

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Pflanzenzüchtung

**Analysis of differential gene expression under water-deficit stress
and genetic diversity in bambara groundnut
[*Vigna subterranea* (L.) Verdc.] using novel high-throughput
technologies**

Florian Stadler

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. A. Gierl

Prüfer der Dissertation: 1. Univ.-Prof. Dr. G. Wenzel
2. Priv.-Doz. Dr. V. Mohler

Die Dissertation wurde am 23.09.2009 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München am 23.12.2009 angenommen.

Teilergebnisse dieser Arbeit wurden vorab veröffentlicht:

Publikation:

Mayes, S.*, Basu, S.*, Murchie, E., Roberts, J.A., Azam-Ali, S.N., Stadler, F.*, Mohler, V., Wenzel, G., Massawe, F., Kilian, A., Bonin, A., Beena, A., Sheshshayee, M.S. (2009): BAMLINK – a Cross Disciplinary Programme to Enhance the Role of Bambara Groundnut (*Vigna subterranea* L. Verdc.) for Food Security in Africa and India. Acta Horticulturae 806, 137-150.

* Have contributed equally to the data presented in this paper.

Tagungsbeitrag:

Stadler, F., Mohler, V., Kilian, A., Bonin, A., Thümmel, F., Mayes, S., Ros, B., Wenzel, G. (2007): Applying biotechnology to underutilised crops: initial results from applying Diversity Arrays Technology (DArT) and Massively Parallel Signature Sequencing to bambara groundnut. New Approaches to Plant Breeding of Orphan Crops in Africa, 19-21 September 2007, Bern, Switzerland. Abstracts 66.

Table of contents

1. Introduction	1
1.1 Bambara groundnut: general information	1
1.2 Water-deficit stress	3
1.2.1 Background	3
1.2.2 The impact of water-deficit on plants	4
1.2.3 Plant adaptation to water-deficit	5
1.2.3.1 Late embryogenesis abundant proteins	6
1.2.3.2 Osmolytes and soluble sugars	6
1.2.3.3 Antioxidants	7
1.2.4 Water deficit stress sensing and signalling	8
1.2.5 Breeding for water-deficit tolerance	10
1.3 The BAMLINK project	11
1.3.1 Genetic diversity	12
1.3.2 Gene expression profiling	14
1.4 Objectives of the work	15
2. Materials and methods	17
2.1 DArT	17
2.1.1 Plant materials	17
2.1.2 DNA isolation	17
2.1.3 Marker discovery and scoring	18
2.1.4 Plasmid isolation and sequencing	20
2.1.5 Statistics and cluster analysis	21
2.2 Gene expression under water-deficit stress	21
2.2.1 Water-deficit stress experiments	21
2.2.1.1 Plant materials	21
2.2.1.2 Growing conditions	23
2.2.1.3 Experimental set-up and sampling	23
2.2.1.4 Physiological measurements	24
2.2.2 RNA isolation	25
2.2.3 cDNA libraries and high-throughput pyrosequencing	26
2.2.4 MPSS data analysis	27
2.2.5 Selection of sequences for microarray analysis	28
2.2.6 Oligonucleotide and microarray design	28
2.2.7 cDNA synthesis and hybridisation	29
2.2.8 Microarray scanning and data analysis	30
3. Results	31
3.1 Genetic diversity	31
3.1.1 Initial discovery array	31
3.1.2 Development of full-size array	33

3.1.3 Genotyping and cluster analysis.....	34
3.1.4 Large-scale genotyping	37
3.1.5 Intra-landrace diversity and ‘exotic’ germplasm	38
3.2 Gene expression under water-deficit stress.....	41
3.2.1 MPSS expression profiling.....	41
3.2.1.1 Data assembly and analysis.....	41
3.2.1.2 Global water-deficit response.....	43
3.2.1.3 Genotypic differences	47
3.2.2 Microarrays	51
3.2.2.1 Signal linearity	51
3.2.2.2 Validation of MPSS	51
3.2.3 Time course experiment.....	53
3.2.3.1 Physiological studies	53
3.2.3.2 Expression kinetics of selected genes	56
3.2.3.2.1 Hierarchical cluster analysis.....	56
3.2.3.2.2 k-Means cluster analysis	58
3.2.3.3 Within-time point analysis	63
3.2.3.3.1 Differential gene expression between treatments.....	63
3.2.3.3.2 Landrace-specific differences	70
4. Discussion	73
4.1 Genetic diversity	73
4.1.1 Experimental approach.....	73
4.1.2 Cluster analysis	77
4.1.3 Intra-landrace diversity and “exotic” germplasm.....	79
4.2 Gene expression under water-deficit-stress.....	81
4.2.1 MPSS expression profiling.....	81
4.2.2 Validation of MPSS through microarrays.....	83
4.2.3 Water-deficit stress reaction in bambara groundnut	85
4.2.3.1 Physiological changes	85
4.2.3.2 Differential gene expression under water-deficit stress	87
4.2.3.3 Differential gene expression between landraces	99
4.3 Conclusion and outlook.....	105
4.3.1 DArT markers and genetic diversity	105
4.3.2 Gene expression profiling	106
4.3.3 Subsumption of the work and future direction.....	108
5. Summary	111
6. Zusammenfassung.....	113
7. Literature cited	115
8. Appendix	135
Danksagung.....	175
Lebenslauf.....	177

