.0.1 package, written as a single function in the R programming language, was used for analysing the resulting microarray data to calculate ER3 and inverse ER2 values (http://microarray.up.ac.za/SSHscreen/; Coetzer et al., 2010). ER3 values (log2(UT/UC) for the forward library clones; and log2(UC/UT) for the reverse library clones)
quantify differential expression of transcripts that give rise to the clones in each library (Berger et al., 2007). Inverse ER2 values \( \log_2(\text{UT/ST}) \) for the forward library clones; and \( \log_2(\text{UC/SC}) \) for the reverse library clones) reflect the relative abundance of transcripts for each clone in the unsubtracted samples. The SSHscreen analysis implemented by the R script (provided in Appendix 2.1) is described in the following sections:

### 2.3.6.1 SSHscreen pre-processing

SSHscreen analysis in R (version 2.8.1) required the libraries of limma (version 2.16.5) and SSHscreen 2.0.1 to be installed in R. The input data to SSHscreen 2.0.1 were the 12 Genepix results files from hybridisation experiments to the cowpea SSH library arrays (gpr files deposited at NCBI GEO series accession number GSE20273), with the Targets file, Spot types file (Appendices 2.2 and 2.3, respectively) and the GAL file (available at GEO accession number GSE20273).

The first step was to weight all spots that had been flagged as poor quality (signal/noise < 3) by the GenePix Pro 5.1 image analysis programme (http://www.axon.com) so that these spots would not be used to calculate the normalisation factors. Background correction, within array normalisation and between-slide normalisation were implemented as described in Coetzer et al. (2010). Within array normalisation was based on data from the four alien controls (globin, its, nptII and gfp) that had been spotted as dilution series on each array to make up a total of 176 control spots/array. Equal amounts of each control fragment had been added to pairs of target samples to be labelled with the Cy3 and Cy5 dyes, and thus hybridisation signals from the control spots could be used for within-slide normalisation. MA plots (Figure 2.6) were exported at each stage of pre-processing.

### 2.3.6.2 SSHscreen enrichment ratio analysis and outputs

ER3 and inverse ER2 values for each SSH library clone were calculated in SSHscreen using the functions of limma for differential gene expression analysis (Smyth, 2004). ER3 values were based on testing the null hypothesis that there is no differential expression for a gene between the UT and UC samples, whereas inverse ER2 values test the null hypothesis that there is no difference in abundance of a clone between the unsubtracted sample (e.g. UT) and its corresponding subtracted sample (e.g. ST). This was achieved in SSHscreen 2.0.1 by implementing the limma function that fits gene-wise linear models through the normalised expression data that was the output of pre-processing. Thereafter, an empirical Bayes approach is used in limma to calculate a moderated t-statistic for each gene, in which the standard errors have been moderated across all the genes on the array. This approach of variance shrinkage improves inference about each gene in experiments in which there are a
low number of replicates (Smyth, 2004). SSHscreen adjusted for multiple testing using Benjamini and Hochberg's method (1995) for controlling the false discovery rate, which computes an adjusted p-value for the hypothesis test of each gene. A prior guess of 50% differentially expressed genes for the SSH libraries was implemented as the default in SSHscreen 2.0.1, and was used in calculating B-statistics for each gene using limma functions. The B-statistic (Lonnstedt and Speed, 2002; Smyth, 2004) can be interpreted as the log-odds that a specific gene is differentially expressed. Thus a positive B-statistic represents more than a 50-50 chance of differential expression. The outputs of SSHscreen were top tables which reported the enrichment ratios for each gene and associated statistics, namely the moderated t-statistic, associated p-value adjusted for multiple testing, and B-statistic (Tables 2.1 and 2.2), MA-plots (Figure 2.6) and a graphical representation of each clone on ER-plots (Figures 2.7 A and B), which were used to select clones for sequencing.

2.3.7 Sequencing
Selected cowpea drought expression library clones were sequenced using the T7 Promoter primer by Inqaba Biotec (SA) or Macrogen (USA). Colonies were sent on LB-agar plates containing 100 µg/ml ampicillin. Non-redundant sequences have been deposited in dbEST at GenBank (Accession numbers GR942571 - GR942610).

2.3.8 Annotation and management of sequences using SSHdb
SSHdb (http://sshdb.bi.up.ac.za) was used as a web-based tool for sequence management of clones in SSH libraries (Coetzer et al., 2010). For each input sequence in FASTA format, SSHdb removed the vector and adaptor fragments after BLASTN (Altschul et al., 1990) searches were performed against the NCBI UniVec database (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html). Further BLASTN searches were carried out against all sequences already uploaded in the database, so that redundant partners in the library (using a BLASTN E-value cut-off value of 10e-10) could be identified. For each redundant partner group, the longest sequence in the group was selected by default as the representative clone. Multiple sequence alignments (generated by ClustalW (Thompson et al., 1994)) for individual redundant partner groups could be viewed and downloaded from SSHdb. For each representative clone, SSHdb performed nucleotide-nucleotide and translated sequence comparisons using BLASTN and BLASTX searches against a local installation of the NCBI non-redundant nucleotide and peptide databases (nt/nr) (Altschul et al., 1990). For cases where the E-value of the top BLASTX hit was less than 10e-10, this hit was automatically selected as the default priority annotation. The top 10 BLASTX and BLASTN hits were stored in the database. SSHdb also implements Blast2GO (Conesa et al., 2005), and the putative GO annotations for representative sequences are
recorded in SSHdb. SSHdb provides two major views of the data, the SSH database view, which shows the annotated representative clones, or the SSH toptable view, which shows the enrichment ratio data for each clone in each library (see Table 2.1). SSHdb can be updated as additional clones are sequenced.

2.3.9 Quantitative PCR
QPCR primer pairs (20-mers) were designed from selected cowpea sequences to amplify products between 120 and 250 bp in length from the SSH cDNA fragment pools (UT, UC, ST and SC) (Appendix 1). QPCR reactions containing 1× Sensimix (Quantace, Celtic Molecular Diagnostics), SYBR Green, 2.5 mM MgCl₂, the appropriate primer pair (200 nM each) and cDNA template in a total volume of 25 µl were set up and run on a Rotor-Gene 3000A (Corbett Research). The enzyme was activated by a hold at 95°C for 10 min followed by 45 cycles of 95°C for 15 s, primer annealing at 56°C for 30 s and extension at 72°C for 6 s. SYBR green fluorescence was measured after the extension step of every cycle. QPCR was performed on serial ten-fold dilutions of a mix of UT and UC cDNAs (templates ranging from 0.5 pg to 50 ng) to construct standard curves for each primer pair. The quantification cycle (Cq) values from the QPCR fluorescent profiles were converted to input nanograms of template using the standard curves. Average nanogram quantities for each gene was normalised relative to the data for the respective sample’s reference gene content.

For ER3 verification, QPCR was performed in duplicate on 50 ng each of cDNA from the two cowpea lines before subtraction (UT and UC). Glyceraldehyde-3-phosphate dehydrogenase C-subunit (gapC) was used as a reference gene. A consensus sequence between the gapC genes of Medicago truncatula [GenBank: AC135505_Mt, exons only], Glycine max [GenBank: DQ192668_Gmax1 and DQ355800_Gmax2] and Pisum sativum [GenBank: PEA-GAPCI] was used to design the GAPC reference gene primers (Appendix 1). An expression ratio of log₂(ng in UT/ng in UC) was calculated for each gene.

For ER2 verification, normalisation of QPCR results between unsubtracted and subtracted cDNA samples (i.e. UT and ST; UC and SC) required the spiking of cDNA samples with equal amounts of an alien gene. QPCR was performed in duplicate on 10 ng of the UT, ST; UC and SC cDNA templates, each spiked with 50 pg of the human beta-globin fragment (prepared using the Globin forward and reverse primers, Appendix 1). Average input nanogram quantities were calculated for each gene and normalised with the globin spike content. Expression ratios of log₂(ng in UT/ng in ST) and log₂(ng in UC/ng in SC) were calculated for each gene.
Reverse transcriptase quantitative PCR (RT-qPCR) reactions were performed in duplicate using 100 ng of RNA isolated prior to SSH library construction from drought-treated IT96D-602 cultivar and control Tvu7778 at two time points, 9 and 12 days, separately. The Sensimix One-step RT-qPCR kit (Quantace) was used with a reverse transcription step at 49°C for 30 min inserted before the cycling profile described above. The quantification cycle ($C_q$) values from the RT-qPCR fluorescent profiles were converted to input nanograms of template using the standard curves.

2.4 Results

2.4.1 Construction of cowpea drought expression SSH library

A cowpea (V. unguiculata, (L.) Walp) drought expression library was generated where the objective was to identify and isolate genes responding to drought stress in cowpea. SSH (Diatchenko et al., 1996) was used to enrich for genes that were differentially expressed between drought stressed and unstressed cowpea plants. Cowpea breeding lines from the International Institute of Tropical Agriculture (IITA) that were previously shown to be drought tolerant (line IT96D-602) and drought susceptible (line Tvu7778) were used (Spreeth et al., 2004). Since these two lines are not genetically similar, genetic differences between the two lines differing in drought tolerance were included in the study. The "tester" for forward library construction was from drought stressed line IT96D-602 (cDNA pooled from plants 9 and 12 days after water was withheld), whereas the "driver" for forward library construction was from control treated line Tvu7778 (cDNA pooled from plants grown for the same time on a normal watering regime). These time points were chosen for maximum drought stress symptoms, before leaves were too senesced for RNA extractions. The aim of this wide subtraction was to be sure to capture sufficient differentially expressed transcripts in the library before testing their involvement in drought responses. These could include not only genes that are induced/repressed by drought stress in drought tolerant IT96D-602 only, but also those that are constitutively expressed at higher/lower levels in IT96D-602 compared to the drought susceptible line Tvu7778.

Formaldehyde-agarose electrophoresis showed the RNA isolated from the cowpea lines to be of good quality (Figure 2.1). Genomic DNA seems to be present (Figure 2.1, lanes 1 to 4), which was removed by DNase-treatment and cleanup (Figure 2.1, lanes 5 and 6, respectively).
Figure 2.1 Formaldehyde agarose (FA) gel electrophoresis of cowpea RNA isolated using Tri-reagent. M: High range RNA ladder (Fermentas SM0423; 200-6000 bp); lanes 1+2: 4 µg Tvu7778 RNA of day 9 and 12, respectively; lanes 3+4: 4 µg IT96D RNA of day 9 and 12, respectively; lane 5: 1 µg Tvu7778 RNA cleaned up with the Plant RNeasy kit (Qiagen); lane 6: 1 µg Tvu7778 RNA treated with the Turbo DNA-free kit (Ambion).

PCR using RNA samples as templates indicated genomic DNA contamination (results not shown). RNA was, therefore, DNase-treated and retested with RT-PCR and PCR for the successful removal of gDNA. A 170 bp fragment is expected with the bean Actin primers from cDNA and 280 bp from gDNA (Appendix 1). Non-specific amplification products of ~500 and ~700 bp were obtained as well. RT-PCR of DNase-treated RNA samples using the bean Actin primers lacked the contaminating 280 bp fragment caused by gDNA contamination (Figure 2.2, lanes 3 and 4 compared to lanes 1 and 2). The cDNA fragment of 170 bp was successfully amplified (Figure 2.2, lanes 3 and 4, indicated by an arrow). The reactions also had less background with the ~500 bp fragment fainter. PCR of DNase-treated RNA samples yielded no visible amplified products (Figure 2.2, lanes 9 and 10), whereas the untreated RNA was contaminated with gDNA that served as template for the amplification of three products (Figure 2.2, lanes 7 and 8). DNase treatment, therefore, successfully removed contaminating gDNA to the level that products were not obtained from PCR with ~400 ng treated RNA as template (Figure 2.2, lanes 9 and 10). DNase treatment of crude RNA (immediately after Tri-reagent isolation) (Figure 2.2, lanes 3 and 9) seems just as efficient as DNase treatment of RNA that has been cleaned up with the RNeasy kit (Figure 2.2, lanes 4 and 10).
Figure 2.2 RT-PCR and PCR of the actin gene using DNase treated and untreated cowpea Tvu7778 RNA as templates. M: Low Range Massruler DNA ladder (Fermentas SM0388; 80-1031 bp); lanes 1-5: RT-PCR; lanes 6-11: PCR. Templates: lanes 1+7: 0.5 µg crude RNA; lanes 2+8: ~0.5 µg RNA cleaned up with the RNeasy kit (Qiagen); lanes 3+9: ~400 ng crude RNA treated with the Turbo DNA-free kit (Ambion); lanes 4+10: ~400 ng RNA treated with the Turbo DNA-free kit and cleaned up with the RNeasy kit; lane 6: 40 ng gDNA; lanes 5+11: negative water controls.

The SSH procedure was performed on cDNA prepared from stressed IT96D-602 RNA (treated) and unstressed Tvu7778 RNA (control). The number of primary PCR cycles were optimised (results not shown), followed by secondary PCR using nested adaptor primers from the PCR-Select cDNA subtraction kit. Good quality forward and reverse subtracted cDNA fragments were generated (Figure 2.3). Distinct patterns of PCR products were obtained from six individual nested secondary PCRs for each of the forward and reverse subtractions (Figure 2.3, lanes 1-12). The secondary PCRs were performed using separate primary PCRs as templates. Discrete bands were observed in the subtracted samples (Figure 2.3, lanes 1-12, pooled together in lane 16 and 17), giving an indication that the SSH was successful, since they represent gene fragments that have been enriched. The subtracted treated (ST, Figure 2.3, lane 16) refers to the forward subtraction and the subtracted control (SC, Figure 2.3, lane 17) to the reverse subtraction. The unsubtracted treated (UT) and control (UC) PCR yielded smears, as expected (Figure 2.3, lanes 14-15), since the two primers have annealing sites on the adapters that are present on both ends of these cDNA fragments.
Figure 2.3 Secondary PCR amplification of subtracted and unsubtracted cDNA fragments produced from drought stressed IT96D-602 and control Tvu7778 cowpea lines that had been subtracted with suppression subtractive hybridisation. M: Molecular weight marker VI (Roche; 154-2176 bp). The secondary PCR products were loaded as follows: Lanes 1-6: six individual subtracted treated (ST) samples; Lanes 7-12: six individual subtracted control (SC) samples; Lane 13: negative water control; Lanes 14-15: 4.5 µl sample of pooled unsubtracted treated (UT) and unsubtracted control (UC), respectively; Lane 16-17: 4.5 µl sample of pooled subtracted treated (ST) and subtracted control (SC), respectively.

The subtracted cDNA fragments were used to construct a cDNA library with a total of 4160 cDNA clones (2144 in the forward and 2016 in the reverse library), which were amplified by PCR and spotted in duplicate onto glass slides for screening and selection of clones for sequencing.

Subtracted and unsubtracted cDNA samples from cowpea, used to construct the SSH libraries, were prepared as Cy3- and Cy5-labelled targets and hybridised to the microarrays. These cDNA samples were UT (unsubtracted treated), UC (unsubtracted control), ST (subtracted treated, i.e. forward library), and SC (subtracted control, i.e. reverse library). High quality microarray images were obtained from hybridisation of pairs of Cyanine™-labelled cDNA targets (UT, UC, ST or SC) to the cowpea drought expression microarrays. The average numbers of spots across the 12 arrays with signal-to-noise ratios > 3 were 83% and 85% for the Cy3 and Cy5 channels, respectively. The average coefficients of variance (%CV) (standard deviation*100/mean intensity) for the background values across the 12 arrays were...
7 and 12% for the Cy3 and Cy5 channels, respectively. This is supported by visual inspection of pseudocolour images of slides. For example, strong hybridisation of Cy3 targets from ST to probes from the forward library spotted in the top six rows of each array block can be observed as green spots in Figure 2.4 A, whereas Cy5 targets from UT hybridise predominantly to probes from the reverse library as red spots (rows 7-11 of each array block) (Figure 2.4 A). The opposite hybridisation pattern is observed in a dye swap slide (Figure 2.4 B), as expected.

**Figure 2.4 Example of Microarray pseudocolour images following hybridisation.** A: Example of a cowpea microarray image following hybridisation with differentially labelled cDNA samples, and scanning with a GenePixTM 4000B scanner (Axon Instruments). In this particular example, subtracted treated (ST) (cDNA prepared from pooled RNA extracted from IT96D-602 cowpea plants drought stressed for 9 and 12 days, and subtracted with cDNA prepared from RNA isolated from control Tvu7778 plants) was labelled with Cyanine™-3 dye, (green pseudocolour). Unsubtracted treated cDNA (prepared from pooled RNA extracted from IT96D-602 cowpea plants drought stressed for 9 and 12 days) was labelled with Cyanine™-5 dye (red pseudocolour). B: Dye swap of the experiment in Figure 2.4 A. Subtracted treated (ST) cDNA was labelled with Cyanine™-5 dye, and unsubtracted treated cDNA was labelled with Cyanine™-3 dye. These differentially labelled cDNA samples were hybridised to the cowpea microarray slide and scanned with a GenePixTM 4000B scanner (Axon Instruments).
2.4.2 Screening the cowpea SSH libraries using SSHscreen 2.0.1

SSHscreen version 2.0.1 ([http://microarray.up.ac.za/SSHscreen/](http://microarray.up.ac.za/SSHscreen/)) was used to analyse the resulting microarray data using limma functions, thereby quantitatively screening the library for significantly differentially expressed clones (Berger et al., 2007). SSHscreen analysis of the microarray data was used to assist in selection of 118 clones for sequencing, based on their statistics of differential expression.

Within-slide normalisation of two-colour microarray data is an important consideration to account for systematic bias due to differences between the Cy3 and Cy5 dyes (Smyth and Speed, 2003). Commonly, loess normalisation is applied (Smyth and Speed, 2003), however, this is based on the assumption that most of the genes on the array are not differentially expressed. This is legitimate for most whole genome microarray experiments, however, it is not appropriate when the array is constructed from an SSH library, which selects for differentially expressed genes. Therefore, spike-in control spot-based normalisation was applied in SSHscreen analysis of the cowpea SSH libraries (Smyth and Speed, 2003). Serial dilutions of four "alien" control probes (green fluorescent protein (*gfp*), human beta-globin (*globin*), bacterial neomycin phosphotransferase II (*nptII*) and a fungal rRNA gene internal transcribed spacer (*its*); see Materials and Methods) were spotted on the glass slides. These probes were chosen since matching sequences were unlikely to be present in the cowpea cDNA samples. Importantly, a "spike-in" control mix of fragments of the genes corresponding to the four control probes *gfp*, *globin*, *nptII* and *its* was prepared in which each of the four genes was present at a different concentration. The spike-in control mix was added in equal amounts to each cDNA target sample prior to labelling. The dilution series of control spots on each array which have hybridised to the spike-in controls (added in equal amounts to the pairs of target cDNAs) can be observed in the raw pseudocolour images as yellow spots in row 12 of most array blocks (Figure 2.4).

Within-slide normalisation using the spike-in control spots was effective in our study, and this was illustrated by boxplots of the control spots across the 12 slides, which showed that the variability of M values in the raw data was diminished considerably by the normalisation (Figure 2.5). The average standard deviations of the M values for the control spots across the 12 slides decreased from 0.18 to 0.10 after normalisation, a similar improvement to that reported in Figure 2 of Fardin et al. (2007), who also applied control spot-based normalisation. This can also be visualised in the MA plots (Figure 2.6), since the control spots (colours other than blue or yellow) were placed on M = 0 line in the MA-plots after normalisation. Clones of the forward and reverse libraries were illustrated by blue and yellow dots in Figure 2.6 panels a-h and i-p, respectively. Dye swap slides showed consistent...
clouds of data points above and below the $M = 0$ line, as expected (compare panel a with b, for example). Effective normalisation would also be expected to decrease the variation of the $M$ values for the clones. The average standard deviations of individual clone $M$ values decreased by 37% and 8% for the ER3 and ER2 slides, respectively. We also calculated the consistency of clone spot intensities across pairs of replicate slides, and Pearson's correlation coefficients ranged from 0.78 (Figure 2.6 i vs. k) to 0.96 (Figure 2.6 f vs. h) with an average of 0.85 (data not shown).

![Figure 2.5 Box plots of M-values of control spots before and after normalisation for all 12 slides used for SSHscreen analysis. Each box corresponds to one array. A: Box plots before normalisation (i.e. only background subtracted M-values of raw data). B: Box plots after within and between slide normalisation (i.e. spike-in control spot loess and A-quantile normalised M-values).](image)
Figure 2.6 MA plots after normalisation of the forward and reverse cowpea SSH libraries. M versus A plots for microarray slides after within and between slide normalisation in SSHscreen 2.0.1. For each comparison of interest, there were four technical replicates of which two were dye-swaps: plots a-d (forward library, UT versus UC), e-h (forward library, UT versus ST), i-l (reverse library, UC versus UT), and m-p (reverse library, UC versus SC). Forward and reverse library clones are indicated by blue and yellow dots, respectively. Control spots are indicated as red, light blue, green or mauve dots. M and A values were calculated as described in (Dudoit et al., 2002, for example (a) $M = \log_2(UT/UC)$; $A = 1/2(\log_2(UT*UC))$; and for example (e) $M = \log_2(UT/ST)$; $A = 1/2(\log_2(UT*ST))$.

Each clone was quantitatively described by SSHscreen in terms of up/down regulation (Enrichment ratio 3 (ER3) values: $\log_2(UT/UC)$ for forward library; $\log_2(UC/UT)$ for reverse library) and rarity/abundance (Enrichment ratio 2 (inverse ER2) values: $\log_2(UT/ST)$ for forward library; $\log_2(UC/SC)$ for reverse library) in the treated sample; and a measure of statistical significance for each result was provided in the form of a moderated t-statistic with an associated p-value (Smyth, 2005).
The results of the SSHscreen 2.0.1 analysis were visualised by ER3 versus inverse ER2 plots for the forward and reverse libraries (Figure 2.7 A and Figure 2.7 B, respectively). Most of the genes in these plots fall in quadrant I, where ER3 > 0 and inverse ER2 < 0, meaning up-regulated by drought stress and rare in the unsubtracted drought stressed cDNA for the forward library (92%; Figure 2.7 A), and down-regulated by drought stress and rare in the control cDNA for the reverse library (52%; Figure 2.7 B). The criterion we chose to score genes as statistically significant differentially expressed (ER3 analysis: UT versus UC comparison) was that the adjusted p-value should be less than 0.05 after the linear model fit and empirical Bayes calculations. The p-value reflects the probability of rejecting the null hypothesis that there is no differential expression between the drought stressed (UT) and control (UC) samples for the forward library, and vice versa for the reverse library. There were 62% (1337/2146) significantly differentially expressed clones in the forward library and 34% (688/2018) in the reverse library using the stringent criterion of adjusted p-value < 0.05. Only the most significant 300 for each library are marked in Figure 2.7 A and B. The quality of the subtraction process in construction of the forward library was reflected in the low number of clones that had negative ER3 values (8%; Figure 2.7 A), whereas the subtraction was less efficient for the reverse library (48% with negative ER3 values; Figure 2.7 B).
Figure 2.7 ER3 versus inverse ER2 plot produced by SSHscreen for the cowpea forward (A) and reverse (B) libraries. A: ER3 for the forward library was calculated as the log-2 ratio of the unsubtracted treated cDNA (UT; drought stressed sample) divided by the
unsubtracted control cDNA (UC). Inverse ER2 was calculated as the log-2 ratio of UT divided by the subtracted treated cDNA (ST; SSH library enriched for genes up-regulated by drought stress). Data points were classified as representing transcripts: up-regulated by stress treatment/rare (Up.Rare: quadrant 1; ER3 > 0 and inverse ER2 < 0); up-regulated by stress treatment/abundant (Up.Abundant: quadrant 2; ER3 > 0 and inverse ER2 > 0); down-regulated by stress treatment/rare (Down.Rare: quadrant 3; ER3 < 0 and inverse ER2 > 0); and down-regulated by stress treatment/abundant (Down.Abundant: quadrant 4; ER3 < 0 and inverse ER2 < 0). The top 300 statistically significant clones are represented on the plot (adjusted p-value < 0.05).

**B:** ER3 for the reverse library was calculated as the log-2 ratio of the unsubtracted control cDNA (UC) divided by the unsubtracted treated cDNA (UT; drought stressed). Inverse ER2 was calculated as the log-2 ratio of the unsubtracted control cDNA (UC) divided by the subtracted control cDNA (SC; SSH library enriched for genes down-regulated by drought stress). Data points were classified as representing transcripts: down-regulated by stress treatment/rare (Down.Rare: quadrant 1; ER3 > 0 and inverse ER2 < 0); down-regulated by stress treatment/abundant (Down.Abundant: quadrant 2; ER3 > 0 and inverse ER2 > 0); up-regulated by stress treatment/rare (Up.Rare: quadrant 3; ER3 < 0 and inverse ER2 > 0); and up-regulated by stress treatment/abundant (Up.Abundant: quadrant 4; ER3 < 0 and inverse ER2 < 0). The top 300 statistically significant clones are represented on the plot (adjusted p-value < 0.05).

SSHscreen also provides an alternative statistic to choose differentially expressed genes, namely the B-statistic, by implementing this function of limma. The B-statistic (Lonnstedt and Speed, 2002; Smyth, 2004) can be interpreted as the log-odds that a specific gene is differentially expressed. This means that a B-statistic of zero corresponds to a 50-50 chance of differential expression, and accordingly a user is generally interested in genes with a positive B-statistic. For the cowpea ER3 analysis, 67% of the clones in the forward and 52% of the clones in the reverse library had positive B-statistics, which are larger numbers of clones than those selected based on the stringent criterion of t-statistic adjusted p-value < 0.05.

**Table 2.1** shows the top 20 cowpea clones sorted by p-value for the forward and reverse libraries extracted from the top tables that were generated from the ER3 analysis in SSHscreen. The most significant up-regulated forward library clone, 46D03-F, had a log-2 fold change of 2.9 (equivalent to the ER3 value). Taking the antilog of the log with base 2, it can be shown that this clone was ~8-fold up-regulated. With a similar calculation it can be shown that the most significant down-regulated reverse library clone, 45C07-R, with a log-2 fold change of 2.4, was ~5-fold down-regulated. The top table also reported the Average expression (A value) and statistics associated with the ER3 value, namely a moderated t-statistic, a p-value, an adjusted p-value and a B-statistic (**Table 2.1**). A top table for the ER2 analysis was also generated by SSHscreen, which reported the statistics of whether the clones represent rare or abundant transcripts in the original treated sample (data not shown).
Table 2.1 Top tables produced by SSHscreen for the forward and reverse cowpea libraries.

**Forward library top table: up/down regulation**

<table>
<thead>
<tr>
<th>ID</th>
<th>logFC(ER3)</th>
<th>AveExpr</th>
<th>t</th>
<th>P.Value</th>
<th>adj.P.Val</th>
<th>B</th>
<th>invER2</th>
</tr>
</thead>
<tbody>
<tr>
<td>46D03-F</td>
<td>2.91</td>
<td>10.89</td>
<td>34.37</td>
<td>5.40E-12</td>
<td>1.10E-08</td>
<td>19.51</td>
<td>-2.77</td>
</tr>
<tr>
<td>25B07-F</td>
<td>3.09</td>
<td>11.13</td>
<td>31.07</td>
<td>1.50E-11</td>
<td>1.60E-08</td>
<td>19.04</td>
<td>-3.06</td>
</tr>
<tr>
<td>13C10-F</td>
<td>2.71</td>
<td>10.62</td>
<td>29.27</td>
<td>2.80E-11</td>
<td>1.90E-08</td>
<td>18.73</td>
<td>-2.75</td>
</tr>
<tr>
<td>07E11-F</td>
<td>2.86</td>
<td>10.18</td>
<td>26.11</td>
<td>9.00E-11</td>
<td>3.20E-08</td>
<td>18.10</td>
<td>-2.84</td>
</tr>
<tr>
<td>13B02-F</td>
<td>2.77</td>
<td>10.59</td>
<td>25.71</td>
<td>1.10E-10</td>
<td>3.20E-08</td>
<td>18.01</td>
<td>-2.76</td>
</tr>
<tr>
<td>07E08-F</td>
<td>2.67</td>
<td>8.97</td>
<td>25.70</td>
<td>1.10E-10</td>
<td>3.20E-08</td>
<td>18.00</td>
<td>-2.51</td>
</tr>
<tr>
<td>26F04-F</td>
<td>2.98</td>
<td>11.46</td>
<td>25.60</td>
<td>1.10E-10</td>
<td>3.20E-08</td>
<td>17.98</td>
<td>-3.11</td>
</tr>
<tr>
<td>25A05-F</td>
<td>2.94</td>
<td>10.82</td>
<td>25.26</td>
<td>1.30E-10</td>
<td>3.20E-08</td>
<td>17.90</td>
<td>-2.81</td>
</tr>
<tr>
<td>25B08-F</td>
<td>2.86</td>
<td>11.25</td>
<td>24.99</td>
<td>1.40E-10</td>
<td>3.20E-08</td>
<td>17.83</td>
<td>-2.80</td>
</tr>
<tr>
<td>13C01-F</td>
<td>2.73</td>
<td>10.41</td>
<td>24.74</td>
<td>1.60E-10</td>
<td>3.20E-08</td>
<td>17.77</td>
<td>-2.77</td>
</tr>
<tr>
<td>05E09-F</td>
<td>2.58</td>
<td>10.86</td>
<td>23.98</td>
<td>2.10E-10</td>
<td>4.00E-08</td>
<td>17.58</td>
<td>-2.27</td>
</tr>
<tr>
<td>25C06-F</td>
<td>2.92</td>
<td>11.08</td>
<td>23.75</td>
<td>2.40E-10</td>
<td>4.00E-08</td>
<td>17.52</td>
<td>-2.92</td>
</tr>
<tr>
<td>25B06-F</td>
<td>3.60</td>
<td>11.17</td>
<td>22.99</td>
<td>3.30E-10</td>
<td>5.20E-08</td>
<td>17.31</td>
<td>-3.31</td>
</tr>
<tr>
<td>13F02-F</td>
<td>2.61</td>
<td>10.99</td>
<td>22.68</td>
<td>3.80E-10</td>
<td>5.20E-08</td>
<td>17.22</td>
<td>-2.84</td>
</tr>
<tr>
<td>13A08-F</td>
<td>2.65</td>
<td>10.84</td>
<td>22.54</td>
<td>4.00E-10</td>
<td>5.20E-08</td>
<td>17.18</td>
<td>-2.82</td>
</tr>
<tr>
<td>13C12-F</td>
<td>2.93</td>
<td>11.03</td>
<td>22.53</td>
<td>4.00E-10</td>
<td>5.20E-08</td>
<td>17.17</td>
<td>-2.89</td>
</tr>
<tr>
<td>3A07-F</td>
<td>3.32</td>
<td>10.42</td>
<td>22.27</td>
<td>4.50E-10</td>
<td>5.30E-08</td>
<td>17.10</td>
<td>-3.11</td>
</tr>
<tr>
<td>33C06-F</td>
<td>3.23</td>
<td>10.61</td>
<td>22.21</td>
<td>4.70E-10</td>
<td>5.30E-08</td>
<td>17.08</td>
<td>-3.22</td>
</tr>
<tr>
<td>06H12-F</td>
<td>2.41</td>
<td>10.54</td>
<td>21.83</td>
<td>5.50E-10</td>
<td>5.80E-08</td>
<td>16.96</td>
<td>-2.65</td>
</tr>
<tr>
<td>08B07-F</td>
<td>1.88</td>
<td>10.01</td>
<td>21.76</td>
<td>5.70E-10</td>
<td>5.80E-08</td>
<td>16.94</td>
<td>-2.40</td>
</tr>
</tbody>
</table>