List of abbreviations

ABA	Abscisic acid
ABRE	Abscisic acid-responsive element
AFLP	Amplified fragment length polymorphism
APX	Ascorbate peroxidase
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
bZIP	Basic region leucine zipper
CAD	Cinnamyl alcohol dehydrogenase
cDNA	Complementary deoxyribonucleic acid
CDPK	Calcium-dependent protein kinase
COMT	Caffeic acid methyltransferase
CTAB	Cetyl trimethylammonium bromide
DArT	Diversity Arrays Technology
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DRE	Dehydration-responsive element
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
GPX	Glutathione peroxidase
HD-ZIP	Homeodomain-leucine zipper
HSP	Heat shock protein
InDel	Insertion/deletion
LEA	Late embryogenesis abundant
LTP	Lipid transfer protein
MAPK	Mitogen activated protein kinase
MAS	Marker-assisted selection
MPSS	Massively parallel signature sequencing
mRNA	Messenger ribonucleic acid
NCED	9- <i>cis</i> -epoxycarotenoid dioxygenase
OD	Optical density
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PIC	Polymorphism information content
QTL	Quantitative trait locus
RuBisCO	Ribulose 1,5-bisphosphate carboxylase/oxygenase
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase
SSC	Standard saline citrate
SSR	Simple sequence repeat
TPM	Transcripts per million
U	Unit
UTR	Untranslated region

1. Introduction

1.1 Bambara groundnut: general information

Bambara groundnut [*Vigna subterranea* (L.) Verdc., syn. *Voandzeia subterranea* (L.) Thouars; Fig. 1], belongs to the *Fabaceae* (= *Leguminosae*) botanical family and, like nearly all legumes of agronomic importance, to the subfamily *Faboideae* (= *Papilionoideae*). The genus *Vigna* Savi, e.g. including cowpea [*V. unguiculata* (L.) Walp.] and mung bean [*V. radiata* (L.) R. Wilczek], and other well-known beans such as *Phaseolus* L. spp., soybean [*Glycine max* (L.) Merrill] and pigeonpea [*Cajanus cajan* (L.) Millsp.], are grouped in the tribe *Phaseoleae*.



Fig. 1. Bambara groundnut [*Vigna subterranea* (L.) Verdc.]. 1: habit of the flowering plant, 2: flower, 3: fruits, 4: seed. From van der Maesen & Somaatmadja (1989).

The plant is an annual herb with a reproductive cycle of usually 90 to 150 days. It develops a tap root with lateral roots in the lower part. Close to the soil surface, bambara groundnut forms creeping, much-branched, indeterminate lateral stems with erect trifoliolate leaves. Depending on the petiole/internode ratio, Doku (1969) distinguished three habit groups among cultivated material – bunch, semi-bunch and spreading. Around 40 days after sowing, the first (pale) yellow flowers open. The species is autogamous; however, cross-pollination has been observed and attributed to ants (Doku, 1968). Fruit set usually begins after a period under short day conditions although qualitative and quantitative differences exist (Linnemann *et al.*, 1995). Pods develop on lengthening peduncles on or beneath the soil surface and contain one or, less frequently, two seeds. A number of morphological differences led to the classification of two botanical varieties. *V.s.* var. *subterranea*, the domesticated form, is characterised by larger seeds and leaves, longer leaf petioles, shorter internodes, a thickened pod shell wall, and a more rapid and uniform germination compared to the wild var. *spontanea* (Harms) Hepper (Hepper, 1963; Pasquet *et al.*, 1999). Both wild and cultivated bambara groundnut have a chromosome number of $2n=2x=22$ (Frahm-Leliveld, 1953; Smartt, 1990).

Bambara groundnut is an indigenous African crop and cultivated in wide parts south of the Sahara at elevations up to 1600m. It is also found in parts of South and Central America, South and South-East Asia and Northern Australia (Kadam *et al.*, 1989; Linnemann & Azam-Ali, 1993). According to the FAO production statistics, 76,300 tonnes of “bambara beans” were produced worldwide in 2007 (FAO, 2009). However, these data are to be treated with care. Only four countries are listed (Burkina Faso, Cameroon, Democratic Republic of Congo, and Mali) and most figures are based on estimates. Significant statistics may indeed be hard to obtain as bambara groundnut is almost exclusively grown by small-scale farmers for subsistence. Only the surplus is sold on local markets. It is assumed that 100 million Africans regularly consume this crop (National Research Council, 2006) and for 1982, global production was estimated at 330,000 tonnes (Coudert, 1984) with Nigeria being the main producing country (100,000 tonnes) and considerable yields in Burkina Faso, Niger, Mozambique, and Ghana (Duke, 1981; Begemann, 1988; Linnemann & Azam-Ali, 1993). It is considered the third most important legume in Africa after cowpea and peanut (*Arachis hypogaea* L.) (Howell *et al.*, 1994). In more arid parts of sub-Saharan Africa like Namibia, it is second only to cowpea (Fleißner, 2006).

Bambara groundnut is primarily grown for its seeds which are a nutritious source of protein for human consumption. Like in other beans, seeds can be processed into versatile foods such as snacks, pastes, porridges, relishes, sauces, or vegetable milk, either solely or in combination with cereals (Kadam *et al.*, 1989; Obizoba & Egbuna, 1992; Brough *et al.*, 1993; Linnemann & Azam-Ali, 1993; Alobo, 1999; Amadi *et al.*, 1999; National Research Council 2006). The occasional use of the haulm as fodder for pigs and poultry has also been reported (Doku & Karikari, 1971).

At present, no improved bambara groundnut cultivars exist. Growers save their own seed for the next season or buy seed from the market, where usually seed mixtures are sold (Massawe *et al.*, 2005). The main criterion for distinguishing seed lots are testa colour and pattern. Plant morphology, however, does not show much apparent variability (Smartt, 1990). As there is no supra-regional market for bambara groundnut, it is unlikely that seeds from differing environments or production systems are mixed. Consequently, this form of unintentional selection leads to the evolution of populations containing a mixture of genotypes with a common appearance and a continuous adaptation to a specific environment. This meets the definition of an (autochthonous) landrace as proposed by Zeven (1998). In bambara groundnut, landraces are usually named after the site of cultivation or collection and seed colour.

Beside several national genebanks with local germplasm collections, the International Institute of Tropical Agriculture (IITA) holds the largest collection with currently 2030 accessions from 26 countries (as at August 2009).

1.2 Water-deficit stress

1.2.1 Background

The availability of adequate amounts of freshwater is an essential requirement for all forms of agriculture. Of all freshwater used by humans, 80-90% are allotted to this area, and crop production accounts for most of that (Savenije, 2000; Hamdy *et al.*, 2003). As such a large withdrawal has already contributed to major ecological impacts in many parts of the world, e.g. salinisation and desertification in Central and South Asia, Australia, parts of Central and South America, Australia, and the Sahel and Southern Africa, it is obvious that the further exploitation of freshwater resources is not feasible. Although, in the light of global climate change, total precipitation on earth is even expected to increase, the area under drought has been observed growing due to higher evaporative demands caused by rising temperatures. Regional droughts are becoming more frequent, prolonged and more intense. The most affected are regions where soil moisture is already limited, i.e. the tropics and subtropics (Solomon *et al.*, 2007). This dilemma is enforced by the fact that most developing countries are located in these areas. While there have been dramatic famines in the recent past, the situation will aggravate in the near future as these countries are characterised by a rapidly growing population.

From the 1960's on, the so-called green revolution boosted crop production by introducing improved crop varieties and agricultural practices and thus alleviated hunger and poverty by an estimated 6-8% in the Tricont (Asia, Africa, and Latin America) (Evenson & Gollin, 2003). However, it is widely acknowledged that conventional plant breeding in world crops like maize, wheat and rice is meeting its limits and will not lead to any further significant yield increases. The use of inorganic fertilisers and pesticides, too, is on the one hand more or

less exploited in the biological sense; on the other hand it is restricted to bigger agro-economic units and pushes the boundaries of small-scale farmers. Furthermore, these actions added to the problem of water scarcity. Therefore, increasing crop production and hence food security in a sustainable, water saving way will be one of the greatest challenges for mankind in the 21st century.

1.2.2 The impact of water-deficit on plants

Abiotic stress factors are estimated to account for losses of 51-82% of the potential yield in annual crops (Bray *et al.*, 2000). The most detrimental one is certainly soil water-deficit, particularly given that environmental stresses as high temperature, freezing and salinity are usually accompanied with or result in water deficit.

There is hardly a physiological process in plants which is not affected when the amount of water transpired exceeds the amount of water available (McKersie & Leshem, 1994). Alterations in the water balance are primarily manifested as a disturbance of photosynthesis. The first response to water-deficit is stomata closure to prevent further tissue dehydration which results in a limited carbon dioxide uptake. Consequently, the declining photosynthetic activity negatively affects further vegetative growth and the redirection of assimilates towards storage or reproductive organs. In the worst case, flowers and fruits may be shed. Another effect is that under more severe water-deficit, excess light energy cannot be sufficiently dissipated by carbon fixation in the Calvin cycle as the main electron sink. Instead, electrons are transferred to oxygen molecules which leads to the formation of reactive oxygen species (ROS). The primary intermediate of oxygen reduction, superoxide (O_2^-), is not highly reactive itself. However, it is subsequently dismutated to hydrogen peroxide (H_2O_2) and can, together with the latter and in the presence of metal ions (e.g. Fe^{2+} and Fe^{3+}), react to form hydroxyl radicals (OH^\cdot). In C_3 plants, H_2O_2 is also produced at high rates in the photorespiratory pathway when CO_2 is limited. Furthermore, electronic excitation of molecular oxygen may involve the formation of singlet oxygen (1O_2) (Bowler *et al.*, 1992; Smirnov, 1993; Asada, 1999, Noctor *et al.*, 2002). These ROS are described to play a role in degrading proteins, e.g. the D1 protein of photosystem II (Giardi *et al.*, 1996) and proteins involved in the Calvin cycle (Maroco *et al.*, 2002), damaging nucleic acids and, most frequently measured, lipid peroxidation (Smirnov, 1993). The breakdown of lipids leads to the impairment of cell membranes and thus the collapse of cellular compartmentation or even cell leakage.

Aside from biochemical aspects, membrane disruption can also occur through mechanical damage when, due to cellular water loss, the vacuole shrinks and the cytosol is subject to internal tension changes (Wilson *et al.*, 1987). In addition, water transport within the xylem can be considerably inhibited through cavitations and embolism (Choat *et al.*, 2003).

Although floral initiation may be promoted, another important effect of water-deficit is irreversibly reduced pollen viability (Turner, 1993) resulting in decreasing yields, especially

in combination with low atmospheric relative humidity and high temperature (Schoper *et al.*, 1987).

Most of these factors mentioned bear on morphological consequences. Growth of both aerial parts and roots may be suppressed. While the latter may restrict water uptake even more, reduced photosynthetically active tissue negatively affects yield. In case water availability falls below the permanent wilting point, the stomata of most mesophytes lose their ability to close under stress. Complete desiccation of tissues and, accordingly, death of leaf tips, whole leaves or the whole plant are the consequences (McKersie & Leshem, 1994).

1.2.3 Plant adaptation to water-deficit

As plants often face water-deficit during their life cycle, be it for a short time during the midday hours or for longer periods in dry seasons, it is clear that they have evolved manifold ways to cope with it. In general, it is possible to divide these strategies into three groups.