**Reverse library top table: up/down regulation**

<table>
<thead>
<tr>
<th>ID</th>
<th>logFC(ER3)#</th>
<th>AveExpr</th>
<th>t</th>
<th>P.Value</th>
<th>adj.P.Val</th>
<th>B</th>
<th>invER2</th>
</tr>
</thead>
<tbody>
<tr>
<td>45C07-R</td>
<td>2.36</td>
<td>11.01</td>
<td>26.14</td>
<td>2.90E-10</td>
<td>5.50E-07</td>
<td>13.66</td>
<td>-1.35</td>
</tr>
<tr>
<td>36E04-R</td>
<td>2.25</td>
<td>10.81</td>
<td>22.19</td>
<td>1.40E-09</td>
<td>7.40E-07</td>
<td>13.20</td>
<td>-1.38</td>
</tr>
<tr>
<td>36B11-R</td>
<td>2.33</td>
<td>11.31</td>
<td>21.49</td>
<td>1.90E-09</td>
<td>7.40E-07</td>
<td>13.10</td>
<td>-1.22</td>
</tr>
<tr>
<td>44C07-R</td>
<td>2.39</td>
<td>11.69</td>
<td>20.93</td>
<td>2.40E-09</td>
<td>7.40E-07</td>
<td>13.02</td>
<td>-1.42</td>
</tr>
<tr>
<td>37B05-R</td>
<td>2.33</td>
<td>10.99</td>
<td>20.55</td>
<td>2.80E-09</td>
<td>7.40E-07</td>
<td>12.96</td>
<td>-1.07</td>
</tr>
<tr>
<td>45B02-R</td>
<td>2.26</td>
<td>11.05</td>
<td>20.32</td>
<td>3.20E-09</td>
<td>7.40E-07</td>
<td>12.92</td>
<td>-1.20</td>
</tr>
<tr>
<td>35F03-R</td>
<td>2.28</td>
<td>11.24</td>
<td>20.11</td>
<td>3.50E-09</td>
<td>7.40E-07</td>
<td>12.88</td>
<td>-1.41</td>
</tr>
<tr>
<td>35H05-R</td>
<td>2.35</td>
<td>11.22</td>
<td>20.08</td>
<td>3.50E-09</td>
<td>7.40E-07</td>
<td>12.88</td>
<td>-1.23</td>
</tr>
<tr>
<td>35E05-R</td>
<td>2.10</td>
<td>10.79</td>
<td>19.99</td>
<td>3.70E-09</td>
<td>7.40E-07</td>
<td>12.86</td>
<td>-1.02</td>
</tr>
<tr>
<td>45E02-R</td>
<td>2.29</td>
<td>10.88</td>
<td>19.88</td>
<td>3.90E-09</td>
<td>7.40E-07</td>
<td>12.84</td>
<td>-1.19</td>
</tr>
<tr>
<td>45G05-R</td>
<td>2.18</td>
<td>10.27</td>
<td>19.26</td>
<td>5.20E-09</td>
<td>9.00E-07</td>
<td>12.73</td>
<td>-1.24</td>
</tr>
<tr>
<td>35H04-R</td>
<td>2.24</td>
<td>11.22</td>
<td>18.85</td>
<td>6.40E-09</td>
<td>1.00E-06</td>
<td>12.65</td>
<td>-1.29</td>
</tr>
<tr>
<td>36A03-R</td>
<td>2.02</td>
<td>10.15</td>
<td>18.21</td>
<td>8.80E-09</td>
<td>1.30E-06</td>
<td>12.51</td>
<td>-1.22</td>
</tr>
<tr>
<td>16D06-R</td>
<td>1.65</td>
<td>11.02</td>
<td>17.88</td>
<td>1.10E-08</td>
<td>1.40E-06</td>
<td>12.44</td>
<td>-1.28</td>
</tr>
<tr>
<td>16C10-R</td>
<td>1.98</td>
<td>11.01</td>
<td>16.74</td>
<td>1.90E-08</td>
<td>2.30E-06</td>
<td>12.17</td>
<td>-1.20</td>
</tr>
<tr>
<td>35A01-R</td>
<td>2.19</td>
<td>10.92</td>
<td>16.73</td>
<td>2.00E-08</td>
<td>2.30E-06</td>
<td>12.16</td>
<td>-1.27</td>
</tr>
<tr>
<td>23G10-R</td>
<td>1.74</td>
<td>10.59</td>
<td>16.68</td>
<td>2.00E-08</td>
<td>2.30E-06</td>
<td>12.15</td>
<td>-1.08</td>
</tr>
<tr>
<td>45C04-R</td>
<td>2.17</td>
<td>10.50</td>
<td>16.11</td>
<td>2.80E-08</td>
<td>2.90E-06</td>
<td>12.00</td>
<td>-1.34</td>
</tr>
<tr>
<td>37C11-R</td>
<td>1.99</td>
<td>11.20</td>
<td>16.06</td>
<td>2.90E-08</td>
<td>2.90E-06</td>
<td>11.98</td>
<td>-1.08</td>
</tr>
<tr>
<td>16D05-R</td>
<td>2.09</td>
<td>11.12</td>
<td>15.85</td>
<td>3.20E-08</td>
<td>3.10E-06</td>
<td>11.92</td>
<td>-1.01</td>
</tr>
</tbody>
</table>
Only the top 20 statistically significant up- and down-regulated genes (before sequencing) are shown (sorted by B statistic).

*ER3 for Forward library calculated as $\log_2(UT/UC)$

#ER3 for Reverse library calculated as $\log_2(UC/UT)$

### 2.4.3 Annotation and management of cowpea SSH library sequences using SSHdb

The top tables (e.g. Table 2.1) and plots (Figure 2.7 A and 2.7 B) from SSHscreen analysis of the forward and reverse libraries were used to effectively select clones for sequencing based on the criteria of most significant differential expression and least likelihood of sequencing the same gene fragment twice. This was achieved by choosing those clones with the lowest adjusted p-value calculated from the ER3 values. Selection of clones that were spatially separated on the SSHscreen ER plots (Figure 2.7 A and 2.7 B) increased the likelihood of sequencing non-redundant clones. Sequence data for 118 clones in FASTA format, as well as SSHscreen top table data for the entire array, were uploaded to SSHdb for interpretation and management of the data (http://sshdb.bi.up.ac.za/). A schematic representation of the flow of data through SSHdb can be viewed in Figure 3 of Coetzer et al. (2010).

SSHdb removed the vector and adaptor fragments by performing BLASTN searches against the NCBI UniVec database (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html). Similarity searches were carried out against all sequences already uploaded in the database, in order to identify clones with the same sequence, i.e. redundant partners in the library, using a BLASTN E-value cut-off of 10e-10. Forty of the 118 sequenced clones were unique, implying that 67% of these sequences were redundant partners (Table 2.2). The largest group had 19 redundant partners. For each of the 40 redundant partner groups, the longest sequence in the group was selected by default as the representative clone. The choice of representative clone could be reviewed by downloading from SSHdb the multiple sequence alignments of redundant partner groups with two or more members (generated by ClustalW).

Following the identification of redundant partner groups, annotation was performed on the representative clones, using Blast2GO (Conesa et al., 2005), thereby inferring putative functions for each group (Table 2.2). BLASTN and BLASTX (Altschul et al., 1990) against the NCBI non-redundant nucleotide database (nt) and the NCBI non-redundant peptide database (nr) was also carried out. For cases where the E-value of the top BLASTX hit was low enough (less than 10e-10), this hit was automatically selected as the default priority annotation. Blast2GO and the top 10 BLASTX and BLASTN hits were stored in the database. For each redundant partner group, SSHdb allowed the top BLAST results to be viewed and
in several cases the priority annotation was changed after manual inspection. SSHdb linked the selected BLAST annotations to SSHscreen top table entries and it was possible to export different combinations of annotation information for selected subsets of clones (e.g. this allowed the construction of Figure 2.8, see later). One could export selected clones as FASTA files with the functional annotation as part of the header, which was particularly useful in preparing the sequences for submission to GenBank, or as a tab delimited text file containing various columns of available annotation information linked to the selected clones (Table 2.2). SSHdb also provides the option to export annotated SSHscreen top tables or Genepix Array List (GAL) files.
Table 2.2 Cowpea drought responsive genes annotated in SSHdb (after sequencing of selected clones). Table of sequenced cowpea drought responsive genes from the forward and reverse libraries with annotations in SSHdb derived from Blast2GO and BLAST analysis, as well as SSHscreen enrichment ratio values. Data is shown for representative clones from each redundant partner group.

**Forward library**

<table>
<thead>
<tr>
<th>Group number</th>
<th>Represen tative clone ID</th>
<th>Sequence description</th>
<th>BLAST2GO annotations</th>
<th>SSHscreen annotations</th>
<th>SSHdb annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>46D11-F</td>
<td>PR1 protein</td>
<td>C: extracellular region</td>
<td>2.10 1.E-05 10.0 -4.04</td>
<td>543 X AAD336 96 PR1a precursor [Glycine max] [PR1]</td>
</tr>
<tr>
<td>5</td>
<td>07F09-F</td>
<td>late embryogenesis abundant protein</td>
<td></td>
<td>1.15 5.E-07 14.9 0.05</td>
<td>473 X AAB387 82 late embryogenesis abundant 5 [G.max] [LEA]</td>
</tr>
<tr>
<td>6</td>
<td>46E07-F</td>
<td>MtN19-like protein</td>
<td>C: cytoplasmic membrane-bounded vesicle</td>
<td>1.10 2.E-04 6.4 -0.81</td>
<td>538 X AAU149 99 MtN19-like protein [Pisum sativum]</td>
</tr>
<tr>
<td>7</td>
<td>29F12-F</td>
<td>kunitz-type protease inhibitor kpi-f4</td>
<td>F: endopeptidase inhibitor activity</td>
<td>1.06 1.E-04 7.0 -3.47</td>
<td>404 X BAA828 40 miraculin [Youngia japonica] [MIR]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8</td>
<td>26B01-F</td>
<td>CPRD2 protein</td>
<td>F: FAD binding F: oxidoreductase activity</td>
<td>0.82</td>
<td>7.0E-04</td>
</tr>
<tr>
<td>9</td>
<td>26B07-F</td>
<td>ammonium transporter</td>
<td>C: membrane P: ammonium transport</td>
<td>0.50</td>
<td>9.0E-04</td>
</tr>
<tr>
<td>10</td>
<td>29D02-F</td>
<td>universal stress protein family protein</td>
<td>C: cytosol P: response to stress C: plastid</td>
<td>0.30</td>
<td>3.0E-02</td>
</tr>
<tr>
<td>11</td>
<td>30B02-F</td>
<td>---</td>
<td>---</td>
<td>0.25</td>
<td>3.0E-02</td>
</tr>
<tr>
<td>12</td>
<td>14B10-F</td>
<td>calmodulin-binding protein</td>
<td>F: calmodulin binding</td>
<td>0.22</td>
<td>1.0E-01</td>
</tr>
<tr>
<td>13</td>
<td>46G04-F</td>
<td>protein</td>
<td>C: cytoplasmic membrane-bounded vesicle F: protein binding C: membrane</td>
<td>0.17</td>
<td>1.0E-01</td>
</tr>
<tr>
<td>14</td>
<td>31A10-F</td>
<td>---</td>
<td>---</td>
<td>-0.09</td>
<td>3.0E-01</td>
</tr>
</tbody>
</table>

*ER3 for forward library calculated as $\log_2$(drought stressed/control)
### Reverse library

<table>
<thead>
<tr>
<th>Group number</th>
<th>Representative clone ID</th>
<th>Sequence description</th>
<th>GO-terms</th>
<th>logFC (ER3)*</th>
<th>adj.P. Val</th>
<th>B</th>
<th>invER2</th>
<th>Vector free sequence (bp)</th>
<th>BLAST Priority (X/N)</th>
<th>BLAST AccNr</th>
<th>BLAST Hit Def</th>
<th>BLAST Eval</th>
<th>Red Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>38A07-R</td>
<td>chlorophyll a/b binding protein</td>
<td>C: membrane C: thylakoid F: protein binding P: photosynthesis, light harvesting C: chloroplast</td>
<td>1.32</td>
<td>2.E-05</td>
<td>10.29</td>
<td>-2.35</td>
<td>432</td>
<td>BLASTX</td>
<td>NP_181539</td>
<td>LHCb4.3[light harvesting complex PSI][A.thaliana] [LHC]</td>
<td>2.E-43</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>37F06-R</td>
<td>ribulose-bisphosphate carboxylase</td>
<td>P: reductive pentose-phosphate cycle F: ribulose-bisphosphate carboxylase activity F: monoxygenase activity P: photosynthesis P: oxidation reduction C: chloroplast C: mitochondrion</td>
<td>0.94</td>
<td>4.E-04</td>
<td>7.73</td>
<td>-0.9</td>
<td>369</td>
<td>BLASTX</td>
<td>AAG24882</td>
<td>ribulose-1.5-bisphosphate carboxylase rbcS1 [G.max]</td>
<td>1.E-23</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>36G04-R</td>
<td>---</td>
<td>---</td>
<td>0.76</td>
<td>1.E-04</td>
<td>8.95</td>
<td>-2.94</td>
<td>669</td>
<td>BLASTX</td>
<td>AAO74628</td>
<td>lipid transfer protein II [Vigna radiata] [LTP]</td>
<td>1.E-30</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Accession</td>
<td>Type</td>
<td>Function Description</td>
<td>E-value</td>
<td>SCORE</td>
<td>Identity</td>
<td>Similarity</td>
<td>Q-values</td>
<td>Accession</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>21F04-R</td>
<td>type III chlorophyll a/b-binding protein</td>
<td>P: photosynthesis, light harvesting C: photosystem II F: magnesium ion binding P: protein-chromophore linkage F: chlorophyll binding C: integral to membrane C: chloroplast thylakoid membrane C: mitochondrion</td>
<td>0.46</td>
<td>2.E-02</td>
<td>-2.44</td>
<td>-1.67</td>
<td>492</td>
<td>BLASTX</td>
<td>1609235 chlorophyll a/b binding protein 8 [Solanum lycopersicum]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>42G12-R</td>
<td>---</td>
<td>---</td>
<td>0.43</td>
<td>2.E-01</td>
<td>-0.05</td>
<td>-3.2</td>
<td>505</td>
<td>---</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>15F11-R</td>
<td>---</td>
<td>---</td>
<td>0.19</td>
<td>2.E-01</td>
<td>0.09</td>
<td>2.99</td>
<td>620</td>
<td>N</td>
<td>AF479110 5.E-176 0.04 5.E-01 -1.5 -2.39 611 X AAD27876 carbonic anhydrase [V. radiata]</td>
<td>5.E-176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>24B10-R</td>
<td>pg1 protein</td>
<td>C: chloroplast</td>
<td>0.18</td>
<td>3.E-01</td>
<td>-0.4</td>
<td>-2.82</td>
<td>569</td>
<td>N</td>
<td>EU196765 0.29 1.E-01 0.49 -2.78 523 BLASTX Q43019 lipid-transfer protein 3 [Prunus dulcis] 26S ribosomal RNA [M. sativa] [26S] 16S chloroplast rRNA [P. vulgaris] carbonic anhydrase [V. radiata]</td>
<td>0.E+00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>36A11-R</td>
<td>carbonic anhydrase</td>
<td>F: zinc ion binding C: chloroplas stroma F: carbonate dehydratase activity P: carbon utilization</td>
<td>0.04</td>
<td>8.E-01</td>
<td>-1.5</td>
<td>-2.39</td>
<td>611</td>
<td>X</td>
<td>AAD27876 0.04 8.E-01 -1.5 -2.39 611 X AAD27876 carbonic anhydrase [V. radiata]</td>
<td>1.E-100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>43B08-R</td>
<td>---</td>
<td>---</td>
<td>-0.05</td>
<td>7.E-01</td>
<td>-1.4</td>
<td>-0.69</td>
<td>144</td>
<td>---</td>
<td>No significant similarity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>36A10-R</td>
<td>---</td>
<td>---</td>
<td>-0.09</td>
<td>4.E-01</td>
<td>-1.1</td>
<td>-4.31</td>
<td>766</td>
<td>X</td>
<td>AAG13810 0.04 8.E-01 -1.5 -2.39 611 X AAD27876 carbonic anhydrase [V. radiata]</td>
<td>7.E-49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>38E10-R</td>
<td>protein</td>
<td>F: protein binding</td>
<td>-0.10</td>
<td>3.E-01</td>
<td>-0.6</td>
<td>-2.53</td>
<td>254</td>
<td>X</td>
<td>NP_190638 leucine-rich repeat family protein [A.thaliana]</td>
<td>2.E-26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Accession</td>
<td>Description</td>
<td>P: Activity</td>
<td>F: Function</td>
<td>C: Location</td>
<td>Score</td>
<td>Fold Change</td>
<td>ID</td>
<td>Avg</td>
<td>FDR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>--------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------------</td>
<td>-------</td>
<td>-----</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>16B07-R</td>
<td>aldo keto reductase</td>
<td>-0.11</td>
<td>2.E-01</td>
<td>-0.1</td>
<td>-3.65</td>
<td>X</td>
<td>NP_565656</td>
<td>1.E-09</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>21H08-R</td>
<td>---</td>
<td>-0.13</td>
<td>5.E-01</td>
<td>-1.1</td>
<td>-2.48</td>
<td>746</td>
<td>X</td>
<td>NP_568178</td>
<td>2.E-23</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>43E12-R</td>
<td>prolyl endopeptidase</td>
<td>-0.13</td>
<td>3.E-01</td>
<td>-0.6</td>
<td>-0.28</td>
<td>147</td>
<td>X</td>
<td>AAL86330</td>
<td>1.E-13</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>21D09-R</td>
<td>at1g06530-like protein</td>
<td>-0.14</td>
<td>1.E-01</td>
<td>0.3</td>
<td>-3.77</td>
<td>437</td>
<td>X</td>
<td>ABH03547</td>
<td>2.E-14</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>24G10-R</td>
<td>unknown [Glycine max]</td>
<td>-0.16</td>
<td>1.E-01</td>
<td>0.6</td>
<td>-1.97</td>
<td>323</td>
<td>X</td>
<td>Q9SSU8</td>
<td>3.E-02</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>43G10-R</td>
<td>cell elongation protein</td>
<td>-0.16</td>
<td>7.E-02</td>
<td>1.0</td>
<td>-0.36</td>
<td>115</td>
<td>X</td>
<td>AAK15493</td>
<td>3.E-13</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>37B02-R</td>
<td>---</td>
<td>-0.34</td>
<td>2.E-02</td>
<td>2.7</td>
<td>-0.82</td>
<td>541</td>
<td>N</td>
<td>EU196765</td>
<td>0.E+00</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>38A04-R</td>
<td>---</td>
<td>-0.40</td>
<td>9.E-04</td>
<td>6.8</td>
<td>-0.16</td>
<td>695</td>
<td>N</td>
<td>AF479105</td>
<td>0.E+00</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ER3 for reverse library calculated as log2(control/drought stressed)*
2.4.4 Cowpea SSH library contains genes known to play a role in plant response to stress

Table 2.2 is a summary of the annotations for the cowpea SSH libraries that were extracted from SSHdb's sequence database. The data for the forward library is sorted by ER3 values, which represents the amount of up-regulation of the transcript in drought stressed IT96D-602 cowpea plants compared to the control treatment. Several genes with known roles in the stress response in other plants were present at high frequency in the cowpea SSH forward library with positive ER3 values, such as glutathione S-transferase (GST), a late embryogenesis abundant 5 protein (LEA), miraculin (MIR), thaumatin (THAU), pathogenesis related protein 1 (PR1), cowpea responsive to dehydration 2 (CPRD2), and a universal stress response protein (Table 2.2). Photosynthesis related genes had positive ER3 values in the reverse library screening indicating that their transcripts were up-regulated in the control treatment, which means they were down-regulated in the drought stressed IT96D-602 cowpea plants (Table 2.2). Table 2.2 illustrates the usefulness of the output from SSHdb, showing the 14 redundant partner groups from the forward library and 26 redundant partner groups from the reverse library. For each group, the representative clone's ID is given, together with the number of redundant partners in that group. Also, each representative clone is labelled with its ER3 value, adjusted p-value, B-statistic and inverse ER2 value calculated by SSHscreen, as well as with a putative function corresponding to the Blast2GO annotations and priority selected BLAST result for each group added by SSHdb. The provision of BLASTN results (as well as BLASTX results) is very useful, since several of the priority annotations were BLASTN hits to rRNA of chloroplast or nuclear origin, indicating that some of the highly abundant non-coding RNA had been retained in the mRNA preparation and was cloned in the SSH library. This is most likely due to priming on A-rich tracts within non-coding RNA or self-priming of rRNA during cDNA synthesis (Bloom et al., 2009).

Interestingly, inspection of the ER plots (Figure 2.7 A and 2.7 B) indicates that the majority of the genes (> 88%) that were cloned in both the forward and reverse SSH libraries have negative inverse ER2 values (present in quadrants I and IV). This indicates that most of the forward library clones were rare in the drought stressed IT96D-602 cowpea plants, and thus were enriched relative to other transcripts in this sample by the normalisation step of the SSH process (Figure 2.7 A). This is because a negative inverse ER2 value (log₂ (UT/ST)) means that the amount of molecules of the gene is greater in ST (i.e. after subtraction) than in UT (before subtraction). The same is true for the reverse library clones, indicating the transcripts are rare relative to other transcripts in the control plants (Figure 2.7 B).
Figure 2.8 shows the value of the ER plots to aid in the choice of non-redundant clones for sequencing. To illustrate this, we plotted the ER3 versus inverse ER2 values for a selection of clones from the eight largest redundant partner groups in the library (42 clones from the forward library and 26 from the reverse library). As indicated by the colour coding in Figure 2.8, clones from the same redundant partner groups clustered together. Drought stress up-regulated clones (ER3 > 0) encoding GST (mauve), THAU (red), PR1 (blue) and MIR (yellow) formed clusters that were relatively distinct, thus the choice of a few clones within each region is likely to capture the sequences for most genes in the library. Redundant partners of drought stress down-regulated clones (ER3 < 0 in Figure 2.8) also clustered together, namely lipid transfer protein (LTP; purple), LHCB4.3 light harvesting complex PSII (LHC; orange) and chlorophyll a/b-binding protein (CHL < 90 bp; green) (CHL > 170 bp; dark green). Clones encoding 26S rRNA (26S; blue) also clustered together with ER3 values close to 0. This indicates that 26S rRNA transcripts are present in similar quantities in the stressed and control cowpea plants, as expected, although non-coding RNA was not expected to be captured in either library.
Figure 2.8 ER3 versus inverse ER2 plot for sequenced clones to illustrate that redundant partners cluster together. The ER-values of clones from the eight largest redundant partner groups in the library were plotted. Clones from the same redundant partner groups clustered together. Groups are colour coded and labelled as follows: glutathione S-transferase GST (mauve), pathogenesis related protein 1a (PR1; blue), Thaumatin (THAU; red), miraculin (MIR; yellow), 26S rRNA (26S; blue), light harvesting complex PSII) (LHC; orange), lipid transfer protein (LTP; purple), chlorophyll a/b-binding protein (CHL < 90 bp; green) (CHL > 170 bp; dark green).

2.4.5 Verification of SSHscreen Enrichment Ratios using qPCR

Representative SSH library clones of six cowpea genes were selected for verification of the SSHscreen enrichment ratios using qPCR. These were three up-regulated genes from the forward library (GST, THAU and LEA; Table 2.2), two down-regulated genes from the reverse library (CHL and LTP; Table 2.2), and 26S rRNA which was not differentially expressed (Table 2.2). qPCR corroborated the direction of gene regulation (ER3 value) calculated by SSHscreen analysis of the microarray data for all six selected genes (Figure 2.9 A; compare blue to purple, yellow and green bars). Firstly, qPCR was carried out on the unsubtracted material used to construct the SSH libraries, the same material used to
determine the ER3 values. The unsubtracted treated (UT) cDNA sample was a mixture of cDNA from drought stressed cowpea IT96D-602 at 9 and 12 days; and the unsubtracted control (UC) was a mixture of cDNA from control cowpea Tvu7778 at 9 and 12 days. After normalisation of the qPCR data using the glyceraldehyde-3-phosphate dehydrogenase C-subunit (gapC) gene, an expression ratio was calculated \( \log_2(\text{drought stressed cowpea/control cowpea}) \). Good correlation between the ER3 values and the qPCR expression ratios was seen for all six genes (Figure 2.9 A; compare blue bars with purple bars). GST, THAU and LEA were up-regulated, CHL and LTP were down-regulated and 26 S rRNA unchanged (Figure 2.9 A). Secondly, since the libraries were constructed from mixtures of cDNA at two time points, RT-qPCR was carried out on the RNA samples from the individual time points before they were pooled for SSH library construction (Figure 2.9 A, yellow and green bars). GST and THAU were up-regulated, and CHL and LTP were down-regulated at both time points, thus corroborating the ER3 values (Figure 2.9 A). Interestingly, LEA was up-regulated at 9 d and down-regulated at 12 d, and thus the transcript abundance measured in the mixtures used to make the SSH library is likely to be an average between the two (Figure 2.9 A). RT-qPCR analysis of 26S rRNA at the two time points gave expression ratios that are essentially unchanged between the two treatments (Figure 2.9 B).

The SSH process aims to equalise the proportion of genes in the final subtracted sample before cloning by enriching for rare transcripts and suppressing the amplification of highly abundant transcripts (Diatchenko et al., 1999). A rare gene before subtraction should be in increased amounts in the subtracted sample and vice versa. The inverse ER2 value (\( \log_2(\text{UT}/\text{ST}) \) for the forward library; \( \log_2(\text{UC}/\text{SC}) \) for the reverse library) provides a measure of this, since clones with inverse ER2 value < 0 are rare before subtraction (i.e. UT < ST or UC < SC). The cowpea drought expression forward and reverse libraries contained mostly rare clones, with inverse ER2 values < 0 (see Figures 2.7 A and 2.7 B; blue bars in Figure 2.9 B). qPCR was also used to verify the SSHscreen inverse ER2 values using the same cDNA samples, and all five genes that were tested (GST, THAU, LEA, CHL and LTP) gave negative \( \log_2(\text{before/after subtraction}) \) values, and closely mirrored the inverse ER2 values, confirming that they were rare in the unsubtracted samples (compare purple with blue bars; Figure 2.9 B). 26S rRNA transcripts are expected to be abundant in any plant cell, however, the amount of rRNA in the cDNA sample derived from the mRNA isolation step is unlikely to be representative, since it is present due to false priming. Importantly, normalisation of the qPCR data for verification of the ER2 values cannot be done with an endogenous housekeeping gene, since no product should be present in the same abundance before and after subtraction. Therefore, equal amounts of an alien gene fragment (human beta-globin) were spiked into the cDNA samples and effectively used for normalisation of the qPCR data.

© University of Pretoria
Figure 2.9 Regulation (A) and relative abundance (B) of selected cowpea genes (qPCR verification). A: Confirmation of differential expression in drought stressed tolerant cowpea (IT96D-602) versus control susceptible cowpea (Tvu7778) observed in microarray studies. The expression ratios for each gene in the microarray experiment are indicated by blue bars and qPCR on cDNA by purple bars. RT-qPCR using total RNA isolated from leaves after 9 and 12 days of stress treatment are indicated in yellow and green bars, respectively. B: Confirmation that transcripts of selected genes had low abundance (i.e. rare) before subtraction. The log2 ratios before and after SSH are presented. Negative log2 ratios indicate that cDNAs have greater signals in subtracted samples compared to unsubtracted samples, indicating that they were rare before subtraction and have been enriched by the SSH process. Results from the microarray experiment are indicated by blue bars and the qPCR results by purple bars. (Error bars = standard deviation of replicate qPCR experiments)
2.5 Discussion

2.5.1 SSHscreen and SSHdb facilitates the screening and annotation of SSH library clones

SSH remains a popular approach for gene discovery based on its advantages of enriching for genes that are differentially expressed between treatments, as well as the recovery of rare transcripts (Diatchenko et al., 1996; Hillmann et al., 2009). SSH has proven particularly useful as a first step in genomics research of non-model organisms that do not have genome sequence information (Crampton et al., 2009; Van den Berg et al., 2007). In this study, we have used two software tools, SSHscreen and SSHdb, which greatly facilitate drought-responsive gene discovery from the non-model crop cowpea using SSH. This approach represents a significant improvement compared to commonly used approaches in which SSH libraries are screened qualitatively using inverse dot blots, and sequence information is stored and managed on an individual researcher's desktop.

SSH libraries, constructed using either a commercial kit or homemade protocols, have the limitation that they often contain clones derived from transcripts that escaped subtraction (i.e. false positives), clones derived from highly abundant RNA species, such as rRNA, and some redundancy (i.e. the same inserts in several clones) (Van den Berg et al., 2004). For example, even though we performed the subtraction effectively, the forward and reverse SSH libraries constructed from cowpea plants in this study were calculated to have 9% and 46% false positives, respectively (negative ER3 values). Despite using mRNA for library construction, 6% of the clones in the reverse library were 26S rRNA, mostly likely due to self-priming or priming on A rich regions by the oligo-dT primer (Bloom et al., 2009). Approximately 67% of the sequenced clones were redundant. This means that sequencing all the clones from an SSH library would be a very inefficient use of resources, since many false positives and redundant clones would be sequenced. Commonly, this is overcome by first screening the SSH library clones as colonies or PCR products on nylon membranes using inverse dot blots (Hein et al., 2004). This, however, does not provide accurate quantification, and the choice of clones to use for normalisation is difficult. This study employs an alternative approach of using a simple R package SSHscreen 2.0.1 to apply appropriate control spot normalisation methods, and calculate differential expression ratios with statistical support after screening the SSH clones on a small number of microarray slides.

SSHscreen is tailored for the screening of SSH libraries as a semi-automated process with appropriate background correction, normalisation, differential expression analysis and false
discovery rate corrections. In a single analysis, the user can submit microarray data for slides containing both forward and reverse libraries, hybridised with cDNA targets for both ER3 and ER2 calculations, and obtain top tables (i) of differentially expressed genes between the unsubtracted treatment and control, and (ii) with information on the relative abundance of cloned transcripts in these unsubtracted samples. Another output of SSHscreen is the ER3 vs. inverse ER2 plot for each library (Figure 2.7 and 2.8).