Plants may escape drought by completing their life cycle before water-deficit occurs. This involves a high degree of developmental plasticity and is of particular significance in environments with periodic rainfalls such as the semi-arid sub-tropics and savannahs. The water stored in the soil is used most efficiently through high rates of growth and gas exchange during the short period of available moisture. At the onset of drought, assimilates are shifted towards developing fruits, leading to successful reproduction before severe stress precludes further plant growth.

Another strategy is the avoidance of tissue dehydration. This is achieved by minimising water loss and/or maximising its uptake. While due to limited carbon resources, enhanced root growth is usually not possible, it has often been observed that the proportion of assimilates invested in the roots decreases less than in leaves and stems, resulting in an increased root/shoot dry matter weight ratio. Deep rooting capacity and fine root branches are a general feature of many dryland crops. Reductions in specific leaf area are not the only way to reduce transpiration. The above-named abortion of tissue may also be regarded as beneficial in this context. Shedding of older leaves allows the reallocation of nutrients to younger ones, stems, roots and fruits. Adaptation also becomes manifest in leaf morphology. A thick layer of cuticular wax may reduce leaf dehydration through non-stomatal water-loss and also decrease radiation load to leaf surfaces by enhanced light reflexion. Trichomes work in a similar way. Leaf wilting, curling, rolling and steepening leaf angles diminish exposure to sunlight and thus alleviate the precarious effects of excess radiation under water-deficit.

The third group is characterised by tolerance to low tissue water potential. Apart from structural adjustment of cells through more rigid cell walls or smaller cells, osmotic adjustment can play an important role here. The accumulation of ions (potassium, sodium and calcium) and compatible solutes in the cells lowers the osmotic potential and help the plant in maintaining water absorption and cell turgor under dehydration. Prominent osmotically active

compounds include proteins and amino acids, methylated quarternary ammonium compounds, hydrophilic proteins, carbohydrates and cyclitols. Furthermore, there is evidence showing osmoregulators being capable of stabilising enzymes.

As another tolerance strategy, plants have evolved effective mechanisms to detoxify reactive oxygen species. A number of antioxidant enzymes as well as non-enzymatic compounds are available (Levitt, 1980; Blum, 1996; Bohnert & Sheveleva, 1998; Reynolds *et al.*, 1999; Chaves *et al.*, 2003; Yokota *et al.*, 2006).

An overview of gene and gene product groups with significant accumulation under water-deficit conditions is presented in the following.

1.2.3.1 Late embryogenesis abundant proteins

The accumulation of non-storage proteins was first described in ripening cotton seeds (Dure *et al.*, 1981), with concentrations at up to 4% of cellular protein (Roberts *et al.*, 1993). Accordingly, these were termed late embryogenesis abundant (LEA) proteins. Their occurrence and abundance in other dehydrated tissues has been shown in many plants (Ingram & Bartels, 1996), but they have also been found in bacteria and lower animals (Stacy & Aalen, 1998; Gal *et al.*, 2004). Traditionally, based on their amino acid motifs, LEA proteins are divided into three major and two or three minor groups (Dure *et al.*, 1989; Bray, 1993; Ramanjulu & Bartels, 2002). However, both grouping and nomenclature are not consistent in the literature (Tunnacliffe & Wise, 2007). Recently, Hundertmark & Hinch (2008) dissected the 51 *Arabidopsis* LEA genes from the NCBI database into nine clusters.

LEA proteins have a biased amino acid composition conferring hydrophilicity and heat stability in solution (Tunnacliffe & Wise, 2007). Furthermore, they usually lack the amino acids cysteine and tryptophan (Bray, 1993). Despite their ubiquitous abundance in water-deficit stressed tissues, little is known about their functions. Dure (1993) proposed LEAs being capable of sequestering ions, possessing enhanced water binding capacity and functioning as chaperones, i.e. molecules that assist other proteins in maintaining or regaining their secondary structure. More recent data additionally suggest LEAs playing a role in the formation of cytoskeletal filaments, interacting with nucleic acids, scavenging ROS, and possibly regulating transcription or signalling (Wise & Tunnacliffe, 2004; Tunnacliffe & Wise, 2007).

1.2.3.2 Osmolytes and soluble sugars

Osmotic adjustment refers to the accumulation of compatible solutes in order to lower the cellular osmotic potential and thus maintain the driving gradient for water uptake under limiting conditions. Three major groups are described to act as compatible solutes: amino acids, quaternary amines and sugars or sugar alcohols.

The amino acid proline is often found to accumulate in dehydrated tissues. Gene expression studies revealed up-regulation of its two anabolic enzymes Δ^1 -pyrroline 5-carboxylate (P5C) synthase and P5C reductase and simultaneous down-regulation of its catabolism through proline dehydrogenase (Yoshida *et al.*, 1997). In addition to its role in osmoregulation, proline functions as a major structural component of plant cell walls (Nanjo *et al.*, 1999).

Glycine betaine is the most common example of a quaternary amine serving as a compatible solute. Two (mutually exclusive) ways of stabilising molecule structures and activities were proposed: either direct interaction with macromolecules or the formation of hydration shells around target complexes. However, there are species-dependent differences. While barley and spinach accumulate glycine betaine in high concentrations, *Arabidopsis* and tobacco do not synthesise this compound (Sakamoto & Murata, 2002).

The most effective osmoprotectant sugar is trehalose. However, in plants, sucrose appears to be the usual soluble sugar (Crowe *et al.*, 1992) although monosaccharides also are considered an important factor. This has been concluded from the coordinated induction of hydrolytic enzymes such as amylases and invertases under water-deficit (Keller & Ludlow, 1993; Pinheiro *et al.*, 2001). In the face of reduced carbon assimilation, concentrations of soluble sugars seem to be relatively constant, whereas starch contents decline (Chaves, 1991). Koster (1991) suggested glass formation being a possible way for sugars protecting cellular structures. Liquids become supersaturated through the presence of sugars and enter the state of plastic solids rather than solutes crystallising and disrupting membranes. Sugars have also been shown to directly protect membranes and proteins *in vitro*, possibly by replacing water molecules and altering physical properties through the formation of hydrogen bonds (Crowe *et al.*, 1992).

Although osmotic adjustment is considered one of the crucial processes in plant adaptation to drought, the accumulation of compatible solutes often is not sufficient to significantly decrease the osmotic potential (Ramanjulu & Bartels, 2002), at least until severe desiccation occurs (Chaves *et al.*, 2003). Therefore, osmolytes may also be involved in other protective mechanisms like scavenging ROS (Zhu, 2001).

1.2.3.3 Antioxidants

As outlined above, the increased formation of ROS is one of the main deleterious consequences of limited water supply. However, ROS are also an inevitable by-product of life for any aerobic organism. Accordingly, mechanisms to detoxify ROS exist in all plants (Bohnert & Sheveleva, 1998). While these usually suffice under normal conditions, the capacity of the antioxidant system is a critical factor for plant performance under stress. As no scavengers of hydroxyl radicals are known, the only strategy is to avoid its generation through inhibiting precursor reactions involving O_2^- and H_2O_2 (Apel & Hirt, 2004). A number of enzymatic and non-enzymatic pathways are available.

Superoxide dismutase (SOD) enzymatically converts O_2^- to H_2O_2 . Three isoforms are known and classified according to their metal cofactors. Copper/zinc-SOD is found in the cytosol and plastids; manganese-SOD is present in mitochondria and iron-SOD in plastids (Bowler *et al.*, 1992; Bohnert & Sheveleva, 1998). H_2O_2 can subsequently be detoxified by various pathways. In the ascorbate-glutathione cycle, H_2O_2 is reduced into H_2O by oxidising ascorbate which is catalysed by ascorbate peroxidase (APX). Ascorbate is regenerated either directly via monodehydroascorbate reductase or via oxidation of glutathione, which is regenerated using glutathione reductase. Both reactions require NAD(P)H as reduction equivalent and thus consume energy. The glutathione peroxidase cycle works in a similar way, but its catalysing enzyme glutathione peroxidase (GPX) uses glutathione directly as the reducing equivalent (Apel & Hirt, 2004). Another H_2O_2 scavenging enzyme is catalase ($2 H_2O_2 \rightarrow O_2 + 2 H_2O$), which is located in peroxysomes. It does not consume reducing power and shows a high reaction rate, but has poor affinity for H_2O_2 (Willekens *et al.*, 1997). In addition to the cellular redox buffers ascorbate and glutathione, which can also directly scavenge ROS without APX and GPX (Noctor & Foyer, 1998), many non-enzymatic antioxidants were described. These include isoprenoids, such as the carotenoids β -carotene and zeaxanthin, tocopherols or carnolic acid (Havaux, 1998; Demmig-Adams & Adams, 2002; Munné-Bosch & Allegre, 2003) and phenylpropanoids, such as hydroxycinnamic acids, flavonols and anthocyanins (Chalker-Scott, 1999; Close & McArthur, 2002; Tattini *et al.*, 2004), which, in contrast to the enzymatic detoxification systems, are also capable to quench singlet oxygen (Smirnoff *et al.*, 1993).

1.2.4 Water deficit stress sensing and signalling

The first step in generating a biochemical response to water-deficit is the recognition of a stimulus at the cellular level. It is still unclear what aspect of water loss is actually perceived. The decrease or loss of turgor itself or its effects on cell wall-plasma membrane interactions or the change in the osmotic potential across the plasma membrane may come into consideration to be the trigger of the stress response (Bray, 1997; Shinozaki & Yamaguchi-Shinozaki, 1997). The hybrid-type histidine kinase ATHK1 from *Arabidopsis thaliana*, a transmembrane protein with two hydrophobic regions, was described as being transcriptionally upregulated in roots as a response of external osmotic changes and displaying functional similarity to osmosensors in yeast (Urao *et al.*, 1999).

After stress sensing, the signal is mediated through a signal transduction cascade involving several protein phosphorylation and dephosphorylation events. A number of Ca^{2+} dependent (CDPK) and mitogen activated (MAPK) protein kinases, and kinases that in turn phosphorylate MAPKs, have been reported relaying the dehydration signals from the plasma membrane to the nucleus (Jonak *et al.*, 1999; Sanders *et al.*, 1999; Ramanjulu & Bartels,

2002). An important role in signal transduction has been attributed to elevated levels of cytosolic Ca^{2+} (Sanders *et al.*, 1999; Knight & Knight, 2001).

These early responses induce the biosynthesis of abscisic acid (ABA) (Bray, 2002), with the key regulatory step being catalysed by 9-*cis*-epoxycarotenoid dioxygenase (Qin & Zeevaart, 1999). ABA is well-known to induce *de novo* expression of both structural and functional genes under water-deficit stress. Shinozaki & Yamaguchi-Shinozaki (1997) proposed the existence of two ABA-dependent pathways. One leads to the expression of genes that do not require protein biosynthesis for their expression. These possess abscisic acid response elements (ABREs), which have a core ACGT-containing G-box and are bound by bZIP transcription factors (Chaves *et al.*, 2003). The second ABA dependent pathway contains genes that do depend on protein synthesis for their expression. MYB and MYC transcription factors fall into this group, but there are also bZIP proteins. Due to the additional transcriptional regulation, genes mediated by this pathway are assumed to react rather slowly to water-deficit conditions. However, the existence of water-deficit-inducible genes that do not require ABA has also been shown. Those genes carry a conserved dehydration responsive element (DRE) in their promoter regions, which does not function as an ABRE, and are inducible by exogenous ABA and cold. Moreover, a class of water-deficit-inducible genes do not respond to ABA and cold treatment (Shinozaki & Yamaguchi-Shinozaki, 1997).