Within-slide normalisation of two-colour microarray data is an important consideration to account for systematic bias due to differences between the Cy3 and Cy5 dyes (Smyth and Speed, 2003). Commonly, loess normalisation is applied (Smyth and Speed, 2003), however, this is based on the assumption that most of the genes on the array are not differentially expressed. This is legitimate for most whole genome microarray experiments, however, it is not appropriate when the array is constructed from an SSH library, which selects for differentially expressed genes. SSHscreen 2.0.1 provides spike-in control spot-based normalisation with the option of giving full or partial weight to control spots when fitting the loess curve. Fardin et al. (2007) state that there are two main considerations for the spikes in this type of normalisation of custom arrays. Firstly, the intensities of the spikes should span the range of intensities of the experimental data, and secondly, the strength of the loess curve with respect to the number of replicates. We applied full weight to the spike-in control spots in our normalisation approach, and effective normalisation can be seen in the boxplots of the control spots across the 12 slides (Figure 2.5; similar to Figure 1 of Fardin et al. (2007)) which shows the variability of M values in the raw data is considerably diminished by the normalisation. The spike-in control spots spanned 90-99% of the range of data intensities (see A values on the MA plots in Figure 2.6), which is sufficient compared to the 75% range reported by Fardin et al. (2007). Reliability of the normalisation and strength of the loess curve through the control spots is further illustrated by our calculations that the average standard deviations of the M values for the spike-in controls across the 12 slides decreased from 0.18 to 0.10 after normalisation, a similar improvement to that reported in Figure 2 of Fardin et al. (2007).

The outputs of SSHscreen are top tables of ER3 and inverse ER2 values which the user can rank based on moderated t-statistics and associated adjusted p-values calculated with limma functions (Smyth, 2005). Clones can be chosen for sequencing based on a positive ER3 value and a user-defined threshold of p-value adjusted for multiple testing (Benjamini and Hochberg, 1995). SSHscreen 2.0.1 also provides the B-statistic as an alternative for choosing clones. The prior estimate of the percentage of differentially expressed genes is set
at 50% default in SSHscreen 2.0.1. Clones with a positive B-statistic have more than a 50% chance of differential expression.

Enrichment ratio 2 is calculated in SSHscreen 2.0.1 as an inverse ER2 value ($\log_2(UT/ST)$) for ease of interpretation in the ER3 versus inverse ER2 plots, since it arranges rare → abundant transcripts from left → right on the plot (Figure 2.7). It gives a measure of whether a clone in the library represents a transcript that was rare or abundant in the original tester sample, based on the theory of the SSH process that normalises the relative amount of transcripts in the final subtracted tester sample that is cloned (Diatchenko et al., 1996). SSHscreen 2.0.1 provides a plot of the ER3 versus inverse ER2 values, which provides another tool in the selection of clones for sequencing. As shown in the current cowpea study (Figure 2.8), redundant clones clustered on the ER3 versus inverse ER2 plot, thus these plots can be used to choose clones for sequencing that are spatially separated. Interestingly, this plot was able to distinguish between longer and shorter clones of CHL (Figure 2.8). It should be noted that clusters do overlap (Figure 2.8), so although this plot serves to improve the efficiency of selecting unique clones, some redundant clones will be chosen.

2.5.2 Microarray-derived SSHscreen enrichment ratios were confirmed by qPCR
The ER3 and ER2 calculations derived from microarray hybridisation signals were validated using an independent technique, qPCR. Three cowpea genes, encoding GST, THAU, and LEA were significantly up-regulated more than 2-fold in the drought stressed cowpea plants compared to the control plants (ER3 value > 1; adjusted p-value < 0.05). qPCR of the UT and UC cDNA mixes prior to subtraction, as well as RT-qPCR of RNA from the individual time points used to make the UT and UC mixes, confirmed the up-regulation of these three genes (Figure 2.9 A). Interestingly, RT-qPCR showed that LEA was up-regulated at 9 d after initiation of drought stress and down-regulated at 12 d after drought stress, and thus the microarray and qPCR of the UT/UC mixes represent the average (Figure 2.9 A). Similarly, ER3 values for two selected down-regulated genes, CHL and LTP were confirmed by qPCR and RT-qPCR. The 26S rRNA escaped subtraction in the construction of the reverse subtraction library and thus is observed to be at equal quantities in the UT and UC samples prior to subtraction (i.e. ER3 ~0) and this was also confirmed by the qPCR results (Figure 2.9 A). The inverse ER2 values for all five selected differentially expressed genes were negative, indicating that their transcripts were rare in the original tester samples and had been enriched during the normalisation step of the SSH process. qPCR confirmed this, indicating that the microarray hybridisations accurately reflect the relative amount of gene fragments in the target cDNA mixes (Figure 2.9 B).
2.5.3 SSHdb links SSHscreen data to sequence annotations for clones

The output from SSHscreen is a priority list of clones to sequence, and thus the efficient management of the sequence information in the context of the SSHscreen results was performed with SSHdb. SSHdb as a web-based tool with no software requirements for the user except an internet browser.

SSHdb proved very effective in managing the sequence information for a set of sequences obtained from the forward and reverse drought stressed cowpea libraries. Each clone was automatically annotated using two approaches, Blast2GO (Conesa et al., 2005) and BLAST similarity searches of sequenced clones against the NCBI non-redundant nucleotide and peptide databases (nt/nr) (Altschul et al., 1990). Several features of SSHdb make it particularly effective for non-model organisms for which there is not an annotated genome sequence available. Blast2GO was designed as an annotation tool for non-model organisms and uses a more robust approach than BLAST to assign putative functional annotations to sequences, as well as significant GO terms (Conesa et al., 2005). Providing BLASTN, as well as BLASTX hits, allows the identification of clones derived from non-coding RNA, which escaped the subtraction. This is a common problem in SSH library construction, as seen in this study with 6% clones derived from 26S rRNA. SSHdb also allows for manual curation of the annotations. The top ten BLAST hits sorted by E value are stored in the database, and the user is given the choice of choosing the representative annotation. Very often with non-model organisms the top hit is to a sequence that is not functionally annotated (e.g. "hypothetical protein", "expressed sequence"), whereas the second hit is to an annotated sequence, which can then provide the user with a working hypothesis of the putative identity of the clone. This was our experience for some of the cowpea clones in this study. In other studies, due to the poorly annotated rice genome in GenBank, the same problem was found with SSH clones from non-model monocots, pearl millet and banana, that had top hits to unannotated rice genes, whereas more useful hits within the top 10 were to sequences from other plants with annotations (Crampton et al., 2009; Van den Berg et al., 2007).

The SSHscreen data for each clone can be inspected in the SSHscreen toptable view, and annotated toptables or GAL files can be exported from SSHdb. This is particularly useful in cases where the same array is to be used later for gene expression profiling in a more in-depth study, e.g. over a time course of drought stress. Such an experiment could be analysed for differentially expressed genes using limma in R, e.g., which would benefit from an annotated GAL file so that it could immediately be seen if differentially expressed clones had been sequenced. In this study, another feature of SSHdb was used to export the
representative sequences of each redundant partner group in FASTA format with the correct header information, so that they could be submitted easily to dbEST at GenBank.

SSHdb is not limited to the management and analysis of sequences from SSH libraries, since it can organise any sequence dataset in FASTA format, including cDNA sequences from next generation sequencing projects. The cDNA Annotation System (CAS) is another generic tool for analysis of cDNA sequences (Kasukawa et al., 2003), however, it requires the complete NCBI database to be loaded and updated on individual desktops, and thus is less user-friendly for collaborative projects, such as ours, in which the co-workers are at different institutions.

2.5.4 Identification of cowpea drought response genes

Some inferences can be made of the cowpea genes identified in this study as role players in the drought response by comparison with studies of stress responses in other plants. A glutathione S-transferase, a late embryogenesis abundant protein 5, and a universal stress response protein have clear links to drought stress responses. Glutathione S-transferase (EC 2.5.1.18; GST, group 1, Table 2.2) is an enzyme that catalyses the conjugation of reduced glutathione, via its sulfhydryl group, to the electrophilic centers on various substrates (Dixon et al., 2002). Glutathione is a tripeptide present in the intracellular space of plants and other organisms, functioning to keep sulfhydryl groups reduced and to remove toxic metabolites. The induction of GST during drought stress in cowpea may protect the plant cells from a build-up of toxic compounds, thus contributing to its drought tolerance.

Late embryogenesis abundant (LEA, group 5, Table 2.2) proteins were initially discovered in desiccating plant seeds, but have subsequently been described in various plants and plant tissues. They are associated with abiotic stress tolerance in plants, namely desiccation, salt and cold stress (Tunnacliffe and Wise, 2007). Their structure changes during dehydration from an unordered conformation, lacking in tertiary structure, to a folded structure which may protect the cell from collapse, stabilising membranes or protecting other proteins by acting as chaperones during periods of water stress. Most LEA proteins fall into three main groups, but two unnumbered groups were discovered in cotton: LEA5 and LEA14 (Galau et al., 1993). These two are the only cloned cotton mRNAs encoding LEA’s that are highly induced in drought stressed leaves. They are predicted to be more hydrophobic and possibly more structured than LEA groups 1 – 3 (Tunnacliffe and Wise, 2007). LEA from the cowpea drought expression library in this study has the characteristic LEA_3 motif (Pfam family PF03242; http://www.sanger.ac.uk/Software/Pfam), and is most similar to the drought-
induced cotton LEA5 and a LEA5 protein identified in desiccating seeds of soybean [GenBank: AAB38782].

The cowpea drought stressed forward library also contained a clone that matched a universal stress protein from *A. thaliana* [TAIR: At5g54430.1] (group 10, Table 2.2). These plant proteins have sequence similarity to UspA that has been well characterised in bacteria. Bacterial UspA is a small serine and threonine phosphoprotein that is induced by several stress treatments, and strains with mutations in this gene are less stress tolerant (Freestone *et al.*, 1997). This may represent an ancient conserved stress mechanism at the cellular level. Iuchi *et al.* (1996) identified genes induced after 5 h of dehydration in detached leaves of cowpea line IT84S-2246-4, and named them "cowpea clones responsive to dehydration" (CPRD). One of these genes (CPRD2) was also isolated in our study (group 8, Table 2.2).

Several pathogenesis-related genes were induced during drought stress in cowpea, namely a PR1, THAU, and a wound induced protein (WIN2) (groups 2 to 4, Table 2.2). Overlap in the responses to biotic and abiotic stresses has been documented (Fujita *et al.*, 2006). This may reflect a structural stabilising role that these proteins may confer to protect against water loss and cellular damage by either stress. THAU has the unique property of being a very sweet protein with a distinct protein structure made up of beta-sheets with a high content of beta-turns and very few alpha-helices.

The reverse library was dominated by clones encoding components of photosynthesis, such as chlorophyll a/b binding proteins (groups 15, 17 and 22, Table 2.2) (de Bianchi *et al.*, 2008; Liu *et al.*, 2008), ribulose-1,5-bisphosphate carboxylase small subunit rbcS1, and the chloroplast genes fructose-bisphosphate aldolase 1 and phytoene synthase (groups 19, 16 and 37, Table 2.2). This reflects a reduction in photosynthesis during drought stress. Similar genes of the photosynthetic apparatus were also down-regulated in leaves of *P. vulgaris* under progressive drought stress (Kavar *et al.*, 2008). They include carbonic anhydrase and the photosynthesis-related genes encoding ribulose 1,5-bisphosphate carboxylase (large and small subunits), chlorophyll a/b-binding protein CP24 precursor and photosystem I light-harvesting chlorophyll a/b-binding protein. Chlorophyll a/b-binding proteins are part of the light-harvesting complex that act as antennae to capture light excitation energy and deliver it to photosystems I and II. In *Arabidopsis*, cab genes were also more than 5-fold repressed under drought stress (Seki *et al.*, 2002).

Three different lipid transfer proteins (LTP; group 20, 21 and 25, Table 2.2) were cloned in the reverse library. Plant LTPs show a highly conserved secondary structure, forming a
hydrophobic pocket capable of carrying a fatty acid, phospholipid or acyl-CoA, and have been shown in vitro to transfer lipids between membranes (Arondel et al., 2000). Drought responsive LTPs have been described in Solanum pennellii (Trevino and Connell, 1998). Down-regulation of LTP during drought stress possibly indicates a need to suppress LTP mediated signaling.

2.6 Acknowledgements

We acknowledge funding from the National Bioinformatics Network (NC), the National Research Foundation (NC, DKB) and a parliamentary grant to the Agricultural Research Council (ARC)-Roodeplaat (IG). We thank Dr. BB Singh for providing cowpea seed and Prof. BD Wingfield (FABI) for providing the Leptographium elegans clone. We thank Jan-Peter Nap, Wageningen University and Research Centre, for critical reading of the manuscript (Coetzer et al., 2010).
Chapter 3

Expression patterns and diversity study of the cowpea LEA5 gene

3.1 Abstract
A cowpea late embryogenesis abundant protein (LEA) gene, belonging to the atypical group 5, was identified in a drought expression library created by SSH and screened with microarrays (Chapter 2, Table 2.2, clone ID 07F09-F). In this chapter, the VuLEA5 protein gene’s structure was determined from the two different lines used in this study, IT96D-602 and Tvu7778. IT96D-602 contains two alleles (VuLEA5-1 and VuLEA5-2) differing by a SNP at position 226 in the coding region. The A→G substitution causes an amino acid change from isoleucine (codon ATA) to valine (codon GTA) at the 76th residue of the 107 amino acid protein. The gene from the Tvu7778 line contains a G at the SNP site (VuLEA5-2), and a single insertion of three nucleotides in its intron that is otherwise identical to IT96D-602. The amino acid sequence of the VuLEA5-1 gene from IT96D-602 was compared to other plant LEA5 proteins and cowpea ESTs. It is most similar to the G. max LEA5 gene [GenBank: AAB38782], with close homology to various other plant LEA proteins from group 5, specifically the D-73 type identified first from cotton (Galau et al., 1993). Cowpea ESTs of the VuLEA5 gene from nine different libraries deposited in GenBank were conserved, with only six positions affected by amino acid substitutions. One of them corresponds to the identified SNP position, and the cowpea libraries could be separated into two groups based on their allele content. Reverse transcriptase quantitative PCR (RT-qPCR) analysis of the expression pattern of the VuLEA5 gene was performed on the cowpea line IT96D-602 during drought stress and various in vitro abiotic stress treatments of detached cowpea leaves. The VuLEA5 gene expression was up-regulated by drought, desiccation, ABA (100 µM), NaCl (250 mM) and especially under cold (4°C) conditions. VuLEA5 protein sequence homology to other studied plant LEA5 proteins suggest some role in the protection of plant cells during any of a range of abiotic stress conditions (desiccation, drought, heat, salt, cold, high light and ABA).

3.2 Introduction
One of the drought responsive genes identified from a cowpea drought expression library (Chapter 2) was a LEA5 protein gene (SSH clone 07F09-F). A BLASTX homology search in
the protein sequence database GenBank found this gene to be most similar to a LEA5 protein identified in desiccating seeds of soybean [GenBank: AAB38782].

The literature available on LEA proteins states a strong case for LEAs to play an important protective role during drought and other abiotic stresses of plants (Baker et al., 1988; Wise and Tunnacliffe, 2004; Tunnacliffe A and Wise MJ, 2007). It was, therefore, decided to characterise this identified VuLEA5 gene further regarding its expression patterns and its sequence diversity in cowpea and other plants.

The objectives reached to achieve this aim were: to investigate possible allele differences in the VuLEA5 gene from the two different cowpea lines used in this study, IT96D-602 and Tvu7778, by cloning it from genomic DNA and cDNA prepared from these lines; to compare the IT96D-602 VuLEA5-1 protein sequence to other plant LEA5 proteins to determine possible related functions; and to compare it to other available cowpea VuLEA5 expressed sequence tags (EST) in the GenBank sequence database to detect more, if any, single nucleotide polymorphisms (SNPs) of this gene in cowpea. The last objective was to characterise the VuLEA5 gene expression patterns in cowpea leaves with quantitative PCR during a greenhouse drought trial and after abiotic stress treatment of detached cowpea leaves.

3.3 Materials and Methods

3.3.1 Cloning of different alleles of VuLEA5 cDNA
The coding sequence of the VuLEA5 gene was amplified from 100 ng DNase-treated RNA isolated from two replicates each of 12 day water-stressed greenhouse grown cowpea (Vigna unguiculata (L.) Walp.) plants of the lines Tvu7778 and IT96D-602. RT-PCR reactions were performed using the C-therm Polymerase one-step RT-PCR system (Roche) in 50 µl reactions. The primers used to amplify the coding region were LEA5_start and LEA5_stop (Appendix 1). PCR products were cleaned up (Qiagen Minelute PCR purification kit) and cloned into pGEM-T Easy (Promega). Ligation reactions were transformed into competent E. coli JM109 cells (Promega) and plated out onto LB-agar using ampicillin antibiotic selection and X-Gal/ IPTG blue-white screening. PCR-positive clones were sequenced using the T7 primer (Appendix 1; Inqaba Biotec). Sequence analyses were performed using BioEdit Sequence alignment editor 7.0.9.0 (Hall, 1999).
3.3.2 Cloning of VuLEA5 genomic sequences from two cowpea lines

The VuLEA5 gene was amplified from genomic DNA from two cowpea lines, IT96D-602 and Tvu7778, using the LEA5_start and LEA5_stop primers (Appendix 1) and MyTaq HS DNA polymerase (Bioline). A standard reaction mixture and a primer annealing temperature of 56°C were used. The PCR products were cloned into pGEM-T Easy (Promega) as described above. PCR-positive clones were sequenced using the M13-26-R primer (Appendix 1; Inqaba Biotec).

3.3.3 Bioinformatic analysis of the VuLEA5 protein

Amino acid comparisons and multiple alignments of VuLEA5-1 and other LEA proteins were performed using the ClustalW program (Chenna et al., 2003) and visualised with BioEdit Sequence alignment editor 7.0.9.0 (Hall, 1999). Phylogenetic analysis was performed using the Neighbor Joining method using the MEGA4.0 program (Tamura et al., 2007). The molecular mass and the isoelectric point (pl) were predicted with tools at http://web.expasy.org. The prediction of a transit peptide of the VuLEA5-1 protein was carried out by the iPSORT algorithm (Bannai et al., 2002; http://ipsort.hgc.jp). The Pfam domain was identified with a tool at http://pfam.sanger.ac.uk. Hydropathy/ predicted hydrophilicity of the VuLEA5-1 protein was performed according to Kyte and Doolittle (1982) (http://web.expasy.org/protscale). Secondary structures were predicted with PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred).

3.3.4 Diversity of cowpea VuLEA5 genes

Expressed sequence tags (EST) of VuLEA5 genes from different libraries prepared from cowpea were downloaded from the GenBank database. Alignments were performed using ClustalW and the MEGA4 program and their phylogenetic relationships represented by constructing trees using the Neighbor Joining method (Tamura et al., 2007).

3.3.5 Plant material and abiotic stress treatments

V. unguiculata (L.) Walp. plants of the drought tolerant IT96D-602 and a susceptible Tvu7778 lines were grown in pots (25 cm diameter), in a mixture of compost, sand and vermiculite (5:2:2, v/v), in a greenhouse at 28/18°C day/ night temperature. Seeds were first coated with a cowpea inoculant containing Bradyrhizobium -sp (Vigna) bacteria (Stimuplant CC., Zwavelpoort). Emerged seedlings were thinned to two plants per pot. A progressive drought trial was implemented when the plants were six weeks old, by withholding watering from treated pots. Control plants were watered and Multifeed applied. Occasional spraying of Hunter against spider mite was performed. Four treatments (IT96D-602 control, Tvu7778...
control, IT96D-602 drought stressed and Tvu7778 drought stressed) consisted each of nine replicate pots, pooled into 3 groups to make up a biological replicate.

At 5, 9 and 12 days after withholding water, and 3 days after re-watering, the second mature leaves from three pooled plants per triplicate biological replicate were harvested. The lateral leaflets (six) were pooled, ground to a fine powder in liquid nitrogen and stored at -70°C for RNA extractions. The three terminal leaflets were pooled for relative water content (RWC) measurements. Three 10mm leaf disks were punched from each with a no. 6 cork borer and the fresh weight (W) of the nine pooled leaf disks determined. After hydration in distilled water for 4 hours and blotting dry on paper towel, the turgid weight (TW) was recorded. The dry weight (DW) was measured after drying in an oven at 60°C for at least 48 hours. The RWC is calculated using the following formula: RWC = (W – DW)/(TW – DW) x 100 where W = fresh weight, TW = turgid weight and DW = dry weight.

For abiotic stress treatments on detached cowpea leaves, the second fully expanded leaf of 7-week old greenhouse grown IT96D-602 cowpea plants were used in all experiments. Detached leaves were desiccated by air-drying the leaves at dim light at 23°C for 17 h. Control leaves were frozen immediately. In recovery experiments, desiccated leaves were re-hydrated in water and harvested 24 h later. For other stress treatments, the leaves were cut off plants and immediately submerged in water. The lower portion of the petiole was cut off and the detached leaves placed in tubes of buffer (10 mM Tris-Cl, pH 7.0) containing 250 mM NaCl. Control leaves were incubated in buffer at 23°C. For the cold stress treatment, leaves in tubes containing buffer were incubated at 4°C. Exogenous ABA was applied by immersing the petioles of detached cowpea leaves in buffer containing 100 µM ABA (Sigma Aldrich). The leaves were treated for 5 h or 24 h and then frozen in liquid nitrogen and kept at -80°C until RNA isolation. The experiment was performed in duplicate.

3.3.6 RNA isolation
Frozen cowpea leaf material was ground with a mortar and pestle in liquid nitrogen. RNA was isolated from approximately 100 mg ground material using the RNeasy Plant kit (Qiagen) according to the manufacturer’s protocol. Contaminating genomic DNA was removed from the total RNA with the Turbo DNA-free kit (Ambion). Genomic DNA contamination was checked with RT-PCR using the bean Actin primers (Appendix 1) and the C-therm Polymerase One-Step RT-PCR system (Roche). PCR was performed using GoTaq DNA polymerase (Promega). The RT-PCR reaction was incubated at 60°C for 30 min before cycling. The primer annealing temperature for both RT-PCR and PCR was 56°C, and
30 cycles were included. All amplification products were analysed on a 2% agarose/0.5x TAE gel.

3.3.7 RT-qPCR
A master mix of 1x Sensimix One Step (Quantace, Celtic Molecular Diagnostics), SYBR Green, RNase inhibitor (0.2 U/µl), MgCl₂ (final concentration of 4 mM), the appropriate primer pair (LEA5, NCED, GAPC, VuELF1B and Gm40S; 200 nM each; Appendix 1) and template in a total volume of 25 µl was prepared following the manufacturer’s protocol. For some experiments the Sensimix SYBR One Step kit was used. Hundred nanograms of each RNA template were added to the reaction mastermix in 0.1 ml strip tubes. Reactions were prepared in triplicate and run on a Rotor-Gene 3000A (Corbett Research). Triplicate standard curves were constructed for each primer using a tenfold dilution (100 to 0.01 ng) of DNase-treated cowpea IT96D-602 RNA pooled from control and water-stressed (9 and 12 d without water) samples. The cycling profile included a hold at 45°C (42°C for the Sensimix SYBR One Step kit) for 30 min for reverse transcription, 95°C for 10 min for enzyme activation and 45 cycles of 95°C for 15 s, primer annealing at 56°C for 30 s and extension at 72°C for 6 s. The SYBR green measurement was taken every cycle after the extension step. The melt curve was constructed by holding for 45 s at 72°C, and 5 s per step while increasing the temperature from 72 to 95°C by 1°C increments.

C_q values were obtained from the fluorescence curves at a threshold of 0.01 and a cut-off of 10 cycles. RT-qPCR data was analysed with the software program qBase (Hellemans et al., 2007). The genes were grouped as target genes (LEA5 and NCED) or reference genes (GAPC, VuELF1B and Gm40S). Efficiency values were calculated for each primer pair from the slopes of the standard curves. The C_q values were converted to input template (in nanogram) using the standard curve of each gene. Reference gene stability was analysed with qBase, whereafter the gene expression values were normalised with two reference genes (GAPC and VuELF1B).

3.3.8 In silico gene expression studies
The Genevestigator tool (http://www.genevestigator.com/gv/) was used to study the expression patterns of specific LEA5 genes from soybean and *A. thaliana*. Expression of *A. thaliana* LEA5 genes during various abiotic stresses was visualised with the Arabidopsis eFPBrowser (http://bar.utoronto.ca) (Winter et al., 2007).
3.4 Results