The signal in this transduction cascade is enhanced by several second messengers. Phospholipase D, activated by Ca^{2+} , catalyses the synthesis of phosphatidic acid (PA) which in turn activates phospholipase C. The latter hydrolyses phosphatidylinositol 4,5-bisphosphate into inositol 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG). IP_3 releases Ca^{2+} from intracellular stores in the cytoplasm. DAG is phosphorylated to PA by DAG kinase (Meijer & Munnik, 2003). ROS can also act as second stress messengers, either through the induction of MAPK cascades, by oxidising components of the signalling pathways or by directly regulating the activity of transcription factors (Kovtun *et al.*, 2000; Apel & Hirt, 2004). In addition, sugars are attributed a role in plant stress signalling and coregulating ABA- and stress-inducible genes (Rolland *et al.*, 2002).

ABA is not the only plant hormone controlling dehydration stress-induced gene expression. Senescence-related ethylene triggers another signalling pathway and affects growth by interacting with ABA (Morgan & Drew, 1997; Sharp & LeNoble, 2002).

It is difficult to describe an integrated model of signal transduction and gene regulation pathways under water-deficit. The complex network of stress responses not only regulates itself through the multiple functions of its components (e.g. Ca^{2+} , hormones, phospholipids, ROS, sugars), but pathways also converge at certain junctions (Knight & Knight, 2001). A prominent example is the *A. thaliana rd29A* gene which contains both a DRE and an ABRE motif (Narusaka *et al.*, 2003) and thus canalises branches. Furthermore, uncoupling water-deficit stress from other stresses does not represent natural situations. Different types of stress may lead to responses with overlapping pathways that interact with each other. The

expressions of *CBF* (also known as *DREB1*) and *DREB2* transcription factors are induced by cold and dehydration, respectively, but share a DRE binding site (Liu *et al.*, 1998).

1.2.5 Breeding for water-deficit tolerance

The above sections show that water-deficit stress is a multidimensional stress. It is obvious that resistance strategies are not mutually exclusive. Instead, evolution has directed plants to find individual ways for making the most out of the dilemma of growth or protection. However, in an agricultural context, mere plant survival, as it is typical for desert succulents or resurrection plants, does not meet the demands of appropriate food production in drought-prone environments. Therefore, water-deficit tolerance has to be defined in terms of yield in relation to a limited water supply (Passioura, 1996).

Although the problem of more frequent droughts has been common for several decades, traditional plant breeding for water-deficit tolerance has been rather ineffective. Reasons for this are to be found in the complexity of the stress itself, its unpredictability and its interaction with other abiotic and biotic stresses. Furthermore, breeding approaches have missed focussing on target environments, leading to the release of cultivars which are superior under favourable conditions but are not adopted by farmers in drought-prone areas, where agriculture is characterised by low inputs in irrigation, fertilisation and crop management (Ceccarelli & Grando, 1996). Hence, a gap between yield potential under optimal condition and actual yields under stress arises and yield stability for differently challenging environments often is not granted (Cattivelli *et al.*, 2008).

Nevertheless, most of the progress made in improving water-deficit tolerance is accredited to conventional breeding. Accelerating the creation of tolerant cultivars through molecular concepts inevitably involves quantitative trait loci (QTLs) due to the polygenic nature of the trait (Reynolds & Tuberosa, 2008). However, the practical application of marker-assisted selection (MAS) for QTLs conferring water-deficit tolerance also bears difficulties. The high variability in stress types (timing, duration and intensity) together with other environmental factors, and the plethora of genes interacting with each other lead to QTLs of low heritability, which may not be valid when detached from the genomic background of the mapping population (epistasis) (Francia *et al.*, 2005). Association mapping may be helpful to overcome this problem as it incorporates thousands of recombination and selection events (Syvänen, 2005) in contrast to segregation mapping. The risk of high genotype-by-environment interaction implies that only major QTLs can be mapped with enough precision (Witcombe *et al.*, 2008). Furthermore, QTL confidence intervals can span several hundred genes, hampering the linkage of molecular markers to functional genes.

Using structural genomics associated with trait-based approaches requires detailed knowledge of the physiological and molecular basis of water-deficit tolerance in order to dissect candidate traits. Functional genomics is therefore most suitable to complement forward

genetics (trait to gene). Such a ‘bottom-up’ approach, i.e. from the gene to the phenotype, allows the direct discovery of genes of significance minimising linkage drags between markers and QTL component genes, can provide information about the traits underlying tolerance and create the basis for choosing genes or gene combinations for genetic engineering (Witcombe *et al.*, 2008).