3.4.1 Sequence and bioinformatics of the VuLEA5 gene

A 477 bp fragment of the cDNA sequence of VuLEA5 was cloned in the drought-expression library following SSH (Chapter 2; SSH library clone ID: 07F09-F; GenBank accession number GR942575, Appendix 3). The VuLEA5 cDNA fragment contains an open reading frame of 324 bp, encoding a peptide of 107 amino acids (Figure 3.1). TheSSH clone contains 79 bp and 74 bp of the 5’ and 3’ UTR, respectively.

```
gaagactcaaagaatggtgtgtgctgcgaatactacacttaaggttttcgcagc  61
    M S P S L S Q A K S L R L
agcaaaaccttatataaaatagtctctctctctctcgaagccaatctctctgtctccttt  121
    L P Q S I S L I P V H R R G Y A V A S D
ctctctcaagtctatctctactactattccccctgcgttacggtgagttatcgagttgctgctgat  181
    V S V R V G L G N N V G R R S G I V G
gtttcggtaaggttgagttgtaataaatgtagggctagagttggaatcgctgggaggt  241
    A E E K P V T R D G A K A Y S D W A P D
gcaagagaagaacctgtgtcacaagagatggtgcacccatcactttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Two alleles were detected in the IT96D-602 line and one in Tvu7778. The alleles differed by a single nucleotide at position 226 from the start codon (results not shown). Line IT96D-602 had four clones with an A (VuLEA5-1) and two clones with a G (VuLEA5-2) at this position. All six clones of the Tvu7778 contained a G at this position (VuLEA5-2). This G→A substitution caused an amino acid change from valine (codon GTA) to isoleucine (codon ATA) at the 76th residue of the 107 amino acid protein. Since cowpea is a diploid organism, and two alleles were detected in line IT96D-602, this line is, therefore, heterozygous for this single nucleotide polymorphism (SNP).

PCR with the LEA5 qPCR primers on genomic DNA showed the presence of an intron in the gene, when compared to RT-PCR products (results not shown). The genomic gene sequences of four clones from each cowpea line were, therefore, cloned and sequenced. Within the 415 bp genomic VuLEA5 gene sequence from the IT96D-602 line, an intron of 88 bp was identified. The Tvu7778 line had an insertion of three nucleotides in its otherwise identical intron (Figure 3.2). The same single nucleotide polymorphism (SNP) as in the two cDNA alleles was detected in the second exon of the genomic gene sequence, at position 226 from the start codon. All four genomic clones sequenced from the IT96D-602 line contained an A at this position (the VuLEA5-1 allele). No clones were therefore obtained from IT96D-602 containing the G-allele (VuLEA5-2). All four genomic clones sequenced from the Tvu7778 line contained the G-allele (VuLEA5-2) (Figure 3.2). The primer binding sites of the qPCR primers (LEA5_F and _R, Appendix 1; blue underlined in Figure 3.2) are not affected by the SNP, highlighted in yellow in Figure 3.2.
Figure 3.2 Sequence alignment of the cowpea genomic LEA5 gene of two lines, IT96D-602 and Tvu7778. The intron is underlined and indicated in red. Nucleotide differences are indicated with yellow highlighting. The binding sites of the LEA5 primers used for qPCR studies are underlined in blue.

The VuLEA5-1 coding sequence from the IT96D-602 cowpea line, containing an A at position 226, was chosen for further analysis and sequence comparisons. It corresponds to the SSH library clone 07F09-F [GR942575] shown to be drought-responsive (Chapter 2, Table 2.2).

The deduced amino acid sequence of VuLEA5-1 was a polypeptide with a predicted molecular mass of 11.7 kDa, with an isoelectric point (pI) at 9.98 (http://web.expasy.org) (Appendix 3). LEA5 genes from A. thaliana also have basic predicted pI values (Bies-Ethève et al., 2008). VuLEA5-1 contained no cysteine residues, but were high in leucine, proline and arginine (9.35% each), with the highest amino acid content being serine (10.28%).

VuLEA5-1 had a putative mitochondrial targeting peptide (mTP) located at the N-terminus, predicted by the iPSORT algorithm (Bannai et al., 2002; http://ipsort.hgc.jp/). No signal sequence was detected. A significant LEA_3 Pfam domain (PF03242), found in LEA proteins, was identified in the VuLEA5-1 sequence (http://pfam.sanger.ac.uk). The Pfam graphic (Figure 3.3) shows the significant matches of the VuLEA5-1 sequence to the LEA_3 Pfam family.
Figure 3.3 Significant match of the VuLEA5-1 protein sequence to the LEA_3 Pfam family. #HMM is the consensus of the HMM (hidden Markov models of a small set of representative members of the family), capital letters indicate the most conserved positions; #MATCH is the match between the query sequence and the HMM. A '+' indicates a positive score, which can be interpreted as a conservative substitution. #PP: posterior probability, or the degree of confidence in each individual aligned residue. 0 means 0-5%, 1 means 5-15% and so on; 9 means 85-95% and a '*' means 95-100% posterior probability. #SEQ: query sequence, coloured according to the posterior probability (green = 100%)

The VuLEA5-1 protein’s predicted hydropathy is hydrophobic for the N-terminal half and more hydrophilic at the C-terminus (Kyte and Doolittle, 1982) (Figure 3.4). The VuLEA5-1 protein secondary structure was predicted with PSIPRED to contain three areas of α-helices (Figure 3.5) (http://bioinf.cs.ucl.ac.uk/psipred).
Figure 3.4 Hydropathicity plot of the amino acid sequence of the VuLEA5-1 protein. The Kyte and Doolittle algorithm (Kyte and Doolittle, 1982) was used with a window size of 7 amino acids.

Figure 3.5 Predicted secondary structure of the VuLEA5-1 protein. PSIPRED was used to predict the secondary structure (http://bioinf.cs.ucl.ac.uk/psipred).
The amino acid sequence of VuLEA5-1 was most closely related to LEA5 from soybean (*Glycine max*) [GenBank: AAB38782], with 81% identity in 100 overlapping amino acids using LALIGN (http://www.ch.embnet.org). Further homology was found with other LEAs of many other plants, using a BLASTP search of the protein databases at GenBank using the translated *VuLEA5-1* amino acid sequence. The soybean and six other homologous plant LEA proteins (peanut, *Populus*, citrus, castor bean, cotton and *Arabidopsis*) were aligned to *VuLEA5-1* using ClustalW (Figure 3.6).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Amino Acid Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cowpea</td>
<td>MSPSL-QAKSLRLLLPSIS-LIPVHRGGVAVSD--SVRVVGNV</td>
</tr>
<tr>
<td>Soybean</td>
<td>MARSLSQAKRIGVLQVSIS-LIPVHRGGVAVSD--SVRVVGLGNI</td>
</tr>
<tr>
<td>Peanut</td>
<td>MARSLSQASRLGALVAEASSSLIPLRRGYAYAVSD--AS--</td>
</tr>
<tr>
<td>Populus</td>
<td>MARSLSL-NAKLVLASLADGLS--LSPVFRGGYAAAPIASAVTASFGGG</td>
</tr>
<tr>
<td>Citrus</td>
<td>MARSLSKAKLLLAIVYDGGIS--LSISRRGYYAAD--LGTIS--</td>
</tr>
<tr>
<td>Castor bean</td>
<td>MARSLS-YYVKK---------VDLS-LPFLRGYSATAH-EAAARGGG--</td>
</tr>
<tr>
<td>Cotton</td>
<td>MARSLSFSPKFLGASVFDALY--VSRISRSYSGAPP--AATASFF--</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>MAARSLSAVKSLCSAASSSLS-CISVLRRSYYATSQVTAAGLSK----</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>Amino Acid Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cowpea</td>
<td>GRRSGIVGGAEEKPVTRDGAKA--------YS-DWADPITGDYRPINTPEID</td>
</tr>
<tr>
<td>Soybean</td>
<td>GRRNGIVGGVEEKKDTRDGSKA--------YSTDWADPVTGYRPINTPEID</td>
</tr>
<tr>
<td>Peanut</td>
<td>NSIVEGKGCGGVR----------DWPASGYRPINTPEID</td>
</tr>
<tr>
<td>Populus</td>
<td>SRTSAMTKMEDGAVAKEDSEA--------YSAWDPVBTGYRPANYVAED</td>
</tr>
<tr>
<td>Citrus</td>
<td>RTGIMEKNDLRPAVREDSGA----------SSAWDPVYTGPYRPENRAEID</td>
</tr>
<tr>
<td>Castor bean</td>
<td>FRSLGVGEQRAAAIKEDSGA-------CISAWDPVBTGYRPANRAEID</td>
</tr>
<tr>
<td>Cotton</td>
<td>CPQPAMGKVERRYKESSESTREYSSAWDPVBTGYRPENCAGAEID</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>GSTRVMKEMQRGLEDQAESEA---------WPDPVTGYRPENRAEID</td>
</tr>
</tbody>
</table>

**Figure 3.6 Amino acid sequence alignment of the cowpea VuLEA5-1 protein and its closest orthologs in soybean, peanut, *Populus*, citrus, castor bean, cotton and *Arabidopsis*.** The sequences used for the alignment and their GenBank references were the following: Cowpea (*Vigna unguiculata*) LEA5 [GR942575]; Soybean (*Glycine max*) desiccation protective protein LEA5 [AAB38782]; Peanut (*Arachis hypogaea*) lea protein 3 [AAZ20280]; Populus (*Populus trichocarpa*) predicted protein [XP_002315512]; Citrus (*Citrus sinensis*) Late embryogenesis abundant protein LEA5 [Q39644]; Castor bean (*Ricinus communis*) putative late embryogenesis abundant protein LEA5 [XP_002516942]; Cotton (*Gossypium hirsutum*) LEA5-A [P46521.1]; and Arabidopsis (*A. thaliana*) AtDI21 [NP_193326.4].
A conserved C-terminal sequence motif (WAPDP(V/I)TGYYRP) was observed in the aligned amino acid sequences. This motif is typical for this family of proteins and is not found in other classes of LEA proteins (Jain and Minocha, 2000). The isoleucine at the SNP position 76 of the VuLEA5-1 allele was conserved only in proteins from citrus (Citrus sinensis [Q39644] and Citrus unshiu [ABD93882]), whereas all the other selected plant LEA5 proteins from the top 20 BLASTP hits (not shown) contained a valine, as in the VuLEA5-2 cowpea allele. The amino acid at this position for the A. thaliana protein orthologs of VuLEA5 was, however, not conserved (Figure 3.7). It was substituted by valine (V, hydrophobic) in the A. thaliana AtDI21 gene (At4g15910; NP_193326.4), a lysine (K, basic) in both LEA 3-like protein (At1g02820; NP_171781) and SAG21 (At4g02380; NP_567231) and glutamine (Q, neutral polar) in the LEA3 family protein (At3g53770, NP_190945).

Figure 3.7 Amino acid sequence alignment of the cowpea VuLEA5-1 protein and its four LEA5 protein orthologs in A. thaliana. The gene ID and GenBank reference numbers of the sequences from A. thaliana were the following: AtDI21: At4g15910 [NP_193326.4]; LEA3-like protein: At1g02820 [NP_171781]; SAG21: At4g02380 [NP_567231]; and LEA3 family protein: At3g53770 [NP_190945].

The phylogenetic relationship between group 5 LEA proteins, obtained from the literature and GenBank, was analysed. A rooted phylogram was constructed with VuLEA5-1 and other group 5 LEA protein sequences (Figure 3.8). The out group was a group 4 LEA protein from Solanum lycopersicum, le25 (AAA34172; Cohen and Bray, 1992).
Figure 3.8 Phylogram of group 5 LEA proteins including VuLEA5-1 from *Vigna unguiculata* (IT96D-602). Amino acid sequences from LEA5 proteins were aligned with ClustalW and the phylogram generated with the MEGA4.0 program (Tamura *et al*., 2007). The evolutionary history was inferred using the Neighbor Joining method. The GenBank accession numbers and a shortened protein name are indicated together with the plant species source in brackets. The VuLEA5-1 protein is indicated by an arrow. Numbers at nodes indicate the percentage bootstrap scores (the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test, 1000 replications). Only values higher than 60% are shown. The evolutionary distances were computed using the Poisson correction method. The bar represents the branch length equivalent to 0.2 amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 302 positions in the final dataset. The GenBank accession numbers represent the following proteins:

<table>
<thead>
<tr>
<th>GenBank accession</th>
<th>Gene and plant species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA34172</td>
<td>le25 (<em>Solanum lycopersicum</em>)</td>
<td>Cohen and Bray, 1992</td>
</tr>
<tr>
<td>AAA34193</td>
<td>ORF (<em>Solanum lycopersicum</em>)</td>
<td>Maitra and Cushman, 1994</td>
</tr>
<tr>
<td>AAA61564</td>
<td>putative desiccation protectant protein, homolog of LEA14 (<em>Glycine max</em>)</td>
<td>Piatkowski <em>et al</em>., 1990</td>
</tr>
<tr>
<td>AAA63615</td>
<td>pcC27-45 dessication-related protein (<em>Craterostigma plantagineum</em>)</td>
<td>Dong and Dunstan, 1996</td>
</tr>
<tr>
<td>AAB01550</td>
<td>late embryogenesis abundant protein (<em>Picea glauca</em>)</td>
<td>Burns <em>et al</em>., 1996</td>
</tr>
<tr>
<td>AAB38782</td>
<td>LEA5 (<em>Glycine max</em>)</td>
<td>Zegzouti <em>et al</em>., 1997</td>
</tr>
<tr>
<td>AAB96796</td>
<td>ethylene-responsive late embryogenesis-like protein ER5 (<em>Solanum lycopersicum</em>)</td>
<td>Dong and Dunstan, 1996</td>
</tr>
<tr>
<td>AAC06242</td>
<td>late embryogenesis abundant protein 5 (<em>Nicotiana tabacum</em>)</td>
<td>Cohen and Bray, 1992</td>
</tr>
<tr>
<td>AAD53078</td>
<td>water stress-induced ER5 protein (<em>Capsicum annuum</em>)</td>
<td>Kim <em>et al</em>., 2005</td>
</tr>
<tr>
<td>AAF21311</td>
<td>GmPM25 seed maturation protein PM25 (<em>Glycine max</em>)</td>
<td>Neuteboom <em>et al</em>., 2002</td>
</tr>
<tr>
<td>AAM28284</td>
<td>late-embryogenesis abundant protein-like protein (<em>Ananas comosus</em>) (pineapple)</td>
<td>Park <em>et al</em>., 2003</td>
</tr>
<tr>
<td>AAN38002</td>
<td>LEA1-like (<em>Capsicum annuum</em>)</td>
<td>Liu R <em>et al</em>., 2010</td>
</tr>
<tr>
<td>AAU87300</td>
<td>LEA (<em>Pinus halepensis</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>AAW31666</td>
<td>putative late-embryogenesis protein-like protein (<em>Ammopiptanthus mongolicus</em>)</td>
<td>Neuteboom <em>et al</em>., 2002</td>
</tr>
<tr>
<td>AAZ20280</td>
<td>lea protein 3 (<em>Arachis hypogaea</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABB16353</td>
<td>MtPM25 seed maturation protein (<em>Medicago truncatula</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABB87099</td>
<td>unknown (<em>Solanum tuberosum</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABC41130</td>
<td>LEA5 protein (<em>Arachis hypogaea</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABD93882</td>
<td>group 5 late embryogenesis abundant protein (<em>Citrus unshiu</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABF29697</td>
<td>late embryogenesis abundant protein 5 (<em>Populus suaveolens</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABF96460</td>
<td>Late embryogenesis abundant protein (<em>Oryza sativa Japonica Group</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>ABG54481</td>
<td>late embryogenesis abundant protein (<em>Tamarix androssowii</em>)</td>
<td>Senthil-Kumar <em>et al</em>., 2007</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>ABG66530</td>
<td>putative late embryogenesis abundant protein (<em>Prosopis juliflora</em>)</td>
<td>George et al., 2007</td>
</tr>
<tr>
<td>ACJ46652</td>
<td>intrinsically disordered protein 1 (<em>Lotus japonicus</em>)</td>
<td>Haaning et al., 2008</td>
</tr>
<tr>
<td>ACV50424</td>
<td>late embryogenesis protein-5 (<em>Jatropha curcas</em>)</td>
<td>Eswaran et al., 2010</td>
</tr>
<tr>
<td>ADM33676</td>
<td>lbLEA14 late embryogenesis abundant protein 14 (<em>Ipomoea batatas</em>)</td>
<td>Park et al., 2011</td>
</tr>
<tr>
<td>ADP23916</td>
<td>late embryogenesis abundant protein (<em>Sesuvium portulacastrum</em>)</td>
<td></td>
</tr>
<tr>
<td>ADQ91844</td>
<td>late embryogenesis abundant protein group 6 protein (<em>Arachis hypogaea</em>)</td>
<td></td>
</tr>
<tr>
<td>ADW66129</td>
<td>late embryogenesis abundant protein LEA5 (<em>Solanum nigrum</em>)</td>
<td></td>
</tr>
<tr>
<td>BAJ34567</td>
<td>unnamed protein product (<em>Thellungiella halophila</em>)</td>
<td></td>
</tr>
<tr>
<td>CAA66948</td>
<td>protein induced upon tuberization (<em>Solanum demissum</em>)</td>
<td>Jackson et al., 1997</td>
</tr>
<tr>
<td>NP_001152662</td>
<td>late embryogenesis abundant protein (<em>Zea mays</em>)</td>
<td>Alexandrov et al., 2009</td>
</tr>
<tr>
<td>NP_171781</td>
<td>late embryogenesis abundant 3-like protein (<em>Arabidopsis thaliana</em>) (At1g02820)</td>
<td></td>
</tr>
<tr>
<td>NP_188888</td>
<td>Atrab28 seed maturation protein (<em>Arabidopsis thaliana</em>)</td>
<td>Arenas-Mena et al., 1999; Borrell et al., 2002</td>
</tr>
<tr>
<td>NP_190945</td>
<td>late embryogenesis abundant 3 (LEA3) family protein (<em>Arabidopsis thaliana</em>) (AT3G53770)</td>
<td></td>
</tr>
<tr>
<td>NP_193326.4</td>
<td>drought-induced 21 protein (<em>Arabidopsis thaliana</em>) (AT4G15910)</td>
<td>Gosti et al., 1995</td>
</tr>
<tr>
<td>NP_567231</td>
<td>senescence-associated protein SAG21 (<em>Arabidopsis thaliana</em>) (AT4G02380)</td>
<td>Miller et al., 1999</td>
</tr>
<tr>
<td>O03983</td>
<td>Probable desiccation-related protein LEA14 (<em>Arabidopsis thaliana</em>) (At1g01470)</td>
<td>Singh et al., 2005</td>
</tr>
<tr>
<td>P09444</td>
<td>LEA D-34 (<em>Gossypium hirsutum</em>)</td>
<td>Baker et al., 1988</td>
</tr>
<tr>
<td>P46518</td>
<td>LEA14-A (<em>Gossypium hirsutum</em>)</td>
<td>Galau et al., 1993</td>
</tr>
<tr>
<td>P46521</td>
<td>LEA5-A (<em>Gossypium hirsutum</em>)</td>
<td>Galau et al., 1993</td>
</tr>
<tr>
<td>P46522</td>
<td>LEA5-D (<em>Gossypium hirsutum</em>)</td>
<td>Galau et al., 1993</td>
</tr>
<tr>
<td>Q39644</td>
<td>Late embryogenesis abundant protein LEA5 (<em>Citrus sinensis</em>)</td>
<td>Naot et al., 1995</td>
</tr>
<tr>
<td>XP_002270639</td>
<td>hypothetical protein isoform 1 (<em>Vitis vinifera</em>)</td>
<td></td>
</tr>
<tr>
<td>XP_002315512</td>
<td>predicted protein (<em>Populus trichocarpa</em>)</td>
<td></td>
</tr>
<tr>
<td>XP_002516942</td>
<td>Late embryogenesis abundant protein LEA5, putative (<em>Ricinus communis</em>)</td>
<td></td>
</tr>
</tbody>
</table>

The Poisson correction model of distance estimates assumes equality of substitution rates among sites and equal amino acid frequencies while correcting for multiple substitutions at the same site. Proteins were resolved into the three types, D-34, D-73 and D-95, that were grouped together into group 5 based on their atypical hydrophobicity by Battaglia et al. (2008) (Figure 3.8).
3.4.2 Diversity of cowpea \textit{VuLEA5} genes

The diversity of the \textit{VuLEA5} gene was studied by comparing EST sequences from different cowpea libraries deposited into GenBank. The libraries were prepared from developing seeds (Pitiuba and line 1393-2-11), mixed tissue and conditions (IT84S-2049, IT97K-461-4, UCR 41, UCR 707, UCR 779 and 524B) and leaf and shoot meristem from mature, pre-flowering tissues of a mix of cultivars (Dan Ila, Tvu11986, Vu7778 and 12008D).

Nucleotide sequence alignment was performed and showed that the gene sequence was highly conserved between all cowpea ESTs. Phylogenetic analysis grouped the \textit{VuLEA5-1} gene identified from IT96D-602 with ESTs from Pitiuba developing seeds and UCR 41 mixed tissue and conditions (Figure 3.9). These three contained the \textit{VuLEA5-1} allele, coding for an isoleucine at position 76, with the rest having the \textit{VuLEA5-2} allele (coding for valine) identified from Tvu7778. A dinucleotide insertion (AC) was also discovered 34 nucleotides upstream of the start codon in all the lines containing the \textit{VuLEA5-1} allele (results not shown).

Only one line contained a few EST submissions with different codons at the position coding for the 76th amino acid. It was line IT97K-461-4, with a changed codon of CCA coding for a Proline, GCC coding for an Alanine or GTC that still coded for a Valine due to 3’ codon wobble (GenBank accessions FG911086, FG910995 and FG910781, respectively) (Figure 3.10).

The translated coding sequences of selected cowpea EST sequences are aligned in Figure 3.10. Amino acid alignment identified only five further amino acid substitutions in the gene’s coding sequence, apart from the differences at position 76 (Figure 3.10). One was from UCR 707 (at position 33; Serine to Alanine change) and the rest were all from IT97K-461-4 (positions 77, 85, 86 and 101 were affected). One EST (FG889157) contained three of the six substitutions contained in this line, explaining its separation from the rest in the phylogenetic tree (Figure 3.9).
Figure 3.9 Phylogenetic tree of cowpea ESTs containing the *VuLEA5* gene, based on the coding nucleotide sequences. *VuLEA5* genes from different cowpea EST libraries were aligned with ClustalW and the phylogenetic tree generated with the MEGA4.0 program (Tamura et al., 2007). The evolutionary history was inferred using the Neighbor Joining method. Numbers at nodes indicate the percentage bootstrap scores (the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test, 1000 replications). Only values higher than 60% are shown. The evolutionary distances were computed as the proportion \( p \) of nucleotide sites at which two sequences being compared are different. Codon positions included were 1st+2nd+3rd+Noncoding. The bar represents the branch length equivalent to 0.001 base substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 324 positions in the final dataset. The GenBank accession numbers are indicated with the cowpea line in brackets.
3.4.3 The expression of *VuLEA5* in response to drought and other abiotic stresses

The expression patterns of the *VuLEA5* gene in cowpea in response to drought stress and other abiotic stresses were analysed by RT-qPCR. The LEA5 qPCR primers did not discriminate between the two *VuLEA5* alleles found in cowpea.

3.4.3.1 Expression of *VuLEA5* in two cowpea lines differing in drought tolerance

Greenhouse-grown cowpea plants were drought stressed by withholding watering, while controls were watered as usual. The two lines (Tv7778 and IT96D) responded to the water stress by reducing their relative water content (RWC) to different extents. The RWC of IT96D-602 decreased less substantially than Tv7778 (Figure 3.11, 77% compared to 62% at 12 days without water). Statistical analysis of RWC percentage differences was performed.
using pair-wise comparisons of the means using Student’s T-test at the 99% confidence level. The RWC percentage of stressed Tvu7778 was significantly lower than stressed IT96D-602 at 9 and 12 days without water. Both lines recovered to near normal values three days after re-watering (Figure 3.11, 86% and 82%, respectively, day 15). Normal values ranged between 88-92% for IT96D-602 and 85-89% for Tvu7778. There was no statistically significant difference in the RWC percentages between control and stressed leaves from both lines.

![Figure 3.11 Relative water content (RWC) of cowpea leaves of two lines during a drought trial.](image)

RNA was isolated from leaf samples taken from control and stressed plants at 9 and 12 days without water. There was still genomic DNA (gDNA) contamination present in RNA samples, as detected by RT-PCR and PCR (Figure 3.12, lanes 3, 4, 11 and 12), even though the Qiagen kit claims to remove it. After DNase treatment using the Turbo DNA-free kit, no gDNA contamination could be detected in 100 ng RNA utilising PCR (Figure 3.12, lanes 13 and 14). The bean Actin primers are able to produce three fragments from cowpea gDNA during PCR (Figure 3.12, lanes 1, 2, 9 and 10). The bean Actin cDNA fragment (170 bp) was strongly amplified during RT-PCR (Figure 3.12, lanes 3-6). The genomic fragment is larger (280 bp) due to the presence of an intron.
Figure 3.12 DNase treatment removes contaminating gDNA from RNA isolated from cowpea using the Plant RNeasy kit (Qiagen). RT-PCR and PCR was performed on gDNA and 100 ng RNA before and after treatment with Turbo DNA-free. Lanes 1-8: RT-PCR and lanes 9-16: PCR reactions of, respectively, gDNA of IT96D-602 (lanes 1, 9) and Tvu7778 (lanes 2, 10); 100 ng untreated RNA of IT96D-602 (lanes 3, 11) and Tvu7778 (lanes 4, 12); 100 ng DNase-treated RNA of IT96D-602 (lanes 5, 13) and Tvu7778 (lanes 6, 14); water control (lanes 7, 15); 200 ng untreated RNA of IT96D-602 (lanes 8, 16).

The two lines (Tvu7778 and IT96D) were assayed for *VuLEA5* gene expression levels using RT-qPCR. Three biological replicates with two technical replicates each were run on 100 ng of the respective RNA samples. RT-qPCR data was analysed with the software program qBase (Hellemans *et al.*, 2007). qBase is relative quantification software from Ghent University for the management and automated analysis of real-time quantitative PCR data. Technical replicates of the three biological replicate samples were pooled for data analysis. Gene expression levels were normalised with the *GAPC* (Glyceraldehyde dehydrogenase C-subunit) reference gene and are presented in Figure 3.13. *VuLEA5* mRNA levels increased in both the drought tolerant and susceptible lines of cowpea after 9 days of drought stress of cowpea plants in a greenhouse (Figure 3.13, compare S9 to C9). Statistical analysis of expression differences was performed using pair-wise comparisons of the means using Student’s T-test at the 99% confidence level. The susceptible line Tvu7778 had significantly higher *VuLEA5* expression than IT96D-602 at 12 days stress treatment, but not at 9 days (Figure 3.13, S12).
3.4.3.2 Optimisation of cowpea reference genes

Two reference genes for RT-qPCR of cowpea samples were added in addition to GAPC previously used (Chapter 2 and Figure 3.13). They were VuELF1B and Gm40S. The soybean GmELF1B gene [TC203623; MGI (Medicago gene index database)] was indicated to be a very stable reference gene during different developmental stages, tissues, photoperiodic treatments etc. (Jian et al., 2008). Since the GmELF1B primers (Appendix 1) did not yield any product with cowpea gDNA or RNA (results not shown), a new primer pair VuELF1B (Appendix 1) was designed from a cowpea ELF1B sequence [GenBank: AB588747]. A 170 bp amplicon was expected from both gDNA and RNA, since the gene contains no intron. RT-PCR yielded the expected 170 bp fragment, but a larger fragment was amplified from the genomic DNA, which means that the gene did contain an intron (results not shown). The soybean 40S ribosomal protein S8 [TC100533; MGI; Mortier et al., 2010] primers (Gm40S; Appendix 1) were able to amplify a 200 bp fragment from both RNA and gDNA (results not shown).
Reference gene stability was analysed with qBase (Hellemans et al., 2007), and showed the Gm40S gene to be unreliable as a reference gene (results not shown). The amplification efficiency of the Gm40S primers was inadequate, with an E-value of 1.49 (Table 3.1). The gene expression values of cowpea leaves treated *in vitro* with abiotic stresses were, therefore, normalised with the remaining two reference genes, GAPC and VuELF1B, which showed satisfactory M and CV values (smaller than 0.5 and 0.2, respectively).

### 3.4.3.3 Expression of *VuLEA5* in cowpea under abiotic stress (*in vitro*)

*VuLEA5* gene expression in cowpea in response to other abiotic stresses was analysed with RT-qPCR. Detached leaves of cowpea IT96D-602 plants retained 87±2% of their original leaf mass after 17 h of desiccation and were fully hydrated to 111±5% of their original leaf mass after 24 h rehydration (results not shown).

RT-qPCR was performed on RNA isolated from the treated leaves. The effectiveness of the *in vitro* abiotic stress treatments was shown by analysing the expression of cowpea *NCED1* (9-cis-epoxycarotenoid dioxygenase; AB030293), a drought inducible gene (Iuchi et al., 2000). *NCED1* codes for a neoxanthin cleavage enzyme involved in ABA biosynthesis under water stress in drought tolerant cowpea. Primers were designed from the coding region. The GAPC, VuELF1B and Gm40S reference genes were also included (Appendix 1). Standard curves were generated for each primer pair, using cowpea RNA as template. They were used to calculate the amplification efficiency (E-values) of the primer pairs (Table 3.1) used during the conversion of C_q values to the amount of input RNA in nanogram (ng). The standard curves were highly linear as indicated by the R^2 values of the linear regression curves. The primers yielded E-values ranging from 0.99 (LEA5) to 1.49 (Gm40S). A value of 1 indicates optimal amplification with a doubling of DNA molecules at every cycle.

#### Table 3.1 Amplification efficiency of primer pairs.

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>E-value</th>
<th>R^2 value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEA5</td>
<td>0.99</td>
<td>0.99824</td>
</tr>
<tr>
<td>NCED</td>
<td>1.21</td>
<td>0.99289</td>
</tr>
<tr>
<td>Gm40S</td>
<td>1.49</td>
<td>0.96784</td>
</tr>
<tr>
<td>VuELF1B</td>
<td>1.22</td>
<td>0.99963</td>
</tr>
<tr>
<td>GAPC</td>
<td>1.05</td>
<td>0.99650</td>
</tr>
</tbody>
</table>

© University of Pretoria
A significant increase in \textit{VuLEA5} expression was observed after leaves from line IT96D-602 were desiccated for 17 h, whereafter \textit{VuLEA5} expression levels dropped after the leaf was rehydrated in water for 24 h (Figure 3.14). A significant increase in \textit{NCED1} expression was also observed after 17 h desiccation, with levels also dropping after 24 h rehydration in water. This is characteristic of many other drought-responsive genes.

![Figure 3.14](image)

**Figure 3.14** Cowpea \textit{LEA5 (VuLEA5)} and \textit{NCED1} expression in IT96D-602 leaves after desiccation and rehydration. Normalised gene expression levels before and after desiccating for 17 h, and 24 h rehydration after 17 h desiccation for \textit{VuLEA5} (green) and \textit{NCED1} (orange, hatched). The error bars indicate the standard deviation of two biological replicates of each treatment. Gene expression values were normalised with the GAPC and \textit{VuELF1B} reference genes.

Both \textit{VuLEA5} and \textit{NCED1} showed increased gene expression following treatment of the detached leaf with ABA, NaCl and cold (Figure 3.15 and 3.16). Cold treatment had the most pronounced effect on the expression of both genes. Gene expression differences of both the \textit{VuLEA5} and \textit{NCED1} genes between the 0 h desiccated or water controls and all the different treatments (desiccation, ABA, NaCl and cold) were statistically significant at the 99% confidence level using pair-wise comparisons of the means using Student's T-test. Only the 5 h ABA (100 \mu M) treatment (Figure 3.15) did not result in statistically significant changes in gene expression of the \textit{VuLEA5} gene compared to the water control.
Figure 3.15 Cowpea \textit{LEA5} (\textit{VuLEA5}) expression in IT96D-602 leaves after 5 h and 24 h treatment with ABA, NaCl or cold. Normalised gene expression levels after treatment with water or ABA (100 \(\mu\)M), NaCl (250 mM) and cold (4\(^\circ\)C) for 5 h (solid green) or 24 h (hatched green). The error bars indicate the standard deviation of two biological replicates of each treatment. Gene expression values were normalised with the GAPC and VuELF1B reference genes.

Figure 3.16 Cowpea \textit{NCED1} expression in IT96D-602 leaves after 5 h and 24 h treatment with ABA, NaCl or cold. Normalised gene expression levels after treatment with water or ABA (100 \(\mu\)M), NaCl (250 mM) and cold (4\(^\circ\)C) for 5 h (solid orange) or 24 h (hatched orange). The error bars indicate the standard deviation of two biological replicates of each treatment. Gene expression values were normalised with the GAPC and VuELF1B reference genes.
3.5 Discussion

3.5.1 Properties and diversity of the *VuLEA5* gene

The *VuLEA5* gene identified during this project is most similar to a LEA5 protein gene identified in desiccating seeds of soybean (*Glycine max*) [GenBank: GMU66316] (Burns *et al.*, 1996) ([Chapter 2, Table 2.2](#)). It also corresponds closely to the drought-induced LEA5 (D-73) cDNA identified in cotton (Galau *et al.*, 1993). BLASTX analysis indicated the closest protein homolog to be the soybean LEA5 protein [GenBank: AAB38782] ([Figure 3.6](#)). The *VuLEA5* protein contains the characteristic LEA_3 domain (Pfam family PF03242; [http://www.sanger.ac.uk/Software/Pfam](http://www.sanger.ac.uk/Software/Pfam)) ([Figure 3.3](#)).

As mentioned before, Battaglia *et al.* (2008) grouped the three families, D-34, D-73 and D-95, together into group 5 based on their atypical hydrophobicity. The group in which the *VuLEA5* gene resides has been classified differently by different authors in the past. Some authors look for conserved and repeated motifs, and places the D-73 type *VuLEA5* into group 6 (Bies-Ethève *et al.*, 2008), while splitting it from group 7 that contains the D-95 type. Others group on their content of Pfam domains, whereby the D-73 type *VuLEA5* would fall into the LEA_3 group ([Chapter 1, Figure 1.3](#); Hundertmark and Hincha, 2008). The D-95 type contains the LEA_2 Pfam domain, while D-34 falls into a group called SMP for seed maturation protein.

Since the three types, D-34, D-73 and D-95, were clearly distinguished on the basis of their amino acid alignments in the phylogram in [Figure 3.8](#), it is suggested that these types remain separated. Their Pfam domain content is suggested as classifying criteria. They all share an atypical hydrophobicity, and probably arose from a common ancestral group 5 gene sequence (Battaglia *et al.*, 2008). Because of this, this group may be considered not to belong to LEA protein family (Tunnacliffe and Wise, 2007).

Hydropathy plots enable the hydrophobicity of a peptide sequence to be visualised along the length of the sequence. A moving window of 7 to 11 residues is suggested by Kyte and Doolittle (1982). Negative values in a hydrophathy plot represent spans of hydrophilic amino acids. The plots of LEA proteins from groups 1 to 3 were found to be almost entirely below the zero line (Dure *et al.*, 1989). The predicted hydropathy, with a window size of 7 amino acids ([Figure 3.4](#)) (Kyte and Doolittle, 1982), of the VuLEA5-1 protein was, therefore, much more hydrophobic than typical LEA proteins. For group 2 and 3 LEA proteins identified from cotton (D-11 and D-29), a hydrophobic C-terminal region was noticed. This is proposed to anchor the protein to hydrophobic surfaces in the cytosol (Baker *et al.*, 1988).
A mitochondrial targeting peptide (mTP) was discovered in the VuLEA5-1 protein sequence, predicted by the iPSORT algorithm (Bannai et al., 2002; http://ipsort.hgc.jp/). Other authors have also found that some LEA proteins are targeted to chloroplasts and mitochondria (Tunnacliffe and Wise, 2007; Hundertmark and Hincha, 2008).

The hydrophobic nature of the N-terminal half of the VuLEA5-1 protein (Figure 3.4) is correlated with its predicted secondary structure containing α-helices (Figure 3.5). Since water is hydrophilic, it excludes the hydrophobic amino acid residues and causes them to interact with each other. The hydrophobic interactions stabilise these helices in solution, to enable the formation of a stable three-dimensional structure that was elucidated for A. thaliana LEA14 [AT1G01470; GenBank: O03983], a LEA5 protein of the D-95 type. In addition to one α-helix, this structure also contained seven β-strands that form two antiparallel β-sheets (Singh et al., 2005). This protein is slightly larger than VuLEA5-1 (151 compared to 107 amino acids). The structure was also determined for MtPM25 [GenBank: ABB16353] from M. truncatula, another group 5 protein of the D-34 type and much larger than VuLEA5-1 (260 amino acids). It was found to form α-helices and β-sheets upon drying by removing water (Boudet et al., 2006).

From the hydrophobic nature of the VuLEA5-1 protein, and the likelihood to form stable three-dimensional structures in solution, but especially upon cell drying, it is proposed that it may share the chaperone activity of a related protein in Lotus japonicus (Haaning et al., 2008; Figure 3.8; GenBank: ACJ46652), to prevent protein misfolding and preserving enzyme activities during abiotic stress conditions.

A LEA protein related to LEA5 was used to root the phylogram of LEA5 proteins (Figure 3.8) It was the le25 protein from Solanum lycopersicum (AAA34172; Cohen and Bray, 1992). Both le25 and a related hydrophobic LEA-like protein from rice (AAD10377, not shown) are reported to be relatively hydrophobic. The le25 protein is, however, categorised into group 4 and corresponds to the cotton D-113 LEA gene (Wise, 2003). It formed an out group when the phylogenetic relationship between group 5 LEA proteins was studied (Figure 3.8). Out groups are useful to polarise the phylogram and to give confidence that the D-73, D-34 and D-95 types of LEA5 are, in fact, sister groupings.

The VuLEA5 gene was cloned from both cowpea lines used in this study (Tvu7778 and IT96D-602) to detect whether there are sequence differences between the two lines that may result in different tolerances to drought stress. In the process, two alleles were discovered, differing by a SNP in the coding region and a triplet insertion in the intron. The SNP results in
the substitution of an isoleucine (VuLEA5-1, line IT96D-602) for valine (VuLEA5-2, line Tvu7778) at residue 76 of the 107 amino acid protein. Since this is a conservative substitution, it might not have any effect on the protein’s function.

To determine the possible significance of this SNP in protein function, the conservation at this site (position 76 of the protein) was investigated for LEA5 proteins from other plants and other available cowpea VuLEA5 ESTs in the GenBank sequence database.

Tao et al. (2006) noted the conserved motif (W(A/V)PDP(V/I)TGYYRPE) among different group 5 LEA proteins when they compared citrus LEA protein sequences to other plant proteins from the LEA5 gene family. Citrus LEA5 proteins contained the WAPDPITGYYRPE motif, which corresponds to the cowpea allele identified in line IT96D-602 (Tao et al., 2006; Naot et al., 1995). The rest of the sequences compared contained the W(A/V)PDPVTGYYRPE motif. The same observation was made during this study, that only citrus LEA5 proteins shared the isoleucine allele with cowpea IT96D-602, while all other proteins from the top 20 BLASTP hits contained a valine (results not shown). The conservation was, however, not absolute, since the protein orthologs of VuLEA5 in A. thaliana contained other amino acids, such as lysine and glutamine, having different physiochemical properties than isoleucine and valine.

The nucleotide and amino acid sequences of cowpea ESTs were highly conserved in a number of EST accessions obtained from nine libraries, with only six amino acid positions affected by substitutions (Figure 3.10). These positions included position 76, where the amino acids proline and alanine were found in addition to the isoleucine and valine in the alleles already discovered. The high conservation may be due to the critical nature of the amino acid sequence of the VuLEA5 gene. A more probable reason, however, is that these ESTs were cloned from cowpea lines that may be closely genetically related, perhaps originating from the same breeding parents.

VuLEA5 amino acid sequence diversity was the highest in the line IT97K-461-4. This suggests the possibility of more than two alleles of this gene. It should be taken into consideration that sequencing errors may be present in deposited ESTs, resulting in false nucleotide polymorphisms. The isoleucine allele (VuLEA5-1) was found in the cowpea lines Pitiuba (developing seed library by F.A.P Campos, Brazil, unpublished) and UCR 41 (from a single plant with a long history of single seed descent from Nicaragua). IT96D-602 was heterozygous at this locus, having both the isoleucine and valine alleles. The valine allele (VuLEA5-2), found in Tvu7778, were also found in lines IT97K-461-4 and IT84S-2049 from a
Nigerian breeding program (J.D. Ehlers, unpublished data); UCR 779 from Botswana; 524B from a California breeding program; UCR 707 from Kenya; a 1393-2-11 developing seed library and an EST library prepared from a mixed pool of cowpea lines Dan Ila, Tvu11986, Vu7778 and 12008D. It is not known how all these lines are related, but it seems as if the isoleucine allele (VuLEA5-1) is from South America and valine (VuLEA5-2) is Africa-specific. Since this SNP occurs in a reasonably variable region of the gene, it might not be under stringent selection pressure, and therefore the possible change in protein properties is not affecting the protein function.