1.3 The BAMLINK project

Out of around 7,000 cultivated edible plant species, only 30 are used to meet 95% of the world’s food energy needs (FAO, 1997). Three cereals, wheat, rice and maize, alone account for more than half of the global plant-derived energy intake. Thus, food security stands on shaky grounds. None of the three crops have their centres of diversity in Africa, which means that any breeding effort may be limited by the non-existence of material adapted to the resource-poor and climatically vulnerable regions of sub-Saharan Africa. Against this backdrop, BAMLINK, a European Union Framework 6 project was launched in 2006, which aims to promote the use of indigenous, under-utilised crops for food security in semi-arid environments. Hammer *et al.* (2001) distinguished between under-utilised crops, which were formerly widely grown and consumed and have fallen or are falling into disuse and neglected crops, which have been ignored by science and development but are still being used in areas where they are well adapted and competitive. In parallel, the less well defined term ‘orphan crops’ is often found in the literature. Bambara groundnut was chosen as a case study for this project and meets the criteria of both under-utilised and neglected crops. Being displaced by the South American peanut, which is similar in habit but different in terms of use and climatic adaptation, no supra-regional markets exist, not to mention improved cultivars. At the moment, there is no research mandate for a Consultative Group on International Agricultural Research (CGIAR) centre. Further reasons limiting the use of bambara groundnut are low and/or unpredictable yields, the long time needed for cooking and processing and stigmatisation as a ‘poor people’s food’ or ‘women’s crop’ (Brough *et al.*, 1993; Mayes *et al.*, 2009). Nevertheless, it bears key features that make it an appropriate crop for developing countries. On the socio-economic side, bambara groundnut has the potential to command a high market price (Coudert, 1984). Furthermore, the seeds have a well balanced nutrient composition. The reported approximate chemical composition is ash 3-5%, fat 6-8%, carbohydrates 53-65% and crude protein 17-21% (Enwere & Hung, 1996; Amarteifio & Moholo, 1998; Onimawo *et al.*, 1998). For six of eight essential amino acids, bambara groundnut scored at or above the WHO reference protein (FAO/WHO, 1973) and was thus among the three plants with highest protein quality in a survey of 24 indigenous plant species of Burkina Faso (Glew *et al.*, 1997). Other authors additionally reported high lysine contents (Nwokolo, 1987; Brough & Azam-Ali, 1992), the most limiting essential amino acid in cereals (Nelson, 1969). Hence, bambara groundnut may serve as an ideal supplement to a cereal-based diet. Its agronomic key traits are symbiotic fixation of atmospheric nitrogen

(Linnemann, 1991), as is common to most legumes, and the potential to produce significant yields under conditions of soil moisture stress where other crops fail (Linnemann & Azam-Ali, 1993; Collinson *et al.*, 1996).

Underpinning a multi-disciplinary, international effort linking agronomic, nutritional and socio-economic aspects is a genetic analysis of bambara groundnut. Within the BAMLINK programme, the focus of the work presented was to exploit the availability of novel high-throughput technologies in order to create molecular information in a rapid and cost-efficient manner.

1.3.1 Genetic diversity

Diversity studies on bambara groundnut were previously carried out using morphological (Schenkel *et al.*, 2002; Ntundu *et al.*, 2006) and biochemical markers (Odeigah & Osanyinpeju, 1998; Pasquet *et al.*, 1999). However, both approaches are capable of detecting only a limited degree of variation (Orozco-Castillo *et al.*, 1994; Johns *et al.*, 1997) and thus provide little insight into the true structure of populations. Furthermore, morphological traits are often subject to environmental influences, which may result in low stability of markers across environments (Alamerew *et al.*, 2004).

These limitations can be overcome by the use of DNA-based molecular markers. However, according to its status as an under-utilised crop, no *ex ante* sequence information exists for bambara groundnut. This complicates the application of genetic marker techniques such as simple sequence repeats (SSRs) or single nucleotide polymorphisms (SNPs), which require laborious and costly preliminary work. Two sequence-independent genetic marker analysis methods, random amplified polymorphic DNA (RAPD; Amadou *et al.*, 2001; Massawe *et al.*, 2003) and amplified fragment length polymorphism (AFLP; Massawe *et al.*, 2002; Singrün & Schenkel, 2003; Ntundu *et al.*, 2004), were successfully implemented in bambara groundnut. Yet these techniques suffer from various constraints, too. Relying on size-separation of DNA fragments using gel electrophoresis, difficulties may arise in accurately determining fragments lengths. Moreover, bands of identical sizes do not necessarily represent the same allele at the same locus (Huttner *et al.*, 2005). Thirdly, throughput is limited for gel-based systems and usually several experiments are required to obtain a full dataset, which bears the risk of scoring experimental variation, even more so, if analyses are conducted in different laboratories.

To deal with these difficulties, Diversity Arrays Technology (DArT) was developed and first published for rice by Jaccoud *et al.* in 2001. Since then, this method has been applied to more than 50 organisms, including mostly major and minor crops, but also animals and microbes (www.diversityarrays.com; as at July 2009). The principle of DArT is illustrated in Figure 2. The first step in DArT involves assembling a group of DNA samples representative of the germplasm to be analysed, further referred to as diversity panel. Pooled samples are then

subjected to a complexity reduction method, i.e. a process which reproducibly selects a defined fraction of genomic fragments (genomic representation). While a number of complexity reduction methods are conceivable, the currently preferred system relies on restriction enzyme digestion, adapter ligation and selective amplification of adapter-ligated fragments. Usually, a combination of a frequently cutting restriction endonuclease (4bp recognition site) and a rare cutter, in most cases *Pst*I (6bp recognition site), are chosen. Thereafter, adapters are ligated to *Pst*I ends and fragments are PCR-amplified using primers complementary to the adapters. Thus, only fragments carrying *Pst*I overhangs are retained and form a genomic representation. These fragments are used to construct an *E. coli* marker discovery library. Individual clones are picked, inserts are amplified and spotted onto glass slides as molecular probes (=discovery array).