### 3.5.2 Expression of the VuLEA5 gene under abiotic stress conditions

VuLEA5 expression was higher in the susceptible line Tvu7778 at 12 days without watering than the tolerant line IT96D-602 (Figure 3.13). This is correlated to the fact that this line experienced a higher water deficit stress, indicated by a lower RWC of Tvu7778 at this stage of stress (Figure 3.11). The expression of protective proteins, such as LEAs, is expected in the type 2 drought tolerance response of cowpeas (Agbicodo et al., 2009; Singh and Matsui, 2002). Type 2 drought tolerant plants can withstand water deficit at low tissue water potential. The drought tolerant line IT96D-602 exhibits a type 2 response to drought (Singh and Matsui, 2002).

Various gene expression studies have been published using detached cowpea leaves treated in vitro with different abiotic stresses (desiccation, NaCl, ABA, heat and cold). They are primarily from the Paris research group of Pham-Thi and Zuily-Fodil, with publications by El Maarouf et al. (1999 and 2001); Matos et al. (2001); Cruz de Carvalho et al. (2001); Diop et al. (2004); d’Arcy-Lameta et al. (2006) and Contour-Ansel et al. (2006). Iuchi et al. (1996) first used this method to identify several drought-responsive genes, termed CPRD, from a drought tolerant cowpea cultivar. Transcripts of a dehydrin LEA (CPRD22) were shown to accumulate within 2 h of the start of dehydration, and by high-salt conditions and exogenous ABA (Iuchi et al., 1996). The experiments performed in this study were designed using information from these papers.

The expression patterns of VuLEA5 were characterised during desiccation, rehydration, cold and exogenous ABA and NaCl treatment of cowpea leaves. Desiccation, cold and NaCl stress all impose an osmotic stress on the plant cell. It is expected that overlapping genes are up-regulated under each of these abiotic stress conditions. ABA is known to be involved in drought-induced signalling pathways, leading to adaptive antioxidant processes (Zhu, 2002). Because of its accumulation during drought stress and its role in plants’ tolerance to water deficit, ABA also induces stress genes. The VuLEA5 gene’s responsiveness to all
these abiotic stress conditions provides confirmation of its putative role during drought responses.

The NCED1 (9-cis-epoxycarotenoid dioxygenase) gene [GenBank: AB030293] codes for a neoxanthin cleavage enzyme involved in ABA biosynthesis under water stress in drought-tolerant cowpea (Iuchi et al., 2000). In these studies it was used as a positive control for drought and other abiotic stress treatments in cowpea (Figure 3.14 and 3.16). Iuchi et al. (2000) found this gene to be stress-inducible, mainly by drought and high salt conditions, but not by cold, heat or exogenous ABA application (100 µM). Transcription of the M. truncatula NCED5 gene was, however, inducible by ABA (Planchet et al., 2011). They treated seedlings with a 10 µM solution of ABA. The NCED1 gene from Stylosanthes guianensis, a pasture legume, was induced by drought, dehydration, salt and cold stress, and mRNA transcript levels coincided with ABA accumulation (Yang and Guo, 2007). Under the conditions of these experiments, enhanced cowpea NCED1 gene expression was also observed after desiccation, 100 µM ABA, 250 mM NaCl and cold (4 °C) treatments.

The minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines were followed (Bustin et al., 2009). Relative quantification of gene expression was performed using the two standard curve method, one each for the gene of interest and the reference gene. This eliminates the difference in amplification efficiencies between the two genes. The standard curves were constructed from serial dilutions of template RNA, pooled from representative treated samples.

The expression of the VuLEA5 gene was revealed to increase under desiccation, ABA treatment, NaCl and low temperature conditions (Figure 3.15). This corresponds to LEA5 genes from other plants that are also stress responsive (examples from soybean (Burns et al., 1996) and cotton (Galau et al., 1993); additionally citrus LEA5 was heat- but not cold responsive (Naot et al., 1995)). Analysing a positive control gene, NCED1, alongside VuLEA5 gave the expected expression levels, providing confidence in the experimental setup (Figure 3.14 and 3.16). Both these genes exhibited the typical response to water stress, namely induction by dehydration and reduction upon rehydration. This behavior was illustrated for various novel drought-inducible genes identified from cowpea by Iuchi et al. (1996). Treating detached leaves in vitro with abiotic stresses is, therefore, a useful approach to test their effects on gene expression.
3.5.3 Expression patterns of LEA5 genes related to VuLEA5

The Genevestigator tool (http://www.genevestigator.com/gv/) was used to study the expression of the soybean LEA5 gene [GenBank: GMU66316], that is most closely related to VuLEA5. The data of hybridisation studies from samples of various tissues and treated conditions to the GM_60K Soybean Genome array can be accessed with this tool. GmLEA5 is highly up-regulated in seeds, specifically the embryo, and the germinating seedling radicle. This is consistent with the role of LEA proteins in the protection of proteins in the desiccating seeds, and the re-establishment of metabolic activity in the germinating seedling.

The VuLEA5 gene share similarities to four A. thaliana genes of the D-73 type, containing the LEA_3 Pfam domain. They are AtDI21 (A. thaliana drought-induced 21, At4g15910), LEA3-like protein (At1g02820), SAG21 (Senescence-associated gene 21, At4g02380) and LEA3 family protein (At3g53770). These four genes form one of nine groups of the fifty-one LEA protein encoding genes in the A. thaliana genome (Hundertmark and Hincha, 2008). They are expressed during various abiotic treatments and in different plant organs (Hundertmark and Hincha, 2008). It is interesting to note that the regulation of VuLEA5 corresponds closely to its relatives in A. thaliana.

AtDI21 is the closest relative of VuLEA5 in the A. thaliana LEA gene family. It is induced by cold, salt and drought, but the most by high light (Hundertmark and Hincha, 2008). AtDI21 expression in the leaves and roots was indicated to be responsive to water deprivation and exogenously applied ABA (Gosti et al., 1995). Expression of this gene during various abiotic stresses was visualised with the Arabidopsis eFPBrowser (http://bar.utoronto.ca) (Winter et al., 2007). Expression is up-regulated especially by cold, moderately by osmotic stress (300 mM mannitol) and slightly by salt (150 mM) after 24 h in an experiment performed by Kilian et al. (2007). Results from their experiment showed, however, that drought did not increase expression significantly. Roots generally had higher transcript levels than shoots. This gene is responsive to ABA, and is expressed in the shoot apex, seed and germinated seeds. Genevestigator did, however, show hybridisation of the AtDI21 probe on the AtH1: 22k GeneChip array to drought-treated samples.

SAG21 is also induced by cold, salt, ABA and during natural and ozone-induced senescence (Miller et al., 1999). This gene was identified as involved in tolerance of plants to oxidative stress. Functional cloning and screening of oxidant-sensitive yeast mutant cells (Deltayap1) with various oxidants were employed in this approach (Mowla et al., 2006). They propose SAG21 to have an important function among LEA genes to protect plants against oxidative stress involving decreased photosynthesis. They also found the gene to be up-regulated by
ABA and dehydration, in addition to the oxidants. During a microarray experiment on *A. thaliana*, Seki *et al.* (2001) found a SAG21 LEA protein homolog to be cold-inducible. Experiments by Fowler *et al.* (2002) could, however, not reproduce the cold regulation of the SAG21 gene. Proposed reasons include differences in plant culture conditions, environmental treatments or the expression-profiling methods used. The Arabidopsis eFPBrowser did not show ABA-inducibility of the gene, but showed oxidative stress (10 µM Methyl viologen) UV-B and wounding in addition to cold, osmotic stress and salt as conditions that up-regulate this gene. Both eFPB and Genevestigator confirmed mature pollen as the major plant part where this gene is expressed.

The LEA3-like protein of *A. thaliana* (At1g02820) is also induced by cold, osmotic stress and salt, and is also ABA-responsive. It is expressed mostly in immature siliques and seed, young leaf, hypocotyl, flower sepals and stamen samples. All three *A. thaliana* LEA5-related genes have cold, osmotic and salt stress-inducibility in common. The regulation of *VuLEA5* is similar, therefore the VuLEA5 protein may also be involved in the protection of cowpea against these abiotic stresses.

### 3.5.4 Related functions indicated by protein sequence similarity

Similarities in protein sequences indicate that they may have similar functions. A list of proteins closest related to VuLEA5, for which information was available, was made from the phylogram in Figure 3.8, and their putative functions, or conditions under which expression is up-regulated, are presented in Table 3.2. Multiple functions of LEA5 proteins are deduced from this table and include protection against desiccation or drought, heat, salt, cold, high light and the stress-hormone ABA.

LEA5 proteins that are more distantly related, but still containing the LEA_3 Pfam domain (PF03242), were also looked at for possible related functions. A putative late-embryogenesis protein-like protein from *Ammopiptanthus mongolicus* [Genbank: AAW31666] was able to enhance *E. coli* cell viability during heat and cold stress (Liu R *et al.*, 2010). An unusual intrinsically disordered protein from the legume *Lotus japonicus* (Figure 3.8; GenBank: ACJ46652) was able to stabilise proteins in vitro when they were dehydrated or frozen. These conditions typically lead to protein misfolding and aggregation, but this protein acted as a chaperone that prevented misfolding and preserved the activities of two model enzymes (Haaning *et al.*, 2008).

In this study, the *VuLEA5* gene was induced by various abiotic stress conditions, of which cold stress and ABA was the most active in inducing transcript accumulation. Since *VuLEA5*
expression is induced by ABA, it is deduced that *VuLEA5* responds to various abiotic stresses through the ABA-mediated stress pathway in cowpea. It is also proposed that this gene may act as protectant during low temperature stress of cowpea.

### 3.6 Acknowledgements

We acknowledge the University of Pretoria for the use of their qBase software. We thank Prof. K Künert, Department of Plant Science, University of Pretoria, for providing the soybean GmELF1B and Gm40S primers used in this study.
Table 3.2 Putative functions of LEA5 proteins closely related to VuLEA5. The GenBank accession numbers and the species source are indicated together with a shortened protein name and putative function from the literature.

<table>
<thead>
<tr>
<th>GenBank accession</th>
<th>Protein name</th>
<th>Specie source</th>
<th>Putative protein function/ expression induced by</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAB38782</td>
<td>LEA5</td>
<td>Glycine max</td>
<td>Desiccation protective protein</td>
<td>Burns et al., 1996</td>
</tr>
<tr>
<td>AAZ20280</td>
<td>lea protein 3</td>
<td>Arachis hypogaea</td>
<td>Peanut seed protein</td>
<td></td>
</tr>
<tr>
<td>ABC41130</td>
<td>LEA5 protein</td>
<td>Arachis hypogaea</td>
<td>Water deficit-induced gene</td>
<td>Senthil-Kumar et al., 2007</td>
</tr>
<tr>
<td>Q39644</td>
<td>LEA5</td>
<td>Citrus sinensis</td>
<td>Drought, heat and salt inducible</td>
<td>Naot et al., 1995</td>
</tr>
<tr>
<td>ABD93882</td>
<td>group 5 late embryogenesis abundant protein</td>
<td>Citrus unshiu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XP_002315512</td>
<td>predicted protein</td>
<td>Populus trichocarpa</td>
<td>Not determined</td>
<td></td>
</tr>
<tr>
<td>XP_002516942</td>
<td>putative LEA5</td>
<td>Ricinus communis</td>
<td>Not determined</td>
<td></td>
</tr>
<tr>
<td>P46521 and P46522</td>
<td>LEA5-A and LEA5-D</td>
<td>Gossypium hirsutum</td>
<td>Drought-inducible</td>
<td>Galau et al., 1993</td>
</tr>
<tr>
<td>BAJ34567</td>
<td>unnamed protein product</td>
<td>Thellungiella halophila</td>
<td>Not determined</td>
<td></td>
</tr>
<tr>
<td>NP_193326.4</td>
<td>drought-induced 21 protein</td>
<td>Arabidopsis thaliana</td>
<td>Inducible by cold, salt, drought, high light and ABA</td>
<td>Gosti et al., 1995</td>
</tr>
<tr>
<td>ADP23916</td>
<td>late embryogenesis abundant protein</td>
<td>Sesuvium portulacastrum</td>
<td>Not determined</td>
<td></td>
</tr>
<tr>
<td>ABG54481</td>
<td>late embryogenesis abundant protein</td>
<td>Tamarix androssowii</td>
<td>Confers drought and cold tolerance to transgenic plants (tobacco and blueberry)</td>
<td>Wang et al., 2006; Zhao et al., 2011</td>
</tr>
</tbody>
</table>

© University of Pretoria
Chapter 4

Functional evaluation of the cowpea LEA5 gene with regard to drought tolerance

4.1 Abstract
The efficacy of the cowpea LEA5 gene (VuLEA5-1) in conferring drought tolerance to a model plant was tested to determine if it is a possible target for improving crop plants using recombinant or breeding strategies. The VuLEA5-1 gene was successfully cloned into the plant transformation vector pMDC32, making use of the Gateway cloning technology from Invitrogen. Nine independent transformation events of transgenic Arabidopsis thaliana plants (ecotype Col-0) expressing 35S::VuLEA5-1 were selected until the fourth homozygous generation. Gene integration into the genome was verified in eight of the transgenic lines by PCR and Southern blot. Copy numbers varied between one and three copies. Gene expression was verified by RT-PCR and RT-qPCR. Western blotting using the α-MtPM25 antibody could not detect the expressed VuLEA5-1 protein, due to sequence differences between these two proteins. Transgenic VuLEA5-1 A. thaliana plants and seedlings did not show any improved growth or recovery, when compared to the wild-type Col-0, following drought and osmotic stress treatments.

4.2 Introduction
4.2.1 LEA genes to improve crops
LEA genes code for LEA proteins, found to accumulate in the seeds of many higher plants when the embryo acquires tolerance to desiccation (Baker et al., 1988; Dure et al., 1989). They are also expressed in the vegetative organs of many plants studied to date, especially during water deficit, suggesting an important protective role during water limitation (Battaglia et al., 2008). LEA proteins are disordered proteins that can act as protein stabilisers (Goyal et al., 2005; Haaning et al., 2008). Because they have a high proportion of hydrophilic amino acids, they can help plants to retain intracellular water during times of low availability.

The role of group 2, 3 and 4 LEA proteins in drought tolerance has been illustrated by transforming these genes into other crops (such as creeping bentgrass) or A. thaliana (Fu et al., 2007; Brini et al., 2007; Dalal et al., 2009). In these experiments, transgenic plants
generally showed better survival under drought, cold- and salt-stressed conditions, compared to non-transgenic control plants, or displayed faster recovery when transferred to normal medium.

Atypical LEA proteins, such as cotton LEA5 and LEA14, tomato ER5 and pcC27-45 from a resurrection plant are hydrophobic, and therefore may function differently from the typical hydrophilic LEA proteins. Nevertheless, they are also involved in the plant dehydration response (Galau et al., 1993; Zegzouti et al., 1997; Piatkowski et al., 1990). LEA proteins corresponding to the cotton D-34, D-73 and D-95 LEA proteins (Baker et al., 1988; Dure, 1993; Galau et al., 1993) were grouped together into group 5 based on their atypical hydrophobicity by Battaglia et al. (2008). The VuLEA5 gene identified in this study is most similar to the D-73 cDNA, responds to various dehydration-related abiotic stresses, and also codes for a hydrophobic protein (Chapter 3).

A recent literature survey revealed that only one LEA group 5 member of the D-73 type has thus far been tested for its role in adaptation to drought and cold stress conditions by transforming it into another plant. It was a LEA5 gene isolated from Tamarix androssowii, a shrub or tree able to tolerate saline soil (Wang et al., 2006). Nucleotide identity of this gene to the VuLEA5-1 gene is 55.8%, while coding for identical or similar amino acids at 70.8% of the protein residues. Transgenic tobacco were produced containing this gene, and were shown to have less cell membrane damage, a decreased number of wilted leaves and an increased height growth under drought stress (Wang et al., 2006). Transgenic blueberry containing the same gene showed enhanced cold tolerance in a recent study (Zhao et al., 2011).

The VuLEA5-1 gene identified during this study was, therefore, chosen as a candidate gene to improve the drought tolerance of crops of importance to resource poor farmers. Although it was beyond the scope of this study to improve a crop plant through genetic engineering, A. thaliana was used as a proof-of-concept of the induction of drought tolerance by the VuLEA5-1 gene. A. thaliana is useful as a model plant for studying the biological effect of transgenes, since it has a short life cycle and is readily transformed without the need for plant tissue culture (Clough and Bent, 1998).

4.2.2 Methods to assess the drought tolerance of plants

Both drought and osmotic stress treatments were applied in this chapter to study the induction of drought tolerance in transgenic A. thaliana by the VuLEA5-1 gene. Drought is defined as the sub-optimal availability of water that limits plant growth and transpiration.
(Neumann, 2008). Desiccation is an extreme form of drought (Bernacchia and Furini, 2004). The abiotic stresses of drought, salt and cold all have the same effect on plants: a reduced water potential ($\psi_W$) inside the plant cell caused by a deficit of cellular water (Verslues et al., 2006). The molecular responses of plants to these abiotic stresses are, therefore, closely related and some of the mechanisms overlap (Zhu, 2002).

Osmotic stress treatments of seedlings can simulate the effects of drought. Chemicals are used to lower the $\psi_W$ of the media. It is a useful approach in studying drought tolerance mechanisms of plants, because seedlings that are grown on a low $\psi_W$ stress media has to equilibrate their tissue $\psi_W$ to it, and therefore cannot avoid the stress.

An improvement in the abiotic stress tolerance of transgenic lines are frequently defined as having a higher fresh weight, retention of green colour or leaf development, compared to the wild type under stress and recovery conditions (Verslues et al., 2006). An approach where stomatal conductance is measured (with a Porometer or through thermal imaging) will only give information on drought avoidance mechanisms, and is not appropriate if a tolerance mechanism, such as LEA protein expression, is being studied. Following is a list of methods, described in the literature, used to assess the drought tolerance of plants or seedlings while comparing them to their wild-type controls:

1. Measurement of growth at low water potential ($\psi_W$). This can be approached through evaluating a variety of characteristics. The plant appearance, size, leaf area, fresh/ dry weight, time of flowering, relative rate of height growth, seed yield, water use efficiency or photosynthetic performance can be measured as growth indicators (Wang et al., 2006; Verslues et al., 2006).

2. Plant survival or recovery after severe stress. For this the biomass accumulation or bud recovery after rehydration (Olvera-Carrillo et al., 2010), survival of seedlings on stress medium (NaCl or mannitol) (Dai et al., 2011), or the ability to resume growth after transfer back to normal medium (Brini et al., 2007) can be measured.

3. Quantify cellular damage caused by oxidative stress. Abiotic stresses cause the accumulation of reactive oxygen species (ROS), which in turn causes lipid peroxidation. Malondialdehyde (MDA) is a product of polyunsaturated fatty acid peroxidation in plants. MDA levels during stress conditions can, therefore, be used as a marker for evaluating the stress tolerance of plants (Gawel et al., 2004; Dai et al., 2011). Oxidative stress can be imposed through e.g. Paraquat, hydrogen peroxide or Cadmium treatment. Subsequent loss of chlorophyll content, electrolyte leakage, ROS accumulation or any ROS-induced chemical changes can be measured (Verslues et al., 2006). The visual
injury of leaf disks can also be assessed by measuring the necrotic lesions (Dai et al., 2011; Rodrigues et al., 2006).

4. Changes in shoot and root growth, such as root bending, length differences and branching characteristics. Low $\psi_W$ stress tends to decrease shoot growth with root growth being equal or slightly enhanced. Salt stress, on the other hand, inhibits root growth. The root bending assay is used to evaluate whether roots are able to continue growing after transfer to a stress medium and inverting the plate, so that the roots point upwards (Zhu et al., 1998; Verslues et al., 2006; Yang et al., 2009). Root mass, root elongation and number of lateral roots can be measured (Dai et al., 2011; Oh et al., 2011; Umezawa et al., 2004).

5. Seed germination rate on stress medium (NaCl or mannitol) (Brini et al., 2007; Dai et al., 2011; Rodrigues et al., 2006; Mukhopadhyay et al., 2004).

4.2.3 Chemicals to induce low water potential ($\psi_W$) of media

The drought tolerance of seedlings can be assessed by growing them on in vitro media with a low $\psi_W$. Since seedlings grown on a low $\psi_W$ stress media has to equilibrate their tissue $\psi_W$ to it, they cannot avoid the stress. It is, therefore, a useful approach in studying drought tolerance mechanisms of plants, since stress conditions are controlled and many replicates can be handled. Various chemicals exist to lower the $\psi_W$ of media. The simplest is the incorporation of NaCl into the medium. Typical concentrations used in the literature range from 100-200 mM NaCl. Due to the toxic effects that the Na$^+$ ions have on plant cells, and the ionic stress that is induced, NaCl is not a good osmotic agent to study plant responses at low $\psi_W$ (Verslues et al., 2006). It is, however, frequently used in the literature for this purpose.

Another solute that is often employed to lower the $\psi_W$ of growth medium is mannitol. Mannitol is a low molecular weight solute that can unfortunately freely penetrate the pores in the plant cell wall and cause plasmolysis. Plasmolysis is the loss of water from the protoplast resulting in a decrease in its volume, while the volume of the cell wall remains unchanged. This leads to a separation of the cell wall and the cell membrane. Cellular damage may occur, thereby eliciting different responses than water loss caused by drying soil. Osmotic stress under plasmolytic conditions should, therefore, be avoided when studying responses to low $\psi_W$. Since mannitol is taken up by plant cells, it has also been shown to have a toxic effect on growth, and can thereby obscure the low-$\psi_W$ response (Verslues et al., 2006).

The high-molecular weight, non-ionic solute polyethylene glycol (PEG), with molecular weight of 6000 g/mol or above, cannot enter the cell and, therefore, causes cytorrhysis rather than plasmolysis. Cytorrhysis is the withdrawal of water from both the cell wall and the cytoplasm.
This is what happens in plant cells when grown in drying soil. High-molecular-weight PEG is, therefore, the best solute to use to impose a low-ψ_W stress that mimics the stress imposed on a plant by drying soil (Verslues et al., 2006).

A method that does not use any chemicals to induce a low ψ_W in A. thaliana seedlings was employed by Umezawa et al. (2004). They applied drought stress to one-week old A. thaliana seedlings by placing them on filter paper for 45 min, whereafter they were transferred to MS medium and grown vertically to assess their root elongation and formation of lateral roots.

One of the aims of this study was to functionally evaluate the role of the VuLEA5 protein in the drought tolerance of plants. In order to reach this aim, the VuLEA5-1 gene was introduced into A. thaliana and tested whether it can improve its tolerance to drought, through controlled drought stress and in vitro osmotic stress experiments.

4.3 Materials and Methods

4.3.1 Construction of the VuLEA5-1 transformation plasmid

The VuLEA5 coding sequence was amplified by RT-PCR from the IT96D-602 line of cowpea and cloned into pGEM-T Easy (Promega) (Chapter 3). A plasmid clone containing the VuLEA5-1 allele (having an A at position 226 of the coding sequence) was chosen for further cloning (Chapter 3, genomic sequence of IT96D-602 represented in Figure 3.2). PCR products were generated from 10 ng of this plasmid using the Expand Long template PCR system (Roche) and the LEA5_start and LEA5_stop primers (Appendix 1). The cycling conditions included denaturation at 94°C for 2 min, cycling 10 times at 94°C for 30 s, 56°C for 30 s and 68°C for 45 s, whereafter another 20 cycles of 94°C for 30 s, 56°C for 30 s and 68°C for 1 min followed. All products were fully extended by a final elongation step of 30 min at 68°C.

The resulting PCR product was inserted into the pCR8/GW/TOPO Gateway entry vector (Invitrogen, Carlsbad, CA, USA) by TOPO-cloning, following the manufacturer’s protocol. The reaction contained 4 µl purified VuLEA5-1 PCR product, 1 µl salt solution and 1 µl TOPO vector. The reaction was mixed gently and incubated at room temperature for 30 min. The total 6 µl TOPO cloning reaction was transformed into competent OneShot TOP10 E. coli cells (Invitrogen) following standard protocols and plated out onto Spectinomycin (100 µg/ml) LB-agar plates. Four colonies were sequenced with the M13-20-F primer (Appendix 1;
Inqaba Biotec). Sequence analyses were performed using BioEdit Sequence alignment editor 7.0.9.0 (Hall, 1999).

The VuLEA5-1 coding sequence was transferred from the correct entry clone to the destination vector pMDC32 (Curtis and Grossniklaus, 2003) using an in vitro LR clonase recombination reaction, according to the manufacturer’s instructions (Invitrogen) (Appendix 4). Entry vector (pCR8-VuLEA5-1 clone 3) and destination vector (pMDC32), 150 ng each, were mixed in a total volume of 8 µl. Two microlitres of the LR Clonase II enzyme mix was added and incubated overnight at 25°C. One microlitre of the Proteinase K solution was added and incubated at 37°C for 10 min. Five microlitres of the LR reaction was transformed into competent E. coli JM109 cells (Promega) following standard protocols and plated out onto kanamycin (100 µg/ml) LB-agar plates. Four colonies were sequenced with the LEA5_start primer; LEA5_R primer and the M13-20-F primer (Appendix 1; Inqaba Biotec).

The correct recombinant plasmid was introduced into Agrobacterium tumefaciens strain GV3101 (pMP90) (Koncz and Schell, 1986) by cold 20 mM CaCl₂ treatment and freeze-thaw and selected on YEP-agar plates (10 g bactopeptone, 10 g yeast extract, 5 g NaCl per liter, pH 7.2; 18% bacto-agar) containing 50 µg/ml each of gentamycin, rifampicin and kanamycin. Plates were incubated inverted at 28°C for 4 days for colonies to appear.

### 4.3.2 Transformation of A. thaliana

A. thaliana plants of the Col-0 ecotype were grown until flowering stage in a growthroom with a light/dark cycle of 16h/ 8h at 25°C/ 18°C, with the relative humidity alternating between 98% and 55% (University of Pretoria, Pretoria, South Africa). It was transformed by Agrobacterium using the floral dip method (Clough and Bent, 1998). A recombinant A. tumefaciens colony transformed with pMDC32-VuLEA5-1 was inoculated into a 10 ml selective overnight LB starter culture. This culture was used to inoculate a 200 ml liquid culture, which was grown at 28°C with shaking for 48 h. A. tumefaciens cells were collected at 5000 rpm for 10 min and resuspended in 5% sucrose to an OD₆₀₀nm of between 0.8 and 1. Silwet was added to 0.05% and inflorescences of A. thaliana Col-0 plants dipped in the solution for 30 s. Excess dip was shaken off and the plants placed overnight on their sides in a covered, moist tray. The plants were placed upright, watered and dipped again after 7 days following the same procedure. Plants were allowed to set seed and seed was collected individually from each plant.

Seeds were surface-sterilised (70% ethanol for 10 min; 10% bleach, 0.1% Triton X-100, 0.1% Sporekill for 15 min; rinse two times with sdH₂O) and resuspended in 1.5 ml sterile
0.2% phytoagar. Seeds were plated onto 90 mm petri dishes containing ½ MS with vitamins (Highveld Biologicals, HP09), 0.8% (w/v) phytoagar and 20 mg/l hygromycin. Seeds were vernalised at 4°C for three days. They were then subjected to 6 h light and 48 h dark in a growthroom at 21°C (Harrison et al., 2006). Scoring of the number of transgenic seedlings was performed after two to five days of growing in the light.

T1 seedlings surviving on 20 mg/ml hygromycin selective medium were transplanted to Jiffies (Jiffy Product International AS, Norway) in a growth room at 22°C. Genomic DNA was isolated from leaf samples using the DNeasy Plant mini kit (Qiagen). PCR screening for the presence of the VuLEA5-1 gene was performed on 2 µl of undiluted genomic DNA using the LEA5_start and LEA5_stop and LEA5_start and M13-26-R primer combinations (Appendix 1) and GoTaq DNA polymerase (Promega). Amplification products were analysed on 1.2% agarose/ 0.5x TAE gels.

T2 seeds were collected and the screening repeated until the T4 generation. Lines giving 100% survival were homozygous and T4 seeds were used for further experiments.

4.3.3 Analysis of transgenic plants
4.3.3.1 PCR
The presence of the VuLEA5-1 transgene was analysed by PCR from leaf tissue samples. Genomic DNA was isolated from one gram of leaf tissue from nine pooled 4-week old T4 A. thaliana plants using the DNeasy Plant Maxi kit (Qiagen). PCR was carried out on 30 ng of genomic DNA using 0.2 µM each of the LEA5_start and LEA5_stop and LEA5_start and M13-26-R primer combinations (Appendix 1) and the 1x EmeraldAmp GT PCR master mix (Takara Bio Inc., Otsu, Shiga, Japan). For amplification of the Cap binding protein-20, the CBP_F and CBP_R primers (Appendix 1) were used. The PCR reactions were run for 30 cycles with a primer annealing temperature of 56°C. PCR products were separated and visualised by agarose gel electrophoresis.

4.3.3.2 Southern blot
VuLEA5-1 transgene integration and copy number was analysed by Southern blot. Seven micrograms of genomic DNA, extracted from the transgenic lines and Col-0 and digested with Xhol or HindIII, were separated on an agarose gel and blotted to a nylon membrane (Osmonics Magnaprobe nylon transfer membrane, Amersham-Pharmacia Biotec, Little Chalfont, UK) using standard protocols (Southern, 1975). The pMDC32-VuLEA5-1 plasmid was digested with Xhol to generate a 1850 bp fragment containing VuLEA5-1 (Appendix 4).
This fragment was spiked into untransformed *A. thaliana* Col-O genomic DNA for determination of the transgene copy number in the transgenic lines during Southern blot.

Molecular weight marker IV (1 µg; Roche) was end-labelled with alkali-stable DIG-dUTP (1 mM; Roche) using Klenow enzyme (3 U; Roche). Unlabelled dATP, dCTP and dGTP (100 mM each) and 1x Klenow buffer were included in the reaction and incubated at 37ºC for 3 h, whereafter the enzyme was heat-inactivated at 65ºC for 15 min. Sixty nanogram labelled marker was loaded in one lane of the agarose gel before electrophoresis and blotting.

The membrane was probed with a DIG-labelled *VuLEA5-1* probe at a hybridisation temperature of 42ºC. The DNA probe was generated using the PCR DIG Probe synthesis kit (Roche Diagnostics GmbH, Germany), and the LEA5_start and LEA5_stop primers *(Appendix 1)*. The denatured probe was added to DIG Easy Hyb solution containing 250 µg/ml denatured salmon testes DNA (Sigma Aldrich). After overnight hybridisation the membrane was washed with two buffers differing in stringency: 2x SSC, 0.1% SDS at room temperature and 0.5x SSC, 0.1% SDS at 65ºC. The membrane was blocked and DIG detected using the DIG Wash and Block Buffer set and the DIG Luminescent detection kit for nucleic acids (Roche Diagnostics GmbH). The membrane was exposed to X-ray film (Hyperfilm ECL High performance chemiluminescence film, Amersham-Pharmacia Biotec) for 40 min and the film developed.

**4.3.4 Analysis of transgene expression**

**4.3.4.1 RT-PCR**

The detection of transgene expression was carried out by gene-specific RT-PCR. Young leaves were collected from 4-week old transgenic *A. thaliana* lines from three pools of three plants each per line. RNA was isolated from 100 mg leaf tissue using the RNeasy Plant kit (Qiagen) with on-column DNase treatment (RNase-free DNase kit, Qiagen). RT-PCR was performed on 50 ng *A. thaliana* RNA using the C-therm Polymerase One-Step RT-PCR system (Roche). Primer pairs amplifying the *VuLEA5-1* gene (LEA5_start and _stop) and Cap binding protein-20 gene (CBP; AT5G44200, GenBank: NM_123787.3) were used *(Appendix 1)*. The RT-PCR reactions were incubated at 60ºC for 30 min before cycling for 25 cycles with a primer annealing temperature of 56ºC. Amplification products were analysed on a 2% agarose gel.

**4.3.4.2 Protein isolation and Western blot**

Protein accumulation was monitored by immunoblotting protein extracted from leaves of 4-week-old T4 *A. thaliana* transgenic and Col-0 plants. Approximately 250 mg leaf material
collected from nine pooled plants were homogenised in 1 ml extraction buffer (20 mM Tris-HCl (pH 6.8), 10 mM EDTA, 1 mM PMSF and 10 µM Leupeptin). After centrifugation at 15000 g for 15 min at 4°C the resultant supernatant was heated for 10 min at 95°C. After incubating on ice for 15 min, the heat-soluble protein fraction was collected as the supernatant after centrifugation at 15000 g for 15 min at 4°C. Proteins were quantified according to Bradford (1976) with the BioRad protein assay kit (Hercules, CA, USA) using bovine serum albumin (BSA) as protein standard.

Ten micrograms of protein per sample were precipitated with 9 volumes ethanol, incubated at -20°C overnight and the pellet collected by centrifugation at 15000 g for 15 min at 4°C. The pellet was washed with cold 90% ethanol, vacuum-dried and resuspended in 25 µl 1x SDS-PAGE sample loading buffer. After heating to 95°C for 10 min, the samples were loaded onto a 4% (w/v) stacking/ 12% (w/v) separating discontinuous SDS-PAGE gel and the proteins separated for 50 min at 150V on a BioRad Mini Protean 3 cell (Laemmli, 1970). The High range Rainbow molecular weight marker RPN756 (1.5 µl, Amersham Biosciences UK Limited, Little Chalfont, England) was loaded as protein ladder. Silver staining of the gel after electrophoresis with silver nitrate was performed according to standard protocols (Blum et al., 1987).

Another gel containing identical samples, but substituting the Rainbow marker with a biotinylated marker (ECL, RPN2107, Amersham Pharmacia Biotec UK Limited, Little Chalfont, England), was run simultaneously. This gel was electroblotted to PVDF-Plus transfer membrane (Osmonics, MSI, Westborough, Ma, USA) at 100V for 1.5 h in 25 mM Tris (pH 8.4), 192 mM glycine and 20% (v/v) methanol using a mini-transblot system (BioRad). The membrane was blocked overnight in Tris-buffered saline (TBS, 10 mM Tris-HCl pH 7.5 and NaCl 150 mM) containing 1.5% Tween-20 under constant agitation on a platform rocker (Bibby Sterilin Limited, Staffordshire, UK). After several washes in TBS-T (TBS containing 0.05% Tween-20) the membrane was incubated for 2 h at room temperature with a rabbit polyclonal antibody against MtPM25 (dilution 1:10000 in TBS-T). This antibody was a gift from Olivier LePrince (UMR Physiologie Moléculaire des Semences, University of Angers, France). Lyophilised serum was resuspended in TBS buffer and dissolved very gently. After washing in TBS-T, the membrane was incubated for 2 h at room temperature with the secondary anti-rabbit IgG-horseradish peroxidase (HRP) conjugate (ECL, diluted 1:5000) and streptavidin-HRP conjugate (ECL, RPN1231, diluted 1:1500). After washing in TBS-T, immunodetection was performed by chemiluminescence using the ECL Western blotting analysis system. The membrane with substrate was exposed to autoradiography film.
(Hyperfilm ECL High performance chemiluminescence film, Amersham-Pharmacia Biotec) for 1, 3, 10 and 30 min and the film developed.

4.3.4.3 RT-qPCR
The quantification of transgene expression was carried out by reverse transcriptase quantitative PCR (RT-qPCR). Young leaves were collected from 4-week old transgenic *A. thaliana* lines from three pools of three plants each per line. RNA was isolated from 100 mg leaf tissue using the RNeasy Plant kit (Qiagen) with on-column DNase treatment (RNase-free DNase kit, Qiagen). Dilutions of 10 ng/µl were prepared for RT-qPCR. RT-qPCR was performed on 50 ng RNA from Col-0 and eight transgenic lines (three biological replicates each) using the primers for *VuLEA5* and several reference genes (Elongation factor 1-alpha (EF1A; AT5G60390), Glyceraldehyde-3-Phosphate dehydrogenase (GAPDH; AT1G13440), Tubulin 2 (TUB2; AT5G62690) and Actin 2 (ACT2; AT3G18780); Appendix 1). Primers were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Triplicate standard curves were constructed for all primer pairs from a serial dilution of pooled *A. thaliana* RNA. A master mix of 1x Sensimix SYBR One Step (Quantace, Celtic Molecular Diagnostics), SYBR Green, RNase inhibitor (0.2 U/µl) and the appropriate primer pair (200 nM each, Appendix 1) in a total volume of 20 µl was prepared following the manufacturer’s protocol. Five microlitres of RNA dilution were added to the reaction mastermix in 0.1 ml strip tubes and run on a Rotor-Gene 3000A (Corbett Research). The cycling profile included a hold at 42°C for 30 min for reverse transcription, 95°C for 10 min for enzyme activation and 45 cycles of 95°C for 15 s, primer annealing at 56°C for 15 s and extension at 72°C for 15 s. The SYBR green measurement was taken every cycle after the extension step. Melt curves were produced on the Rotor-Gene by increasing the temperature from 72 to 95°C with 1°C increments. Cq values were obtained at a threshold of 0.01 and a cut-off of 10 cycles. E-values were calculated from the slopes of the standard curves. Expression data was analysed with qBase to determine the highest relative expressing line of *VuLEA5*-1.

4.3.5 Drought stress studies
4.3.5.1 Drought stress treatments of transgenic *A. thaliana* plants in soil
Drought stress was induced in soil-grown *A. thaliana* plants (Brini *et al*., 2007). Homozygous transgenic *A. thaliana* seeds were surface-sterilised and germinated on ½ x MS + vitamins (Highveld Biologicals HP09) media containing 0.8% plant agar plates, and the 7 d old seedlings transferred to soil media (Jiffy) and grown in a growthroom with a light/dark cycle of 16h/8h at 25°C/18°C, with the relative humidity alternating between 98 and 55%.
Chapter 4

After 3 weeks of growth with normal water supply, drought stress was imposed by withholding water for 10 days. Control plants were watered as normal. Drought-stressed plants were rewatered after 10 days and the survival rate was calculated 3 days after rewatering. Each transgenic line was represented by 24 replicate plants.

4.3.5.2 Seed germination *in vitro* of transgenic *A. thaliana* in the presence of salt

Homozygous T4 seed of *VuLEA5-1* transgenic and wild-type Col-0 *A. thaliana* were surface-sterilised and plated on ½ x MS + vitamins (Highveld Biologicals HP09) media containing 0.8% plant agar and 50, 100 and 200 mM NaCl at 30 seeds per plate, and the number of germinated seedlings recorded after 7 and 15 days (Brini *et al.*, 2007; Dalal *et al.*, 2009, Dai *et al.*, 2011).

4.3.5.3 Osmotic stress treatment of *A. thaliana* seedlings *in vitro*

Homozygous T4 seed of *VuLEA5-1* transgenic and wild-type Col-0 *A. thaliana* were surface-sterilised and plated on ½ MS + vitamins (Highveld Biologicals HP09) media containing 0.8% plant agar. Seeds were vernalised for 48 h at 4°C and incubated in a growthroom at 22°C under continuous lighting. Seedlings (5 – 6 d old) were transferred to control media or media containing osmotic stress agents: NaCl (100 mM and 200 mM), mannitol (200 mM and 350 mM), or plates equilibrated overnight with a solution of PEG-8000 (Sigma Aldrich) to obtain a final water potential ($\psi_W$) of -0.5 MPa, -0.7 MPa or -1.2 MPa (van der Weele *et al.*, 2000; Verslues *et al.*, 2006; Brini *et al.*, 2007; Dai *et al.*, 2011). Square petri dishes (10 cm) were used and incubated vertically. The elongation of the primary root was determined by scanning the plates at different time-points and measuring the roots with ImageJ 1.44p software (National Institutes of Health, USA; [http://imagej.nih.gov/ij](http://imagej.nih.gov/ij)). For recovery experiments, stressed seedlings were transferred onto control MS media. The fresh weights of seedlings (in pools of six) were determined after the recovery period. For measuring hypocotyl length, seeds were germinated and grown in the dark, covered in foil, on media containing 0.5% sucrose.

4.4 Results

4.4.1 Construction of the *VuLEA5-1* transformation plasmid

The *VuLEA5-1* coding sequence was cloned from the IT96D-602 cowpea line into a Gateway transformation plasmid for *A. thaliana* transformation. The Gateway vector system is a plant transformation vector system that makes use of recombination-based transfer of a gene of interest between an entry and destination vectors. The *VuLEA5-1* allele, containing an A at position 226 and coding for an isoleucine amino acid at position 76, was chosen for *A.
 transformations, since it corresponded to the relevant SSH library clone (07F09-F) and it was unique to the drought tolerant cowpea line IT96D-602. The sequence of the *VuLEA5-1* gene and the amino acid sequence it encodes are indicated in Appendix 3.

The cloning strategy of inserting *VuLEA5-1* into the plant transformation plasmid is indicated in Appendix 4. The *VuLEA5-1* coding sequence was amplified with PCR using a high fidelity DNA polymerase, Expand Long template (Roche). It contains a mixture of Taq DNA polymerase, which is thermostable and adds 3’ A overhangs, and Tgo DNA polymerase, which has 3’-5’ exonuclease proofreading activity. The combination of polymerases still result in PCR products adequately A-tagged for T-A cloning. The PCR product was column-purified and TOPO-cloned into the pCR8/GW/TOPO Gateway entry vector. This vector contains the TOPO cloning site for rapid and efficient cloning of Taq-amplified PCR products. The vector is linearised with single 3’-T overhangs and Topoisomerase I covalently bound to it, activating the vector for cloning a PCR product while releasing the topoisomerase. It also contains the **att**L1 and **att**L2 sites for recombination-based transfer of the gene of interest into any Gateway destination vector. The entry clone containing the insert in the correct orientation relative to the **att**L1 and **att**L2 recombination sequences were selected from the sequencing results (results not shown). The *VuLEA5-1* gene was transferred into the Gateway destination vector pMDC32 via a LR Clonase reaction. The Gateway LR Clonase II enzyme mix contains the bacteriophage lambda recombination proteins Integrase (Int) and Excisionase (Xis), the *E. coli*-encoded protein Integration Host Factor (IHF) and reaction buffer. It catalyses in vitro recombination between an **att**L-flanked gene in an entry clone and an **att**R-containing destination vector to generate an **att**B-containing expression clone.

\[
\text{**att**L1-gene-**att**L2} \times \text{**att**R1-ccdB-**att**R2} \leftrightarrow \text{**att**B1-gene-**att**B2} \times \text{**att**P1-ccdB-**att**P2}
\]

entry clone \hspace{1cm} destination vector \hspace{1cm} expression clone \hspace{1cm} pDONR

Within the T-DNA borders, pMDC32 contains the *ccdB* gene flanked by **att**R sites, and it provides the 2x35S promoter for constitutive expression of the inserted gene, a nos terminator, and the hygromycin resistance gene as selectable marker of transformed plants. Outside the T-DNA borders are an *E. coli* origin of replication and a kanamycin resistance gene, that enables replication and selection in *E. coli*. Selection for recombinant clones occurred by transformation into a *recA, endA* *E. coli* strain, JM109, which does not contain the *ccdA* gene, to facilitate negative selection with the *ccdB* gene. Cells containing the recombinant vector were also selected by plating on medium containing kanamycin.

© University of Pretoria
Positive colonies were identified by colony PCR with the LEA5_start, LEA5_stop and M13-26-R primers (Appendix 1). The recombinant pMDC32-VuLEA5-1 plasmid was detected in all four screened colonies (Figure 4.1, lanes 1-4 and 7-10). A 327 bp product was expected from a positively transformed clone with the LEA5_start and LEA5_stop primers (Figure 4.1, lanes 1-4), and 784 bp from the LEA5_start and M13-26-R primer combination (Figure 4.1, lanes 7-10). The M13-26-R primer binding site originates from the pMDC32 plant transformation vector, therefore no product was expected with the LEA5_start and M13-26-R primer combination with the pGEM-VuLEA5-1 plasmid (Figure 4.1, lane 11).

Four clones were sequenced (results not shown), and plasmid was isolated from a correct pMDC32-VuLEA5-1 clone and transformed into the A. tumefaciens strain GV3101 (pMP90) for A. thaliana transformation.

![Figure 4.1](image-url) **Figure 4.1** PCR colony screen of *E. coli* JM109 transformants of pMDC32-VuLEA5-1. M: MWM VI (Roche); Lanes 1-6: Primers LEA5_start and LEA5_stop; Lanes 7-12: Primers LEA5_start and M13-26-R; Lanes 1-4 and 7-10: colony supernatant from pMDC32-VuLEA5-1 colonies 1 to 4, respectively; Lanes 5 and 11: Positive control containing 10 ng pGEM-VuLEA5-1 plasmid; Lanes 6 and 12: negative water control.

### 4.4.2 Transformation of *A. thaliana*

Flowering *A. thaliana* Col-0 plants were dipped into a culture of the *A. tumefaciens* containing the pMDC32-VuLEA5-1 plant transformation plasmid. Seed (T1) was collected from transformed plants and hygromycin selection applied *in vitro*. Transgenic seedlings were transplanted to Jiffies in a growthroom and grown to maturity.

PCR screening of gDNA isolated from T1 leaf samples identified the transgene in all but one transgenic line. Line 20 did not produce the expected amplification products with both primer pairs tested (Figure 4.2 A and B; lane 10). The same product sizes were expected as in Figure 4.1. The LEA5_start and M13-26-R primer combination (Figure 4.2 B) was included...
to serve as a control for the transgenic event, in case the LEA5 primer pair hybridised to and amplified an endogenous *A. thaliana* LEA gene.

Figure 4.2 PCR screening of T1 *A. thaliana* lines transformed with the *VuLEA5-1* gene. Genomic DNA was isolated from leaf samples and screened with PCR using the following primer pairs: **A**: LEA5_start and LEA5_stop primers; **B**: LEA5_start and M13-26-R primers. M: 50 bp DNA ladder; lanes 1 to 13: templates are *A. thaliana* genomic DNA of the T1 lines 1, 5, 6, 8, 12, 14, 15, 17, 18, 20, 21, 22 and 23; lanes 14 and 15: untransformed *A. thaliana* 1 and 2; lane 16: 10 ng pMDC32-*VuLEA5-1*-plasmid template as positive control; lane 17: negative water control.

Hygromycin selection was applied on successive generations until T4 plants homozygous for the *VuLEA5-1* transgene were obtained. It was difficult to select for hygromycin-resistant *A. thaliana* seedlings, since both resistant and non-resistant seedlings are green after selection on medium containing hygromycin. However, by including a 2 d dark growth period in the selection protocol (Harrison *et al.*, 2006), the resistant seedlings’ hypocotyls (~10 mm; **Figure 4.3 A**) elongate much more than the non-resistant hypocotyls (~3 mm; **Figure 4.3 B**). They can then easily be identified and scored after only a few days of growth.
Figure 4.3 Difference in shoot lengths of *A. thaliana* seedlings selected for hygromycin resistance. Transgenic *A. thaliana* seed containing the hygromycin resistance gene (A), and non-transgenic Col-0 seed (B), were plated on ½ MS medium + vitamins containing 20 mg/l hygromycin. Plated seeds were vernalised at 4°C for 3 d, subjected to 6 h light in a growthroom, kept in the dark for 2 d, grown in the light for 2 d whereafter 10 seedlings each were photographed.

Homozygous T4 progeny were identified for eight independent transgenic lines. They were lines 1-5.9, 5-6.5, 6-2.7, 9-7.2, 12-3.8, 20-2.4, 21-2.9 and 23-7.7. These lines were molecularly characterised and screened for improved drought tolerance. *A. thaliana* Col-0 (T4 generation) were included as control in all cases. Line names were abbreviated to the first number, namely 1, 5, 6, 9, 12, 20, 21 and 23.

4.4.3 Analysis of transgenic plants

4.4.3.1 PCR analysis of *VuLEA5-1* transgenic *A. thaliana*

The *VuLEA5-1* transgene was detected with PCR in 7 of the 8 transgenic T4 *A. thaliana* lines (Figure 4.4 A). Line 20 does not contain the *VuLEA5-1* gene, but retained hygromycin resistance due to segregation from the antibiotic selectable marker. Line 20 in the T1 generation also did not show PCR amplification with the *VuLEA5-1* gene primers. The transgene must have segregated from the hygromycin antibiotic selective marker in this line. The primer combination amplifying the *VuLEA5-1* gene from its start to stop codon produced non-specific amplification products from Col-0 and line 20 (Figure 4.4 A, lanes 6, 9 and 10). The *VuLEA5-1* gene primers are probably binding to the endogenous *A. thaliana* LEA gene sequences. PCR reactions with a combination of the LEA5_start primer and a primer binding to a region outside of the gene in the T-DNA, namely M13-26-R, was included to increase the specificity of amplifying the transgene (Figure 4.4 B). No amplification products were
obtained from Col-0 and line 20 (Figure 4.4 B, lanes 6, 9 and 10). To prove that equal quantities of gDNA were used from each line, the *A. thaliana* Cap-binding protein 20 gene (AtCBP20; AT5G44200; GenBank: NM_123787.3) was amplified using the CBP primers (Figure 4.4 C). Amplification products of 660 bp with similar intensities were obtained from all lines tested, including the untransformed Col-0 and line 20.

**Figure 4.4 PCR amplification of the VuLEA5-1 transgene and Cap-binding protein 20 gene in T4 generation transgenic and untransformed *A. thaliana* lines.** Genomic DNA (30 ng) isolated from transgenic *A. thaliana* lines 1, 5, 6, 9, 12, 20, 21 and 23 (lanes 1-8) and untransformed Col-0 (lanes 9 and 10) were amplified by the primer combinations indicated: **A:** LEA5_start and LEA5_stop primers; **B:** LEA5_start and M13-26-R; **C:** CBP_F and CBP_R. Lane 11: positive plasmid control; lane 12: negative water control. M: 50 bp DNA ladder.
4.4.3.2 Southern blot of *VuLEA5-1* transgenic *A. thaliana*

Restriction digested genomic DNA was separated by overnight agarose electrophoresis. The DNA ran evenly in straight lanes (Figure 4.5). Digestion of 7 micrograms of DNA with *Hind*III was more complete than *Xho*I. Higher molecular weight smears are visible in all the lanes containing *Xho*I-digested gDNA.

![Image](image-url)

**Figure 4.5** Restriction enzyme digested genomic DNA of *VuLEA5-1* transgenic *A. thaliana* lines after overnight electrophoresis on an agarose gel. Lane 1: Molecular weight marker IV (Roche, DIG-labelled); lanes 2-3: 7 µg untransformed Col-0 *A. thaliana* digested with *Xho*I and *Hind*III, respectively; lanes 4-19: 7 µg of the following transgenic *A. thaliana* lines gDNA digested with *Xho*I and *Hind*III, respectively: lanes 4-5: line 1; lanes 6-7: line 5; lanes 8-9: line 6; lanes 10-11: line 9; lanes 12-13: line 12; lanes 14-15: line 20; lanes 16-17: line 21; lanes 18-19: line 23; lane 20: one copy of the *VuLEA5-1* fragment spiked into 7 µg *Xho*I digested Col-0 gDNA.
A Southern blot was successfully performed to detect the *VuLEA5-1* transgene in the genome of transgenic *A. thaliana* lines (Figure 4.6). DIG-labelled *VuLEA5-1* was used to probe genomic DNA that have been digested with two different restriction enzymes and blotted to a membrane. *XhoI* excises a 1850 bp fragment from the T-DNA containing the transgene, and *HindIII* cuts only once within the T-DNA borders of the *VuLEA5-1* transgenic plasmid (see Appendix 4 for plasmid map). This enzyme was, therefore, chosen to determine the copy number of the transgenic event. One to three insertions of the transgene were detected in the T4 generation homozygous transgenic lines. Line 20 did not contain any *VuLEA5-1* insert (Figure 4.6, lanes 14 and 15), which was as expected, since it also did not show PCR amplification with *VuLEA5-1* primers since the T1 generation. The transgene must have segregated from the hygromycin antibiotic selective marker in this line. Line 20 was, therefore, used as a negative control during drought stress studies. Lines 9 and 12 contained single copy insertions (Figure 4.6, lanes 11 and 13), with lines 6, 21 and 23 having two copies (Figure 4.6, lanes 9, 17 and 19). Lines 1 and 5 seemed to have triple copy insertions (Figure 4.6, lanes 5 and 7).

![Figure 4.6 Genomic Southern blot analysis of the *VuLEA5-1* gene in transgenic *A. thaliana* lines.](image)

Genomic DNA from Col-0 and lines 1, 5, 6, 9, 12, 20, 21 and 23 were digested with restriction enzymes, separated by overnight agarose electrophoresis, transferred to a membrane and probed with a DIG-labelled *VuLEA5-1* probe. Lane 1: Molecular weight marker IV (Roche, DIG-labelled); lanes 2-3: 7 µg untransformed Col-0 A.
transgenic *A. thaliana* lines gDNA digested with *Xho*I and *Hind*III, respectively: lanes 4-5: line 1; lanes 6-7: line 5; lanes 8-9: line 6; lanes 10-11: line 9; lanes 12-13: line 12; lanes 14-15: line 20; lanes 16-17: line 21; lanes 18-19: line 23; lane 20: one copy of the *VuLEA5-1* fragment spiked into 7 µg *Xho*I digested *Col-0* gDNA.

### 4.4.4 Analysis of transgene expression

#### 4.4.4.1 RT-PCR of *VuLEA5-1* transgenic *A. thaliana*

The *VuLEA5-1* transcript was shown to accumulate in seven transgenic *A. thaliana* lines using semi-quantitative RT-PCR (Figure 4.7). Equal loading of RNA was verified by the amplification of the Cap-binding protein 20 (CBP) gene. A product of 237 bp was expected from cDNA, since an intron was present in the gene. Transgenic line number 20 was hygromycin resistant, but did not contain the *VuLEA5-1* gene when the T1 generation was screened with PCR (Figure 4.2, lane 10). This correlates with the lack of the *VuLEA5-1* transcript, as identified with RT-PCR on RNA isolated from the T4 generation (Figure 4.7, lane 8).

![Figure 4.7 Transcript accumulation patterns of the *VuLEA5-1* transgene in transgenic *A. thaliana* lines.](image)

The Cap binding protein-20 (CBP) transcript was used as a loading control. The arrows indicate the expected transcript sizes of 327 bp and 237 bp, for *VuLEA5-1* and *CBP*, respectively.

#### 4.4.4.2 Western blot analysis of transgenic *A. thaliana*

A Western blot was performed on the *A. thaliana* lines to detect the expression of the *VuLEA5-1* protein. Heat-stable protein fractions were prepared from all the lines, and the protein concentrations ranged between 46 and 73 µg/ml. To load 10 µg protein for Western blotting, the proteins were concentrated by precipitation. Different protein precipitation methods have different yields for different sized proteins (results not shown). Acetone and ethanol loses high molecular weight proteins, but have better resolution on a SDS-PAGE gel.
at the smallest molecular weight range (results not shown). Since the expected molecular weight of the 107 amino acid recombinant VuLEA5-1 protein was 11.7 kDa, ethanol precipitation was chosen to precipitate the heat-soluble protein extracts of the transgenic lines for Western blotting. Equal loading of 10 µg protein from each line was verified by a SDS-PAGE gel, followed by silver staining (Figure 4.8). No differences were observed in the protein banding patterns of the untransformed Col-0 and transgenic *A. thaliana* lines (compare Figure 4.8 lane 2 with lanes 3 to 10).

**Figure 4.8 Silver stained gel of heat-soluble proteins prepared from untransformed *A. thaliana* and *VuLEA5-1* transgenic lines.** Lane 1: High range Rainbow molecular weight marker. Ten microgram heat-soluble proteins of each line were loaded in the lanes as indicated: lane 2: Col-0; lanes 3 to 10: transgenic line 1, 5, 6, 9, 12, 20, 21 and 23, respectively. Molecular marker mass is indicated in kilodaltons.

A rabbit polyclonal antibody against MtPM25 (α-MtPM25) was obtained from Olivier LePrince (UMR Physiologie Moléculaire des Semences, University of Angers, France). Western blotting with the α-MtPM25 antibody resulted in a hybridising signal in all lanes at 14.4 kDa, with two faint signals in all lanes at 97 kDa and between 14.4 and 20.1 kDa (Figure 4.9). No differences were detected in the patterns of hybridising proteins between the untransformed Col-0 (Figure 4.9, lane 2) and the *VuLEA5-1* transgenic lines (Figure 4.9, lanes 3 to 10).
Figure 4.9 Western-blot analysis of VuLEA5-1 in untransformed *A. thaliana* and *VuLEA5-1* transgenic lines. Ten microgram heat-soluble proteins of each line were hybridised with rabbit antibodies against recombinant MtPM25. Samples were loaded in the lanes as indicated: lane 1: High range Rainbow molecular weight marker; lane 2: Col-0; lanes 3 to 10: transgenic line 1, 5, 6, 9, 12, 20, 21 and 23, respectively. Molecular marker mass is indicated in kilodaltons.

4.4.4.3 RT-qPCR quantification of *VuLEA5-1* expression in transgenic *A. thaliana* lines

Reference gene stability of four reference genes (EF1A, GAPDH, TUB2 and ACT2; Appendix 1) were analysed with qBase from a RT-qPCR experiment using RNA pooled from *A. thaliana* Col-0 and a transgenic line (results not shown). Satisfactory M and CV values (smaller than 0.5 and 0.2, respectively) were obtained when EF1A was paired with either TUB2 or ACT2. The EF1A and TUB2 genes were, therefore, chosen as reference genes for quantification of *VuLEA5-1* expression levels in transgenic *A. thaliana* lines.

*VuLEA5-1* gene expression levels of the T4 homozygous transgenic *A. thaliana* lines were assayed by RT-qPCR and normalised with the EF1A and TUB2 reference genes. The data from three biological replicates were analysed separately with qBase, whereafter the expression levels in each line were averaged. These results are presented in Figure 4.10.
The highest relative expressing line of the \textit{VuLEA5-1} mRNA was line 9, followed by line 12. Line 9 had a 43-fold higher expression level than the lowest expression line, line 6, followed by line 12 with 25-fold higher expression. The next highest expressing line was line 5, with 9-fold higher expression. Line 21 and line 1 had similar expression levels, three times higher than line 6, with line 23 only 1.2 times the expression levels of line 6.

4.4.5 Drought stress studies

4.4.5.1 Drought stress treatments of transgenic \textit{A. thaliana} plants in soil

Controlled drought stress experiments were performed to study the drought tolerance of transformed \textit{A. thaliana} plants compared to control plants. Twenty-four replicate plants of each selected transgenic line were grown for 3 weeks and subjected to a drought stress. Withholding water for 10 days imposed a severe drought stress on \textit{A. thaliana} plants (Figure 4.11 B). Plants that became severely wilted couldn’t recover upon rewatering and died (Figure 4.11 C). Table 4.1 contains the number of surviving plants of each line, 3 days after rewatering. Three of the transgenic lines (lines 9, 12 and 21) showed higher survival rates than the negative control (line 20) and wild-type (Col-0).
Figure 4.11 Drought stress and recovery of *A. thaliana* plants transformed with the *VuLEA5-1* gene (line 12). Three-week old transgenic plants of line 12 were subjected to drought stress for 10 days (B), whereafter they were rewatered (C). Transgenic plants of the same age, that were not subjected to the drought stress, are indicated at (A).

Table 4.1 Number of surviving *A. thaliana* plants after rewatering 10-day drought stressed plants.

<table>
<thead>
<tr>
<th><em>A. thaliana</em> line</th>
<th>Number of survivors (of 24 replicates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col-0 (wild type)</td>
<td>4</td>
</tr>
<tr>
<td>Line 5</td>
<td>0</td>
</tr>
<tr>
<td>Line 9</td>
<td>7</td>
</tr>
<tr>
<td>Line 12</td>
<td>7</td>
</tr>
<tr>
<td>Line 20 (negative control)</td>
<td>0</td>
</tr>
<tr>
<td>Line 21</td>
<td>6</td>
</tr>
</tbody>
</table>

4.4.5.2 Seed germination *in vitro* of transgenic *A. thaliana* in the presence of salt

Seed germination of all lines was completely inhibited by NaCl (200 mM) incorporated into the medium (results not shown). Mannitol also caused highly variable seed germination rates (results not shown). Seedlings were, therefore, germinated on control medium before being transferred to osmotic agent-containing medium in subsequent experiments.

4.4.5.3 Osmotic stress treatment of *A. thaliana* seedlings *in vitro*

4.4.5.3.1 Root elongation

When osmotic stress was applied to *A. thaliana* seedlings *in vitro* during this study, a reduction in primary root elongation was observed. The severity of the stress influenced the extent of the reduction. Figure 4.12 illustrates the typical root growth when 5-day old seedlings were transferred to MS medium containing various osmotic stress agents and...
grown vertically for 1 week. Low $\psi_W$ stress imposed by infusing MS plates with a solution of PEG-8000 decreases root elongation less at -0.5 MPa than at -1.2 MPa, which halts root elongation almost completely. NaCl stress caused the death of seedlings, probably due to the toxic ionic stress. A high concentration of mannitol (350 mM) induces the formation of secondary roots.

![Figure 4.12 Root elongation of A. thaliana seedlings after one week on control (MS) and different osmotic stress in vitro mediums.](image)

The vegetative stress response of seedlings to osmotic stress was determined by transferring them to in vitro media infused with PEG-8000 to reduce the $\psi_W$ (van der Weele *et al.*, 2000; Verslues *et al.*, 2006). This causes an osmotic stress that is comparable to plants growing in drying soil. Primary root elongation was quantified as a measure of growth, and compared between the *VuLEA5-1* transgenic seedlings and wild-type Col-0. In all cases,
statistical analysis of root elongation percentage differences between Col-0 and the transgenic lines were performed using pair-wise comparisons of the means using Student’s T-test at the 95% confidence level.

Root elongation of each line on stress medium was normalised to the elongation of the same line on control MS medium, with the elongation on control medium set to 100%. On osmotic stress media with a $\psi_W$ of -0.5 MPa, root elongation was reduced to an average of 64% for all the lines. Lines 9 and 21 showed a 10% higher root elongation percentage compared to Col-0. This difference was statistically significant at the 95% confidence level using pair-wise comparisons of the means using Student’s T-test (Figure 4.13). On media with a $\psi_W$ of -1.2 MPa, root elongation was reduced severely to an average of only 4% of the growth on control MS medium. Line 21 again showed a statistical significant higher root elongation percentage compared to Col-0, but this difference was only 1%.

Figure 4.13 Comparison of osmotic stress response of VuLEA5-1 transgenic A. thaliana lines and wild type Col-0, as induced by PEG-8000 infused plates with a $\psi_W$ of -0.5 MPa or -1.2 MPa. Five-day-old seedlings were transferred from normal MS medium to MS medium or media infused with a PEG-8000 solution with a $\psi_W$ of -0.5 MPa (PEG -0.5, hatched blue) or -1.2 MPa (PEG -1.2, red). Root elongation after seven days incubation in the light is presented as the percentage relative to elongation on MS medium. Error bars represent the standard deviation ($n = 25$ to 30).
When the assay was repeated with an intermediate osmotic stress of -0.7 MPa, root elongation reduced to 15% or 11% of control values at different measuring times (4 or 8 days of stress; Figure 4.14). None of the transgenic lines showed statistically significant differences in root elongation at the 95% confidence level, when the means were compared pair-wise to Col-0 using Student’s T-test.

![Figure 4.14 Comparison of osmotic stress response of VuLEA5-1 transgenic A. thaliana lines and wild type Col-0, as induced by PEG-8000 infused plates with a $\psi_W$ of -0.7 MPa. Six-day-old seedlings were transferred from normal MS medium to MS medium or media infused with a PEG-8000 solution with a $\psi_W$ of -0.7 MPa. Root elongation after four days (hatched green) and eight days (solid green) incubation on stress media is presented as the percentage relative to elongation on MS medium. Error bars represent the standard deviation ($n = 28$ to 30).](image)

The effect of the osmotic stress agent mannitol on root elongation was evaluated and compared to PEG-8000. Root elongation of seedlings were reduced to an average of 59% and 74% of the growth on control MS medium by 200 mM mannitol and -0.5 MPa PEG-8000 treatments, respectively (Figure 4.15). Line 12 showed higher root elongation than Col-0 at both treatments, but the result was not statistically significant. None of the transgenic lines resulted in statistically significant changes in root elongation at the 95% confidence level, except line 5, that was significantly reduced compared to Col-0 when grown on PEG-8000 - 0.5 MPa stress medium (Figure 4.15). Osmotic stress by -0.5 MPa PEG-8000 was milder.
than during a previous experiment, which showed a statistically significant 10% increase in root elongation by lines 9 and 21 (Figure 4.13). Different assay conditions and media preparations from one experiment to the next, combined with the variable responses of different lines, make it clear that there are no specific lines that continuously perform better than the wild-type Col-0 during osmotic stress assays.

Figure 4.15 Comparison of osmotic stress response of VuLEA5-1 transgenic A. thaliana lines and wild type Col-0, as induced by 200 mM mannitol or PEG-8000 infused plates with a $\psi_W$ of -0.5 MPa. Six-day-old light-germinated seedlings were transferred from normal MS medium to MS medium, media supplemented with 200 mM mannitol (Man 200, solid orange) or media infused with a PEG-8000 solution to a $\psi_W$ of -0.5 MPa (PEG -0.5, hatched blue). Root elongation after six days incubation in the light is presented as the percentage relative to elongation on MS medium. Error bars represent the standard deviation ($n = 8$ to 30 for mannitol and $n = 10$ to 29 for PEG-8000).

4.4.5.3.2 Hypocotyl elongation

Elongation of A. thaliana seedling hypocotyls under dark conditions were measured as an indicator of growth under osmotic stress conditions (van der Weele et al., 2000). Results were highly variable, with no significant improvement in hypocotyl elongation of any transgenic line compared to Col-0 (Figure 4.16).
4.4.5.3.3 Fresh weight after growth on osmotic stress medium

The fresh weight of seedlings transferred to osmotic stress medium was determined after fourteen days of growth. Measurements of pools of six seedlings were replicated three to five times. The fresh weight was presented as a percentage relative to seedlings grown on control (MS) medium (Figure 4.17). The fresh weights of seedlings grown on osmotic stress medium were reduced to an average of 49%, 36% and 19% of the growth on control MS media by PEG-8000 -0.5 MPa, mannitol 200 mM and mannitol 350 mM treatments, respectively. Results were highly variable with large standard deviations. Statistical analysis of fresh weight percentage differences between Col-0 and the transgenic lines were performed using pair-wise comparisons of the means using Student’s T-test. None of the transgenic lines resulted in statistically significant changes in fresh weight at the 95% confidence level, except line 5, that was significantly reduced compared to Col-0 when grown on PEG-8000 -0.5 MPa stress medium. This corresponded with the root elongation results shown in Figure 4.15.
Figure 4.17 Fresh weight of *A. thaliana* seedlings after 14 days osmotic stress. Six-day-old seedlings were transferred from normal MS medium to osmotic stress medium. Stress was imposed by infusing plates with a PEG-8000 solution to a $\psi_W$ of -0.5 MPa (PEG -0.5; hatched blue) or by including mannitol into the medium (200 mM and 350 mM; Man 200 (solid orange) and Man 350 (hatched orange), respectively). The fresh weight after 14 days is presented as a percentage relative to growth on control (MS) medium. Error bars represent the standard deviation ($n = 3$ to 5).

4.4.5.3.4 Fresh weight after recovery from osmotic stress

The fresh weights of osmotic-stressed seedlings (pools of six) were determined after a recovery period to evaluate their ability to resume growth after stress (Figure 4.18). No transgenic line showed improved growth compared to Col-0 after 11 day-stressed seedlings (on media with $\psi_W$ of -0.7 MPa) were transferred to recovery medium (MS) and allowed to grow for 13 days. Line 9 had a statistically significantly reduced fresh weight (42 mg) at the 95% confidence level compared to Col-0 (63 mg).
Figure 4.18 Fresh weight of *A. thaliana* seedlings after 13 days recovery from osmotic stress imposed by PEG-8000 (-0.7 MPa) for 11 days. Six-day-old seedlings were transferred from normal MS medium to plates infused with a PEG-8000 solution with a $\psi_W$ of -0.7 MPa. After 11 days on osmotic stress medium, seedlings were transferred to MS plates for recovery. The fresh weight of six pooled seedlings from the different lines after 13 days is presented in milligram (mg). Error bars represent the standard deviation ($n = 5$).

### 4.5 Discussion

#### 4.5.1 Over-expression of the *VuLEA5-1* gene in *A. thaliana*

LEA genes are suggested to improve crop plants’ abiotic stress tolerance by helping to avoid water deficit-associated injuries, improve survival rates and display improved recovery after stress (Fu *et al.*, 2007; Brini *et al.*, 2007; Dalal *et al.*, 2009). A LEA5 gene was identified from cowpea (*VuLEA5-1*) and tested in the model plant *A. thaliana* as a proof-of-concept to improve the drought tolerance of crops. Since the 76th position of *VuLEA5-1* and *VuLEA5-2* differs by a conservative amino acid substitution of isoleucine for valine, they were not expected to have any difference in protein function.

Transformation of *A. thaliana* took place by the *Agrobacterium* floral-dipping method (Clough and Bent, 1998). *A. thaliana* transformants were selected on multiple rounds of selective media until the progeny were homozygous for the selectable *Hygromycin-B* marker gene linked to the transgene. Eight independent transgenic events were selected to homozygosity.
The VuLEA5-1 transgene was required to be in both copies of the chromosome where it was integrated, so that it does not segregate in subsequent progeny generations. It was important that drought stress studies were performed on lines that stably inherit the transgene. Using the approach by Harrison et al. (2006), there was a distinct difference in hypocotyl length between transgenic and non-transformed seedlings germinated on hygromycin-containing media (Figure 4.3). Germination ratios of progeny seed was used to determine the zygosity status of the transgene in the genome. Plants heterozygous for a single insertion of the transgene produced seed in which the selectable marker segregated into 75% resistant and 25% non-resistant progeny seed. A homozygous transgenic plants’ seed was 100% resistant, since the selectable marker and linked transgene did not segregate.

Molecular characterisation of transgenic A. thaliana lines included PCR and Southern analysis for copy number detection. PCR successfully confirmed the presence of the VuLEA5-1 transgene in the genomic DNA extracted from transgenic lines (Figures 4.2 and 4.4). The Southern blot was probed with the complete VuLEA5-1 gene sequence labelled with the DIG-antigen (Figure 4.6). The VuLEA5-1 probe did not hybridise to fragments in the untransformed control (Figure 4.6, lanes 2 and 3). The probe was, therefore, sufficiently heterogeneous from the endogenous LEA5-like genes present in A. thaliana. When the VuLEA5-1 gene coding sequence was compared to its closest relative in the A. thaliana LEA gene family, AtDII1 [At4g15910], only 56% of the nucleotides were identical (LALIGN; http://www.ch.embnet.org).

Incomplete digestion of genomic DNA by Xhol (Figure 4.5) may be explained by the frequency of its digestion sites, or the methylation sensitivity of this restriction enzyme. Xhol is sensitive to CpG methylation, which is used by eukaryotes as a heritable and tissue specific method for differentiation and modulation of gene expression. The promoters of genes that are not expressed were found to be methylated (Siegfried and Cedar, 1997). Methylation can interfere directly with the binding of transcription factors, or through binding to transcriptional repressors called methyl-CpG-binding proteins (Nan et al., 1998).

4.5.2 Western blot analysis of VuLEA5-1 transgenic A. thaliana

A Western blot was employed to detect the expression of the VuLEA5-1 protein in transgenic A. thaliana. It depended on the availability of an anti-LEA5 antibody. LEA antibodies that were found in literature were mostly produced against group 2 (dehydrin) and group 3 LEA proteins (synthetic peptide of group 2 dehydrin consensus sequence (Close et al., 1993); group 3 barley HVA1 (Xu et al., 1996), wheat group 3 (Ried and Walker-Simmons, 1993)). They also detected non-specific bands in the plant protein extracts and could possibly cross-
react with LEA proteins from other groups due to the repeating amino acid motifs forming alpha-helices in most LEA proteins’ secondary structures. The only group 5 LEA antibodies mentioned in publications was produced against the D-34 and D-95 types. An antibody was produced against the CaLEA6 protein from *Capsicum annuum*, which is related to other D-95 LEA proteins (Kim et al., 2005). Another antibody was produced against a *Medicago truncatula* seed maturation protein (SMP), called MtPM25, and the research group leader was willing to share the antiserum (Boudet et al., 2006; Boucher et al., 2010). MtPM25 [GenBank: ABB16353] was identified in the heat stable proteome of *M. truncatula* radicles to be a LEA protein linked to desiccation tolerance (Boudet et al., 2006). It falls within the group 5 LEA proteins, but contains a different Pfam functional group (PF04927) than the VuLEA5-1 protein (PF03242). It encodes a 260 amino acid protein with a theoretical predicted and experimentally determined molecular mass of 25 kDa. The nomenclature according to Dure (1993) of this LEA protein group is D-34 based on the cotton proteins he discovered.

Battaglia et al. (2008) grouped three atypical, hydrophobic LEA proteins (LEA12, LEA5 and LEA14) that were previously found in cotton into group 5. Their cDNA's were labelled D-34, D-73 and D-95 (Baker et al., 1988; Dure, 1993; Galau et al., 1993). VuLEA5-1 is most similar to the D-73 cDNA. Although the MtPM25 and VuLEA5-1 proteins can both be grouped into group 5 by Battaglia et al. (2008), Hundertmark and Hincha (2008) showed that the *A. thaliana* LEA gene sequences of this group are clearly segregated into three lineages, corresponding to the cDNA's found in cotton. They all probably arose from a common ancestral group 5 gene sequence. The expected molecular mass of the transgenic VuLEA5-1 polypeptide is 11.7 kDa.

Western blotting using the α-MtPM25 antibody could not detect the transgenically expressed VuLEA5-1 protein (Figure 4.9). The sequence difference between these two proteins is proposed as the reason. When the amino acid sequences of MtPM25 and VuLEA5-1 were aligned, only 19 amino acids were identical (Figure 4.19). The two largest continuously aligned segments (36 and 56 amino acids) each contained 24 and 29 identical and similar amino acids, respectively.
### CLUSTAL 2.1 multiple sequence alignment

<table>
<thead>
<tr>
<th>MtPM25</th>
<th>MSQEOPRPOQAGPDKYGDVLPVSQGDSQKPITPEDAAMMQSAESRVLGQTQPGGASV 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>VuLEA5-1</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>MtPM25</td>
<td>MQSAAATNEQAG1VGHKDVTDIGRGTETQVPGRRITETGVQVEQPTPQV 120</td>
</tr>
<tr>
<td>VuLEA5-1</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>MtPM25</td>
<td>VGLTGAVRESALTIEEALEATAHTVGDKPVEQSDASAIQAAEVRATGSNVITPGGASMA 180</td>
</tr>
<tr>
<td>VuLEA5-1</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>MtPM25</td>
<td>QSAAAFNAECQREEEIKRMNVLTGATAKLPADKAAATRQDAAGVASEMRNNDATATPG 240</td>
</tr>
<tr>
<td>VuLEA5-1</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>MtPM25</td>
<td>GVAASVAAAARLNENVGNVM 260</td>
</tr>
<tr>
<td>VuLEA5-1</td>
<td>-------------------------------------------------------------</td>
</tr>
</tbody>
</table>

#### Figure 4.19 Amino acid sequence alignment of MtPM25 [ABB16353] and VuLEA5-1.

*A. thaliana* contains a complement of 51 LEA gene family-members. The six *A. thaliana* LEA proteins that corresponds to MtPM25 are classified into the LEA group called seed maturation family proteins (SMP) (Hundertmark and Hincha, 2008). The proteins that were detected in Figure 4.9 were probably native *A. thaliana* SMP proteins expressed in both the untransformed and transgenic lines. The six *A. thaliana* SMP (or D-34) proteins code for proteins with molecular weights between 17 and 27 kDa (Hundertmark and Hincha, 2008). MtPM25 is most similar to an *A. thaliana* SMP encoded by the gene At3g22490 (Arenas-Mena *et al.*, 1999; Borrell *et al.*, 2002) [GenBank: NP_188888]. This gene codes for a 262 amino acid protein with a calculated molecular weight of 26.6 kDa. The protein, called Atrab28, accumulates mostly in seed tissues and is not inducible in vegetative tissues by ABA, osmotic stress or dehydration. A nuclear localisation signal (NLS) consisting of a five amino acid motif QPKRP targets the protein to the nucleus (Borrell *et al.*, 2002). Sequence alignment of MtPM25 and Atrab28 (At3g22490) indicates 62.5% identity in 259 amino acid overlap using LALIGN (only 29 amino acid mismatches, the rest were identical or similar in properties) (Figure 4.20). It is unclear why nonspecific proteins of 14.4 kDa, with two faint signals in all lanes at 97 kDa and between 14.4 and 20.1 kDa, were immunodetected with the α-MtPM25 antibody. The *A. thaliana* SMP protein of 26.6 kDa was not expected due to its absence in vegetative tissues.
Singh et al. (2005) were able to elucidate a stable three-dimensional structure for *A. thaliana* LEA14, a group 5 LEA member of the D-95 type. There is, therefore, a possibility that the transgenic *VuLEA5-1* protein was not present in the heat-soluble protein extracts prepared from the transgenic *A. thaliana* lines due to heat denaturation. The MtPM25 protein, however, is heat stable, since it was identified from the heat stable proteome of *M. truncatula* (Boudet et al., 2006). Possible improvements to the methods employed in this study should include an ELISA of the extracts before and after heat treatment to determine in which extract, if any, the α-MtPM25 antibody detects the protein. A positive control extract could also be prepared from *M. truncatula* to verify the activity of the α-MtPM25 antibody.

4.5.3 RT-qPCR quantification of *VuLEA5-1* expression in transgenic *A. thaliana* lines

Semi-quantitative RT-PCR proved that the *VuLEA5-1* transcript was being produced in the transgenic *A. thaliana*, in all lines except line 20 (Figure 4.7). Expression of the transgenic protein could not be detected in *A. thaliana* lines with the antibody available for Western blot (α-MtPM25) (Figure 4.9). RT-qPCR was, therefore, employed to quantify the *VuLEA5-1* gene transcripts in transgenic *A. thaliana* (Figure 4.10), in order to select lines for drought stress studies. Since the *VuLEA5-1* gene was cloned into the pMDC32 transformation vector under the control of the 2x35S constitutive promoter, the aim was to identify the transgenic lines expressing the highest level of *VuLEA5-1* mRNA transcript, for further drought tolerance studies. The integration position may have an effect on the transcription levels of a transgene (Mirza, 2005). The T-DNA integration region may be methylated or highly repetitive, which
results in inactivation of the transgene. Copy number may also affect gene transcription levels. It is interesting to note that the two highest expressing lines, namely 9 and 12 (Figure 4.10), each contained only a single integration event of the transgene into the genome (Figure 4.6, lanes 11 and 13 for lines 9 and 12, respectively). Higher copy number integration events probably resulted in partial gene silencing, due to competition for the transcription initiation complex for the same promoter. Transgenes are able to silence homologous, unlinked loci, especially when they are present in high copy numbers (Matzke et al., 1994). The mechanism is through epigenetic trans-inactivation and DNA methylation, causing decreased steady-state mRNA levels (Assaad et al., 1993). The possibility is stated that the homologous DNA of the multiple copies pairs, then interacts with flanking heterologous DNA, which induces condensation of chromatin into a non-transcribable state.

Lines 5, 9, 12 and 21 were chosen as the highest expressing lines for in vitro osmotic stress assays. They contain either 1 (both lines 9 and 12), 2 or 3 copies (lines 21 and 5, respectively) of the transgene inserted into the genome.

**4.5.4 Drought stress response of transformed A. thaliana plants**

Various controlled drought and osmotic-stress experiments were performed on the VuLEA5-1 transgenic and wild-type Col-0 plants, to determine whether the introduction of VuLEA5-1 into A. thaliana could improve its tolerance to drought. An approach similar to that of Brini et al. (2007) was followed to characterise the possible improvement in drought tolerance of homozygous transgenic A. thaliana lines.

Both drought and osmotic stress treatments lead to cellular water deficit. It was hypothesised that the transgenic lines containing the VuLEA5-1 gene will tolerate drought and osmotic stresses better than control plants, due to the protective role of the LEA protein. This was deduced from the examples of creeping bentgrass that was transformed with a barley LEA3 member (hva1 gene; Fu et al., 2007), and A. thaliana transformed with a wheat dehydrin (group 2 LEA protein) (Dhn-5 gene, Brini et al., 2007). It was expected that the VuLEA5-1 protein would enable the transgenic plants to continue growing despite experiencing cellular water deficit, or that they would recover from drought stress faster and more complete, due to critical enzyme functions that are protected during cell water deficit, and that can resume activity immediately after rehydration.

The creeping bentgrass was assessed for leaf wilting and turf quality (green colour) through a visual scoring system, and transgenics showed less water deficit injuries compared to wild type (Fu et al., 2007). The RWC of transgenics were also significantly higher than wild type,
with a difference of 12% at both 6 and 9 days of water stress. Transgenic *A. thaliana* containing the *Dhn-5* gene (Brini *et al.*, 2007) showed stronger growth when grown in the presence of salt (50 mM to 200 mM NaCl) or during water deficit. Total leaf area was quantified and was less reduced in transgenics than wild type (20-25% reduction compared to 32% of the wild type at 50 mM NaCl). The RWC of transgenic plants was 13% higher at 4 days and 22% higher at eight days of water deficit stress, compared to wild type. Transgenic seedlings grown *in vitro* in the presence of 200 mM mannitol showed faster recovery than wild type, when root growth was measured. No quantitative data is, however, presented. The protection levels of these LEA-transgenic examples are, therefore, not very high, but authors could show it to be statistically significantly different than the untransformed control. There are many studies reported, e.g. Dalal *et al.* (2009), where no quantitative assessment was made on the tolerance of transgenic plants to drought and osmotic stresses.

Transformed plants’ responses to drought were scored on their survival rate after rewatering. Water-stress treatments of soil-grown plants lead to highly variable results. Plants of different sizes grew at different rates and used different amounts of water, therefore they did not dry out at the same rate when watering was withheld. Withholding water until plants became severely wilted resulted in them not recovering after rewaterting. From various repeated experiments it became evident that *A. thaliana* plants that wilted past a critical point were not able to revive. Three of the transgenic lines (lines 9, 12 and 21) showed higher survival rates than the negative control (line 20) and wild-type (Col-0) (*Table 4.1*). A higher number of replicated drought trials are, however, needed to make a definite conclusion.

### 4.5.5 *In vitro* osmotic stress response of transformed *A. thaliana* seedlings

Different osmotic agents were used to test the osmotic stress response of transgenic *A. thaliana* seedlings *in vitro*. High-molecular-weight PEG-8000 solutions were used to lower the $\psi_W$ of *in vitro* medium. Since gelling of solidified agar medium is prevented by the incorporation of large molecular weight compounds, a method of introducing PEG solutions into gelled media by diffusion was applied in this study (van der Weele *et al.*, 2000; Verslues *et al.*, 2006). Solutions containing PEG should not be autoclaved, since this changes the $\psi_W$. PEG-8000 was thus added to autoclaved liquid media and stirred in a covered container until dissolved. No detectable contamination was found following this approach. Water potentials of plates prepared this way are stable for at least a few days as long as plates are not allowed to dry out (Verslues *et al.*, 2006). By growing *A. thaliana* seedlings on agar-solidified nutrient medium equilibrated with PEG-solutions, the seedlings were subjected to a constant and defined $\psi_W$. Orienting the plates vertically allows the roots to grow on the surface of the
medium, thereby minimising the possibility of oxygen limitation of the roots (van der Weele et al., 2000).

The *in vitro* germination assay involves the germination of seeds on MS media containing osmotic agents, such as NaCl and mannitol, and the percentage germination scored after two weeks. This assay has been used in various studies to study the protective effect that different groups of transgenically expressed LEA proteins contribute to transgenic *A. thaliana* (Brini et al., 2007; Dai et al., 2011). Brini et al. (2007) found a 50-60% decrease in germination capacity of Dhn-5 transgenic *A. thaliana* seeds on 100 mM NaCl, while wild type seed germination was reduced by 90%. Higher ion content (Na\(^+\) and K\(^+\)) was found in leaves of salt-stressed plants. The authors suggest that DHN-5 could facilitate ion sequestration into the vacuole in transgenic plants, to limit ion toxicity and create a higher solute content in the vacuole so that plants can survive under low soil water potential. Seed germination of *A. thaliana* transformed with a group 7 LEA gene, *MpAsr*, showed less sensitivity to osmotic stress media (Dai et al., 2011). The authors suggest it is due to the hydrophilicity of the *MpAsr* protein.

A factor that was considered was that seed age differences affect germination rates, therefore, all lines that were compared were harvested on the same date. Germination assays of *VuLEA5-1* transgenic seedlings did not give reproducible results. NaCl (100 mM and higher) inhibited germination in all lines. Examples from the literature indicate that LEA proteins are not always able to improve seed germination percentages of transgenic *A. thaliana*. Over-expressing AtLEA4-5 did not improve germination ability of homozygous transgenic *A. thaliana* lines (Olvera-Carrillo et al., 2010). Also Dalal et al. (2009) could not show a difference in germination rate between wild type and *A. thaliana* transformed with another group 4 LEA gene from *Brassica napus*.

Subsequent experiments were, therefore, performed where seedlings were germinated on control media first and then transferred to osmotic stress medium when they were approximately one week old. Seedlings were assessed for growth during osmotic stress by measuring root and hypocotyl elongation and the fresh weight of the plants (Figures 4.13 to 4.18). A higher seedling growth rate was expected from the transgenic seedlings compared to the controls. A *Brassica* group 4 LEA resulted in higher root and shoot growth of the transgenic *A. thaliana* seedlings than wild type on MS media containing NaCl or mannitol (Dalal et al., 2009). However, no quantitative data was presented to support this claim. Root growth of *A. thaliana* seedlings was more vigorous in the light than in the dark (van der Weele et al., 2000), therefore light-grown plants were used for root-elongation and fresh
weight studies.

Previous studies with _A. thaliana_ showed that primary root elongation was stimulated slightly by moderate water deficit under light growing conditions (van der Weele _et al._, 2000). Roots are known to be less affected by water deficit than shoots. Water deficit increases the ratio of root: shoot dry weight in plants, and this was found to be true also for _A. thaliana_. This phenomenon of continued growth under water deficit conditions allows the plant to search for water in the drying soil. Seedlings may be especially dependent on this ability in order to establish in soil under water deficit conditions. According to a previous report, severe deficit reduced root elongation of wild type _A. thaliana_, but only partially, since at the lowest $\psi_W$ (-1.2 MPa) elongation was still 50% of the control (van der Weele _et al._, 2000). This study, however, found that low $\psi_W$ (-1.2 MPa) stress severely inhibited root elongation, to approximately 4% of the control (Figure 4.13). The same ecotype was used during both studies (Col-0), but genotype and seedling viability differences could have occurred through local selections during propagation. The difference in root elongation sensitivity to low $\psi_W$ stress can also be caused by differences in media composition used during the two studies.

The hypocotyl growth of _A. thaliana_ seedlings was shown during previous studies to have a hyperbolic dependence on $\psi_W$ (van der Weele _et al._, 2000). Hypocotyl elongation under light growing conditions was too low to be measureable by the time seedlings were transplanted to low $\psi_W$ stress medium. Seedlings were, therefore, germinated and grown on different $\psi_W$ stress media in the dark during this study. A carbon source was supplied in the form of sucrose, to sustain seedling growth under non-photosynthesising conditions. Hypocotyl elongation results were highly variable and no significant improvement in hypocotyl elongation was observed for any of the _VuLEA5-1_ transgenic lines compared to Col-0 (Figure 4.16). Hypocotyl elongation was also very slow, growing only 5.7 mm (on average) on control medium in 17 days, which may account for the high variability. Experimental results reported by van der Weele _et al._ (2000) shows hypocotyl elongation of 13 mm after only 3 days.

No obvious difference was detected between wild type and transgenic _A. thaliana_ containing the wheat dehydrin gene _DHN-5_, a group 2 LEA, when grown on mannitol, but recovery was faster when transferred to normal medium (Brini _et al._, 2007). The fresh weights of seedlings after a recovery period from osmotic stress were, therefore, also measured. No transgenic line had improved recovery compared to Col-0 (Figure 4.18). Line 9, however, had a statistically significantly reduced fresh weight compared to Col-0 (Figure 4.18). This can be attributed to the fact that independent transgenic events may have different growth rates.
The transgene was inserted into different genome positions and may disrupt genes involved in growth, or the overproduction of a transgenic protein may result in fewer resources available for growth.

In conclusion, transgenic VuLEA5-1 A. thaliana plants and seedlings did not show any improved growth or recovery to drought and osmotic stress treatments, when compared to the wild-type Col-0. Another example from literature has also found that over-expressing LEA genes do not always result in improved stress tolerance. LEA genes from the resurrection plant Craterostigma plantagineum did not confer higher stress tolerance to transgenic tobacco (Iturriaga et al., 1992). It is proposed that one LEA gene, in isolation, may not be sufficient, or that it did not function in the heterologous plant as in their native context (Bray, 1993).

During this study, quantitative growth measurements suggested by literature of transgenic A. thaliana seedlings were employed as a measure of tolerance to osmotic stress conditions simulating drought. The type of measurements following drought stress treatments employed may not have picked up any improvement, which may potentially have been very slight. LEA proteins are not the only protective measure employed by plants challenged by drought. By over-expressing a single LEA gene, only a subset of proteins or cellular compartments may be protected through specific interactions, which may not be enough to confer tolerance to the whole seedling during drought stress. Another possibility that cannot be excluded is silencing of the transgene after the transcription level. The lack of confirmation by Western blot of protein expression, and no phenotypic improvement in drought tolerance, may indicate that over-expression of the functional protein did not occur. Silencing mechanisms in the plant at the post-transcriptional and translation levels may have prevented it.

From the expression patterns of VuLEA5 in cowpea (Chapter 3, Figure 3.15), the function of VuLEA5 may be to enhance protection against cold stress, which were not analysed in this study. Measurements of growth parameters during osmotic stress in vitro, simulating drought stress in soil, were not designed to detect any of these functions. This may be investigated in a follow-up study.
4.6 Acknowledgements

We acknowledge the FMG group, Department of Genetics, University of Pretoria for *A. thaliana* reference gene primers (ACT2). Elongation factor 1-alpha (EF1A) was suggested by Mr. N Olivier (Plant Science Department, University of Pretoria).

We thank the FMG group, Department of Genetics, University of Pretoria, for providing the Gateway destination vector pMDC32 used in this study and Dr. C Papdi (Hungary) for providing the *Agrobacterium tumefaciens* strain GV3101 (pMP90).
Concluding remarks

In this study, a customised cDNA library was constructed containing cowpea genes expressed during responses to drought (Chapter 2). Highly responsive individual genes were identified by screening the library with microarrays. This library helped collaborators develop the SSHscreen-SSHdb software pipeline, which greatly facilitates gene discovery in any biological system, particularly non-model organisms. An improved selection of clones for sequencing can be made after screening the library on a small number of microarrays. The pipeline also manages the sequence data and helps with the annotation of the genes.

The combination of these methods allowed the identification of multiple drought-responsive genes from cowpea, an organism without an annotated genome sequence. The first hypothesis was therefore proven true, that drought stress of cowpea would result in differential expression of genes involved in the protection against drought stress.

The SSH library was constructed by subtracting genes expressed by a drought susceptible line Tvu7778 from a tolerant line IT96D-602 that was drought-stressed. We were interested in genes up-regulated by drought and the genetic differences that exist between the two lines that differed in drought tolerance. Some genes may be constitutively expressed at higher or lower levels, or are completely novel, in the tolerant line IT96D-602 compared to the drought susceptible line Tvu7778. An alternative way of identifying genes induced under drought would be to construct the SSH library with one cowpea line only, treated with drought or watered as normal.

An important group of stress-regulated genes, those coding for transcription factors that regulate the transcription of target genes involved in stress adaptation and tolerance, were unfortunately not identified in the list of sequenced clones following SSH. However, not all the clones were sequenced, therefore the remaining library could still contain cDNA coding for transcription factors.

A LEA5 gene was identified to be up-regulated in the cowpea drought expression library (Chapter 2). LEA proteins play important protective roles during seed desiccation, drought and other abiotic stresses of plants. Even though the LEA gene family has been widely studied, the mechanisms whereby these proteins function are complex and not completely understood. The cowpea LEA5 protein is atypical in that it is more hydrophobic and more structured in solution than typical hydrophilic LEA proteins. Classifications of LEA groups
have changed over time, with different authors either grouping LEA5 proteins together based on their hydrophobicity, or splitting them into their three different types (D-34, D-73 and D-95) identified first from cotton. The cowpea LEA5 protein is of the D-73 type. It is suggested that these types remain separated, based on their Pfam domain content and protein length differences. The three types are also clearly distinguishable, based on their amino acid alignments in the phylogram in Figure 3.8. Due to their atypical hydrophobicity, when all other groups of LEA proteins are highly hydrophilic, it has been suggested that this group does not belong to the LEA protein family (Tunnacliffe and Wise, 2007). However, due to similarities in expression patterns and related functions, the continued grouping of LEA5 proteins in the LEA protein family is recommended. In defense of this recommendation, not all hydrophilic plant proteins are classified as LEAs, and not all LEA proteins are even expressed during late embryogenesis of seeds.

Due to the protective role of LEA proteins on cellular structures during drought (Battaglia et al., 2008), these proteins are thought to form part of an active response during plant drought tolerance, rather than drought avoidance. VuLEA5 was over-expressed in A. thaliana in order to test its role in drought tolerance of plants (Chapter 4). It was hypothesised that the transgenic lines containing the VuLEA5-1 gene would tolerate drought and osmotic stresses better than control plants, due to the protective role of this LEA protein on cellular structures that are damaged during drought.

Type 1 drought avoidance and type 2 drought tolerance mechanisms of plants were introduced in Chapter 1. The type 2 drought tolerance response is the most attractive source of genes to confer tolerance to other crops. Incorporating genes leading to drought avoidance will be detrimental to crop yield, since the plants halt their growth and photosynthesis is reduced due to closed stomata. True tolerance mechanisms protect growing cells from water stress injury, through the production of protective solutes and proteins, such as LEA proteins, and detoxification of reactive oxygen species (ROS).

Various methods of quantitative growth measurements during drought and osmotic stress from the literature were employed to investigate whether the transgenic plants had improved drought tolerance compared to the wild type. Due to its small size it is challenging to collect growth data on this plant. Plant growth in soil during drought trials was variable, therefore, in vitro seedling stress tests were performed in the presence of osmotic agents that simulate drought conditions in the soil. Tested conditions included seed germination and seedling growth on in vitro medium supplemented with NaCl, mannitol or infiltrated with PEG-8000 to reduce the medium’s water potential. Measured parameters included hypocotyl and primary...
root elongation, and fresh weight of seedlings after growth or recovery from osmotic stress medium. No statistically significant improvement in growth and recovery after stress could however be attributed to \textit{VuLEA5}.

This LEA protein alone could, therefore, not improve the drought tolerance of transgenic \textit{A. thaliana} lines. The second hypothesis, that the cowpea \textit{LEA5} protein gene plays an important role in the drought tolerance of plants, is, therefore, proven false. It has been suggested by previous authors that the function of a LEA protein in isolation may not be sufficient to confer drought tolerance, or it may not function in the heterologous plant as in their native context (Bray, 1993). LEA proteins are not the only protective measures employed by plants challenged by drought. By over-expressing a single LEA gene, only a subset of proteins or cellular compartments may be protected through specific interactions, which may not be enough to confer tolerance to the whole seedling during drought stress. It is suggested that LEA genes should be used in combination with other genes when a transgenic plant needs to be created that is more drought tolerant.

The type of drought tolerance assays employed during this study may have been unable to pick up any improvement in drought tolerance. Cell protection may potentially be very slight and other assays may be necessary to detect different damage levels at a cellular level. Abiotic stresses cause ROS accumulation, which causes lipid peroxidation. Malondialdehyde (MDA) is a product of polyunsaturated fatty acid peroxidation in plants. MDA levels during stress conditions are, therefore, a marker for evaluating stress tolerance of plants (Gawel et al., 2004). Electrolyte leakage is another marker quantifying cell membrane damage (Dai et al., 2011).

\textit{A. thaliana} was cited as a useful model plant for studying the biological effect of transgenes (Clough and Bent, 1998). For the particular transgene transformed during this study, \textit{VuLEA5-1}, it was perhaps not an optimal model. Since it has a short life cycle, it may adopt drought avoidance mechanisms during drought stress. The protective effect of LEA proteins, leading to drought tolerance, may then be irrelevant. Also, its genome contains a complement of 51 LEA gene family members (Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008), with four being \textit{LEA5} homologs expressed during similar conditions as the cowpea \textit{VuLEA5-1}. The effect of the over-expressed \textit{VuLEA5-1} transgene may have been masked by the abundant endogenous \textit{LEA5} transcripts already present in the drought-stressed \textit{A. thaliana} plants.

The role of LEA5 could also have been tested against other stress conditions, including UV
light, ethylene and wounding. A tomato cDNA with high homology to the atypical group 5 LEA proteins, ER5, was shown to be ethylene- and wounding-responsive (Zegzouti et al., 1997).

Another way to study the role of LEA proteins is to purify them and use them in \textit{in vitro} protection assays of enzymes that are sensitive to denaturation due to dehydration or freezing. The results from such studies should be interpreted with caution (Tunnacliffe and Wise, 2007). The fact that they are sometimes able to protect enzymes from denaturation \textit{in vitro} is not necessarily due to their unique protein identity. It may be a non-specific protein stabilisation effect without any useful desiccation protection effect \textit{in vivo}. LEA proteins are not the only hydrophilic proteins that are up-regulated in response to dehydration of plants. Plants dealing with water loss launch an array of protection and repair mechanisms, of which LEA proteins make out only one facet. How the LEA proteins interact with other defence strategies may be the key to desiccation protection. It was, therefore, important in this study to generate over-expressing transgenic lines to study the \textit{in vivo} role of the VuLEA5 protein (Chapter 4). The interaction of the over-expressed VuLEA5 protein with other drought-tolerance molecules or proteins induced in the plant may have provided the required protection to perform better under drought.

Of all the abiotic stress conditions tested in this study, cold stress and ABA was the most active in inducing VuLEA5 transcript accumulation (Chapter 3, Figure 3.15). It shares this cold stress-inducibility with three of its closest relatives in \textit{A. thaliana}. A related LEA5 protein of the D-73 type, a gene from \textit{Tamarix androssowii}, has only recently been shown to enhance cold tolerance of transgenic plants (Zhao et al., 2011). Enhanced activity of peroxidase (POD) and superoxide dismutase (SOD) enzymes and significantly less lipid peroxidation under low temperature stress conditions have been observed in transgenic blueberry. The role of this LEA in protection against low temperature stress was hypothesised to be through the production of POD and SOD, but the authors did not give any mechanisms by which the LEA protein may induce these enzymes. A previous study using the same gene was able to enhance the drought tolerance of transgenic tobacco (Wang et al., 2006).

The hydrophobic nature of the VuLEA5 protein, and the likelihood to form stable three-dimensional structures in solution, but especially upon cell drying (Chapter 3), relates it to an unusual intrinsically disordered protein from the legume \textit{Lotus japonicus} (GenBank: ACJ46652). This protein acted as a chaperone that prevented misfolding and aggregation, and preserved the activities of two model enzymes \textit{in vitro} when they were dehydrated or
frozen (Haaning et al., 2008). Transgenic *A. thaliana* seedlings were not subjected to cold stress during this study, so the possible protective effect of VuLEA5 under cold conditions was not assayed. Measurements of growth parameters during osmotic stress *in vitro*, simulating drought stress conditions in soil, were not designed to detect this possible function. A possible role of VuLEA5 in the protection of cellular structures subjected to cold stress is, therefore, suggested as a follow-up study.

Over-expressing dehydration-induced genes may result in reduced growth under normal growing conditions (Park et al., 2011). Constitutive over-expression of a transgene under normal conditions, when the gene product is not required, uses valuable energy that may put the plant at a disadvantage. Using a stress-activated promoter instead of a constitutive promoter (such as 35S CaMV) might minimise the negative effects on plant growth. A promoter that is ABA-responsive is suggested, as this phytohormone has been shown in various studies to be up-regulated in response to water deficit (Iuchi et al., 2000). It regulates plant responses to various abiotic stresses, such as water deficit, cold and salt. ABA-responsive elements, or ABRE’s, are sequence motifs found in the promoters of many stress-responsive genes, and are recognised by families of transcription factors to initiate transcription during specific stress conditions (Zhu, 2002). An example where an ABA-inducible promoter was used in the generation of a transgenic crop is creeping bentgrass, that was transformed with a barley LEA3 member gene, *hva1* (Fu et al., 2007). The ABA-inducible promoter was derived from the ABA-response complex (ABRC3). Transformants had low transgene expression during well-watered conditions, but the transgene was up-regulated during water deficit. It will, therefore, be wise to incorporate a stress-inducible promoter in the construct used to generate the final transgenic crop plants.

Transcription factors (TF) are early-response components of the drought signal transduction pathway, and can be applied in transgenic plants to generate plants more tolerant to water stress. They have the potential ability activate a cascade of downstream delayed response drought-protective genes. A successful example is the over-expression of an *A. thaliana* DREB (dehydration response element binding) TF, the *DREB1A/CBF3* cDNA, in *A. thaliana*. Transgenic plants showed increased tolerance to all abiotic stress conditions, namely dehydration, salt and freezing, since a wide range of stress responsive genes were switched on (Kasuga et al., 1999; Liu et al., 1998). In another *A. thaliana* example, the ectopic expression of the CBF1 transcription factor lead to the constitutive expression of downstream delayed-response genes, which lead to enhanced freezing stress tolerance of the plants (Jaglo-Ottosen et al., 1998).
The field of LEA protein research is complicated by the many gene family members present in plants and their diverse expression patterns, putative functions and distribution within plant organs. This study has added some expression characteristics and provides the functional characterisation of a novel LEA5 protein, identified from cowpea, to the current knowledge. Another contribution of this study has been to add to the identification and functional annotation of drought responsive cowpea genes, a crop whose genomics has become important in recent years due to its drought tolerant attributes.
References


References


de Bianchi S, Dall'Osto L, Tognon G, Morosinotto T and Bassi R (2008). Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits


© University of Pretoria


Russouw PS, Farrant J, Brandt W and Lindsey GG (1997). The most prevalent protein in a heat-treated extract of pea (Pisum sativum) embryos is an LEA group 1 protein; its conformation is not affected by exposure to high temperature. Seed Science Research 7: 117-123.


Umezawa T, Yoshida R, Maruyama K, Yamaguchi-Shinozaki K and Shinozaki K (2004). SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling


Crop breeding for drought tolerance is very important in South Africa. Biotechnology can be applied to improve a crop’s drought tolerance through candidate gene identification, followed by transformation or molecular breeding strategies. Cowpea was utilised as a gene source due to its inherent drought tolerance. A cDNA library, containing 4160 clones and enriched for drought responsive cowpea genes, was generated following suppression subtractive hybridisation (SSH) on two lines differing in their drought tolerance. Microarray hybridisation was used to screen the library for differentially expressed genes. Microarray data was analysed with a custom software package called SSHscreen. The SSH procedure was successful, since it enriched the library for transcripts differentially expressed by drought stress and rare transcripts.

Selected clones were sequenced and several cowpea genes responding to drought stress were identified. Among these was a gene from the well-studied late embryogenesis abundant (LEA) family. Other stress-induced genes identified included a glutathione S-transferase, a thaumatin, a universal stress protein and a wound induced protein. A lipid transfer protein and several components of photosynthesis were down-regulated by the drought stress. SSHdb, a web-accessible database, was used to manage the clone sequences, group redundant clones, annotate the sequence information, and combine the SSHscreen data with sequence annotations obtained from BLAST and Blast2GO. A high frequency of redundancy existed in the library, with only 39 unique sequences among the 118 sequenced clones. SSHscreen plots proved to be a useful tool for choosing anonymous clones for sequencing, since redundant clones cluster together on the enrichment ratio plots. The gene expression patterns of selected cowpea genes were verified with reverse-transcriptase quantitative PCR (RT-qPCR).

The gene encoding an atypical group 5 LEA protein (VuLEA5) was selected for further characterisation. LEA proteins play an important protective role during drought and other abiotic stresses of plants, but group 5 proteins are not as well studied as the other LEA groups. VuLEA5 was shown to be highly up-regulated by drought, abscisic acid (ABA), salt and cold in cowpea leaves.

VuLEA5 was over-expressed in the model plant Arabidopsis thaliana to study its role in tolerance to drought. The integration of the transgene into the genome was analysed by PCR and Southern blot. mRNA expression was detected by RT-PCR and RT-qPCR, and
recombinant protein expression by Western blot. Transgenic *A. thaliana* lines were investigated for improved drought tolerance compared to the wild type. Seedling stress tests were performed *in vitro* in the presence of osmotic agents (high molecular weight polyethylene glycol) that simulate drought conditions in the soil. Quantitative growth parameters (hypocotyl and primary root elongation) of homozygous transgenic T4 seedlings were measured. No statistically significant improvement in growth and recovery after stress could, however, be attributed to the cowpea LEA5. A number of progressive drought trials of plants growing in a growth room did not indicate any quantitative improvement in their ability to withstand drought. This LEA5 protein alone could, therefore, not improve the drought tolerance of transgenic *A. thaliana* lines. Its possible role in cold stress protection is suggested as a follow-up study.
## Appendix 1

Oligonucleotide primers used for the amplification of cowpea and *A. thaliana* genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Origin/ GenBank accession</th>
<th>Primer name</th>
<th>Sequence (5' – 3')</th>
<th>PCR product expected from cDNA gDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13-26-R</td>
<td></td>
<td></td>
<td>M13-26-R</td>
<td>CAGGAAACAGCTATGAC</td>
<td></td>
</tr>
<tr>
<td>SP6</td>
<td></td>
<td></td>
<td>SP6</td>
<td>ATTTAGGTGACACTATAG</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td></td>
<td></td>
<td>T7</td>
<td>TAATACGACTCACTATAGG</td>
<td></td>
</tr>
<tr>
<td>26S_F</td>
<td><em>V. unguiculata</em> 26S ribosomal RNA</td>
<td>Library clone 38G04-R</td>
<td>26S_F</td>
<td>GGAATCGAGAGCTCCAAGTG</td>
<td>198</td>
</tr>
<tr>
<td>26S_R</td>
<td></td>
<td></td>
<td>26S_R</td>
<td>GTTGATTCGGCAGGTGAGTT</td>
<td></td>
</tr>
<tr>
<td>ACT2_F</td>
<td><em>A. thaliana</em> Actin 2</td>
<td>AT3G18780; NM_180280.1</td>
<td>ACT2_F</td>
<td>CATCACGATTGGTGCTGAGA</td>
<td>200 200</td>
</tr>
<tr>
<td>ACT2_R</td>
<td></td>
<td></td>
<td>ACT2_R</td>
<td>TCATACGGTGCGAGATACCT</td>
<td></td>
</tr>
<tr>
<td>BeanActinF</td>
<td>Bean-specific actin</td>
<td>Consensus from: AY338231, AY338230 and AB067722</td>
<td>BeanActinF</td>
<td>ACTGAGGCACCGCTAACC</td>
<td>170 280</td>
</tr>
<tr>
<td>BeanActinR</td>
<td></td>
<td></td>
<td>BeanActinR</td>
<td>GTGTGACTGACACCATCACC</td>
<td></td>
</tr>
<tr>
<td>CBP_F</td>
<td><em>A. thaliana</em> Cap-binding protein 20</td>
<td>AT5G44200; NM_123787.3</td>
<td>CBP_F</td>
<td>AGCTCGTGGATTACGGTACT</td>
<td>237 660</td>
</tr>
<tr>
<td>CBP_R</td>
<td></td>
<td></td>
<td>CBP_R</td>
<td>TCTCCGTTCTCATGTCACTA</td>
<td></td>
</tr>
<tr>
<td>CHL_F</td>
<td><em>V. unguiculata</em> Photosystem II (LHCII) type I chlorophyll a/b-binding protein</td>
<td>Library clone 44C07-R</td>
<td>CHL_F</td>
<td>CTCATCCACGCTCAGAGCAT</td>
<td>240</td>
</tr>
<tr>
<td>CHL_R</td>
<td></td>
<td></td>
<td>CHL_R</td>
<td>CTGGACGAAGAAGCCGAACA</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Species</td>
<td>Description</td>
<td>Accession Numbers</td>
<td>Primer Sequences</td>
<td>Length</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>EF1A</td>
<td><em>A. thaliana</em></td>
<td>Elongation factor 1-alpha</td>
<td>AT5G60390; NM_125432.3</td>
<td>EFA_1: AGGCTGCTGAGATGAACAAG  EFA_2: CCAGTGGTGAGTCATGAT</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPC</td>
<td>Bean-specific</td>
<td>Glyceraldehyde-3-Phosphate dehydrogenase C2</td>
<td>Consensus from: AC135505_Mt (exons only), DQ192668, DQ355800, PEAGAPCI</td>
<td>GAPC_F: ATCAGCCAAGACTGGAGAG  GAPC_R: ACGGAATGCCATACCAGTCA</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td><em>A. thaliana</em></td>
<td>Glyceraldehyde-3-Phosphate dehydrogenase C2</td>
<td>AT1G13440; NM_101214.3</td>
<td>GAPDH_5: GTCACGCTCAGTGGAAAGCACC  GAPDH_6: GCCGCATCTTTGCTTGGGCG</td>
<td>239 416</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Globin</td>
<td>Human beta-globin</td>
<td></td>
<td>NM_000518</td>
<td>Globin_F: GGAGAAGTCTGCGTTACTG  Globin_R: GCCATGAGCCCTCACCCTAG</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm40S</td>
<td><em>G. max</em></td>
<td>40S ribosomal protein S8</td>
<td>TC100533 (MGI)</td>
<td>Gm40S_F: GCCAGCGTGCTAACACTAAG  Gm40S_R: AAGAGTCTGAGTGCCACAAG</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GmELF1B</td>
<td><em>G. max</em></td>
<td>Elongation factor 1-alpha</td>
<td>TC203623 (MGI)</td>
<td>GmELF1B_F: GGTGAAAGGCCAGGGGACA  GmELF1B_R: TCTTACCCTTGAGCGTG</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GST</td>
<td><em>V. unguiculata</em></td>
<td>Glutathione-S-transferase</td>
<td>Library clone 25B06-F</td>
<td>GST_F: GCTGGTGAAGGTGTGGGATA  GST_R: CCACGATGGTCTGCCTACTTA</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEA5</td>
<td><em>V. unguiculata</em></td>
<td>LEA5 (coding)</td>
<td>Library clone 07F09-F</td>
<td>LEA5_start: AAGATGTCTCTTCTTCTCTTCTTCT  LEA5_stop: CTATTGTGGCGATTTGAAACT</td>
<td>327 418</td>
</tr>
<tr>
<td>Gene/Protein</td>
<td>Organism</td>
<td>Library Clone</td>
<td>Accession</td>
<td>Primer Forward</td>
<td>Primer Reverse</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>LEA5</td>
<td><em>V. unguiculata</em></td>
<td>LEA5_F</td>
<td></td>
<td>LEA5_F</td>
<td>LEA5_R</td>
</tr>
<tr>
<td>LTP</td>
<td><em>V. unguiculata</em></td>
<td>Library clone</td>
<td></td>
<td>LTP_F</td>
<td>LTP_R</td>
</tr>
<tr>
<td>NCED</td>
<td><em>V. unguiculata</em></td>
<td>AB030293</td>
<td></td>
<td>NCEDcod_F</td>
<td>NCEDcod_R</td>
</tr>
<tr>
<td>THAU</td>
<td><em>V. unguiculata</em></td>
<td>Library clone</td>
<td></td>
<td>THAU_F</td>
<td>THAU_R</td>
</tr>
<tr>
<td>TUB2</td>
<td><em>A. thaliana</em></td>
<td>AT5G62690; NM_125664.3</td>
<td></td>
<td>TUB_5</td>
<td>TUB_6</td>
</tr>
<tr>
<td>VuELF1B</td>
<td><em>V. unguiculata</em></td>
<td>AB588747</td>
<td></td>
<td>VuELF1B_F</td>
<td>VuELF1B_R</td>
</tr>
</tbody>
</table>
Appendix 2

SSHscreen R script, Targets file and SpotTypes file used for ER3 analysis of cowpea SSH libraries

**Appendix 2.1 SSHscreen R script used for ER3 analysis of cowpea SSH libraries.** R script used for ER3 analysis of both forward and reverse cowpea SSH libraries together after loading the limma 2.16.5 and SSHscreen 2.0.1 libraries in R 2.8.1. The working directory contained the Targets file (**Appendix 2.2**), SpotTypes file (**Appendix 2.3**), the 12 Genepix Results (gpr) and the GenePix Array List (GAL) file (available at GEO accession number GSE20273).

```r
# Cowpea
# 12 slides
# ER3 analysis with F and R library

library(SSHscreen)

cowpeaSSH<- SSHscreen(path = "/Users/Nanette/SSHscreen/cowpea_F_and_R", source = "genepix", norm.plot = TRUE, mfrow = c(4,2), legend = TRUE,bc.method = "normexp", wa.method= "printtiploess", ba.method = "Aquantile", irregular = TRUE, ndups = 2, spacing = 1, spot.ave = FALSE, method = "ER3", toplist = "all", adjust = "fdr", negflags = 0, offset = 50, weights = TRUE, library = "both", sort = "p", cutoff = 0.05, proportion = 0.5)

write.table(cowpeaSSH$tt.ud.F,file="/Users/Nanette/SSHscreen/cowpea_F_and_R/cowpea$tt.ud.F.txt",sep="\t")
write.table(cowpeaSSH$tt.ar.F,file="/Users/Nanette/SSHscreen/cowpea_F_and_R/cowpea$t$t.ar.F.txt",sep="\t")

write.table(cowpeaSSH$tt.ud.R,file="/Users/Nanette/SSHscreen/cowpea_F_and_R/cowpea$tt.ud.R.txt",sep="\t")
write.table(cowpeaSSH$tt.ar.R,file="/Users/Nanette/SSHscreen/cowpea_F_and_R/cowpea$t$t.ar.R.txt",sep="\t")
```

© University of Pretoria
Appendix 2.2 Targets file used for ER3 analysis of cowpea SSH libraries. Targets file listing the raw microarray slide data used for the SSHscreen ER3 analysis described in Appendix 2.1.

<table>
<thead>
<tr>
<th>SlideNumber</th>
<th>FileName</th>
<th>Cy3_F</th>
<th>Cy5_F</th>
<th>Cy3_R</th>
<th>Cy5_R</th>
</tr>
</thead>
<tbody>
<tr>
<td>673</td>
<td>FR673_UT_UC.gpr</td>
<td>UT</td>
<td>UC</td>
<td>UT</td>
<td>UC</td>
</tr>
<tr>
<td>18</td>
<td>FR018_UC_UT.gpr</td>
<td>UC</td>
<td>UT</td>
<td>UC</td>
<td>UT</td>
</tr>
<tr>
<td>669</td>
<td>FR669_UT_UC.gpr</td>
<td>UT</td>
<td>UC</td>
<td>UT</td>
<td>UC</td>
</tr>
<tr>
<td>668</td>
<td>FR668_UC_UT.gpr</td>
<td>UC</td>
<td>UT</td>
<td>UC</td>
<td>UT</td>
</tr>
<tr>
<td>34</td>
<td>F034_ST_UT.gpr</td>
<td>ST</td>
<td>UT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>F022_UT_ST.gpr</td>
<td>UT</td>
<td>ST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>R017_SC_UC.gpr</td>
<td></td>
<td>SC</td>
<td>UC</td>
<td></td>
</tr>
<tr>
<td>672</td>
<td>R672_UC_SC.gpr</td>
<td></td>
<td>UC</td>
<td>SC</td>
<td></td>
</tr>
<tr>
<td>671</td>
<td>F671_ST_UT.gpr</td>
<td>ST</td>
<td>UT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>674</td>
<td>F674_UT_ST.gpr</td>
<td>UT</td>
<td>ST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>670</td>
<td>R670_SC_UC.gpr</td>
<td></td>
<td>SC</td>
<td>UC</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>R019_UC_SC.gpr</td>
<td></td>
<td>UC</td>
<td>SC</td>
<td></td>
</tr>
</tbody>
</table>

Appendix 2.3 SpotTypes file used for ER3 analysis of cowpea SSH libraries. SpotTypes file defining the control and cDNA spot types used for the SSHscreen ER3 analysis described in Appendix 2.1.

<table>
<thead>
<tr>
<th>SpotType</th>
<th>Name</th>
<th>ID</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>blank</td>
<td>BLANK</td>
<td>*</td>
<td>NA</td>
</tr>
<tr>
<td>cDNA_F</td>
<td>*-F</td>
<td>*</td>
<td>blue</td>
</tr>
<tr>
<td>cDNA_R</td>
<td>*-R</td>
<td>*</td>
<td>yellow</td>
</tr>
<tr>
<td>control_Globin</td>
<td>control_Globin</td>
<td>*</td>
<td>brown</td>
</tr>
<tr>
<td>control_NPTII</td>
<td>control_NPTII</td>
<td>*</td>
<td>red</td>
</tr>
<tr>
<td>control_GFP</td>
<td>control_GFP</td>
<td>*</td>
<td>green</td>
</tr>
<tr>
<td>control_ITS</td>
<td>control_ITS</td>
<td>*</td>
<td>purple</td>
</tr>
</tbody>
</table>
## Appendix 3

**Nucleotide and protein sequences of a LEA5 gene identified from cowpea**

**GenBank accession of SSH library clone 07F09-F**

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;GR942575</td>
<td>60</td>
</tr>
<tr>
<td>GAAGACTCAAAGAAATGGTGTTGAGCTGCTGCCAAATACACACTTAGGTTTTTCGCAG</td>
<td>60</td>
</tr>
<tr>
<td>CAGCAAACATTATAATAAGATGTCCTCCTCTCTCGCAAGCCAAATCTCTCGCTCCT</td>
<td>120</td>
</tr>
<tr>
<td>TCTTCCCTAGCTATCTCACAATTCTCTGATCGGCGAGGTTATGCAGTTGCGTCTG</td>
<td>180</td>
</tr>
<tr>
<td>TGTTTCCGGAAGTTGAGGGTAGGTAATAGTAGGGCGGAAGTTGCGGAGGGAATCGTGGGAGG</td>
<td>240</td>
</tr>
<tr>
<td>TGCACAAGAAAGCGCTGACAGAGATGGTGCAAAAGCCTATTCTGACTGGGCCCCAGA</td>
<td>300</td>
</tr>
<tr>
<td>CCCAATAACGGGTGACTACAGGCCCATCAACCACACCCCTGGAAATTGAGCGGAGCT</td>
<td>360</td>
</tr>
<tr>
<td>CCGACGGATGTTGCTTAAACCAAGTTGCAAATCGGCAACAAATAGACTTTTTAAAAACCATAAT</td>
<td>420</td>
</tr>
<tr>
<td>ATCTCCAATTTTCCACCTACGATTGGCTTGCTTGAAAGGGAAATCTTGTGTTTTCG</td>
<td>477</td>
</tr>
</tbody>
</table>

**Coding sequence of the VuLEA5-1 gene from cowpea IT96D-602**

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;VuLEA5-1</td>
<td>60</td>
</tr>
<tr>
<td>ATGTCTCTTCTCTCTCGCAAGCCAAATCTCTCGCTCCTCTCTCAGTCTATCTCA</td>
<td>60</td>
</tr>
<tr>
<td>CTAATTCTGTGTCATCGCCGAGTTGATGCGCTCGTGGTTTCCGGAAGGGTTGGA</td>
<td>120</td>
</tr>
<tr>
<td>TTGGGTAATATTGAGGGCGTAGGAGGATCGTTGGAGGTGAGATGGCGAGAAAGGCTGTA</td>
<td>180</td>
</tr>
<tr>
<td>ACAAGAGATGGTGCAAAAGCCTATTCTGACCTGGCCCCAGACCAATAACGGGTGACTAC</td>
<td>240</td>
</tr>
<tr>
<td>AGGCCCATCAACCCACCCCTGAAATTTGAGCGGAGCTCGACGGATGGTTGGTTAAC</td>
<td>300</td>
</tr>
<tr>
<td>CACAAGGTCAATCCGCCACAAATAG</td>
<td>324</td>
</tr>
</tbody>
</table>

**Amino acid sequence of the VuLEA5-1 protein**

<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;VuLEA5-1_aa</td>
<td>60</td>
</tr>
<tr>
<td>MSPSLSQAKSLRLLPQSLISLIPVHRROYAVASDVSVRVGLGNNVVRSGSVGAAEKPV</td>
<td>60</td>
</tr>
<tr>
<td>TRDGAKAYSDWAPDPITGDYRPINHTPEIDPVELRRMLLNHKFKSPQ-</td>
<td>107</td>
</tr>
</tbody>
</table>
Appendix 4
Plasmid maps and cloning strategy for transforming A. thaliana with a cowpea LEA5 gene

Schematic diagram of the LR Clonase recombination reaction in which the VuLEA5-1 coding region is transferred from the pCR8/GW/TOPO entry vector into the pMDC32 destination vector to create the pMDC32-VuLEA5-1 transformation vector.