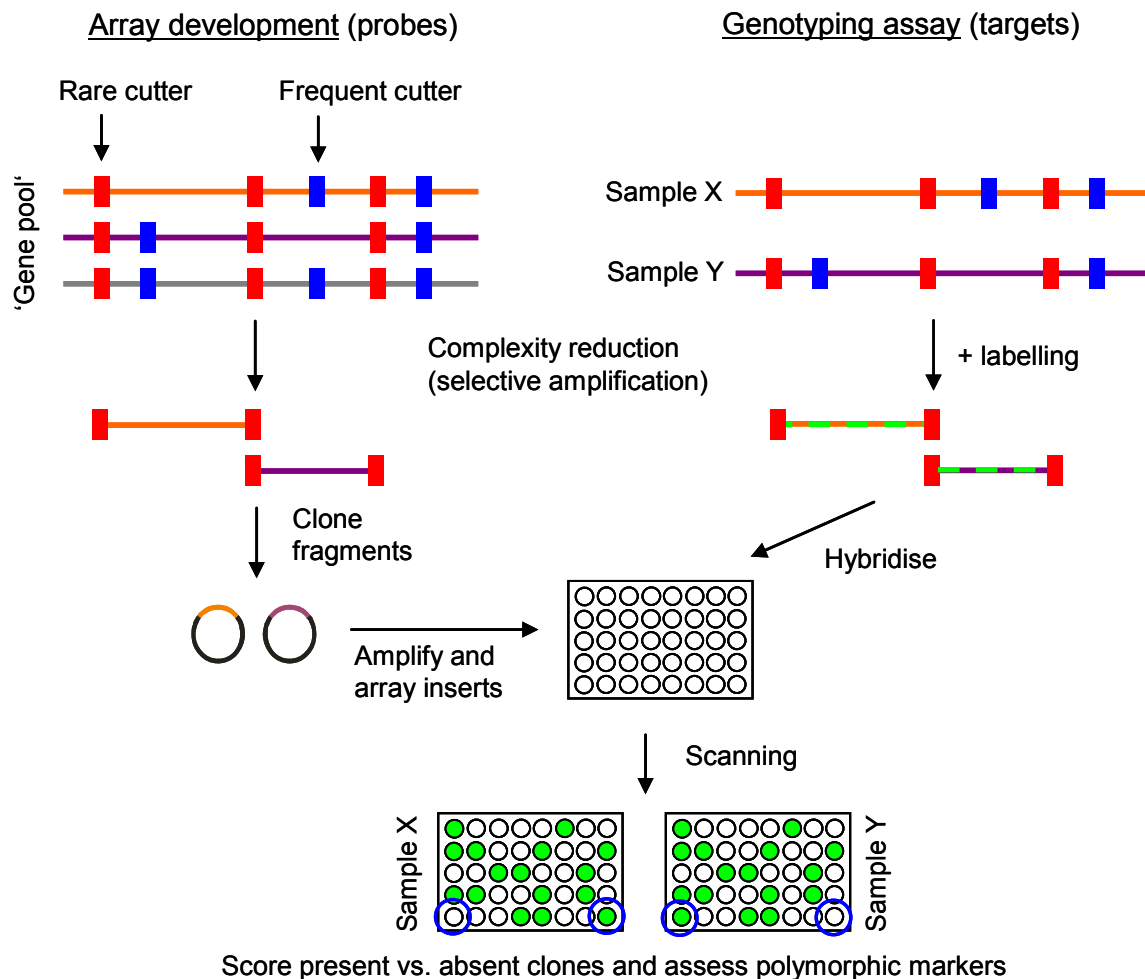


Fig. 2. Simplified scheme of DArT array development and genotyping (after Jaccoud *et al.*, 2001 and Kilian *et al.*, 2005). Explanations are given in the text.

In the same way, genomic representations are prepared from individual genotypes and labelled using a fluorescent dye. These targets are then hybridised to the discovery array, scanned and scored as present or absent using specifically designed software tools. By comparing hybridisation profiles from different individual genomes, clones are identified as polymorphic markers if hybridisation differences are found between genotypes. Thus, DArT delivers biallelic markers behaving in a dominant (present vs. absent) way. The microarray format allows for typing of thousands of loci in parallel, which, compared to other genetic marker systems, significantly reduces the costs per data point once the platform is developed (Wenzl *et al.*, 2004; Huttner *et al.*, 2005; Kilian *et al.*, 2005).

1.3.2 Gene expression profiling

So far, no gene expression studies were conducted in bambara groundnut. Similarly to diversity analysis, a technology platform is needed that allows for generation of information specific for bambara groundnut. However, most tools for gene expression analysis rely on the availability of molecular probes, such as Northern blot hybridisation (Alwine *et al.*, 1977), *in situ* hybridisation (Lawrence & Singer, 1985), real-time quantitative PCR (Heid *et al.*, 1996) and cDNA array technologies (Schena *et al.*, 1995).

The methods mentioned above are based on quantitatively measuring the intensity of fluorescently or radioactively labelled hybridised nucleic acids. The concept of gene expression analysis by massively parallel signature sequencing (MPSS) was introduced around the turn of the millenium (Brenner *et al.*, 2000; Reinartz *et al.*, 2002) and pursues a different strategy. In brief, mRNA is converted to cDNA and individual strands are ligated to microbeads. These are then arrayed in a plate format and sequenced in a parallelised assay. Counting the number of transcripts finally provides a digital measure of gene expression. In its principle, this approach resembles expressed sequence tags (EST) sequencing projects (Adams *et al.*, 1991). Due to the absence of cloning into bacterial vectors, physical separation of clones and individual template processing, MPSS achieves a far greater throughput than conventional Sanger sequencing with respect to sequence tag abundance. However, the initial technology produced signatures of only 16 to 20 nucleotides per cDNA strand (Brenner *et al.*, 2000) compared to typically 300 to 400 nucleotides for ESTs submitted to the National Center for Biotechnology Information (NCBI) EST database (Boguski *et al.*, 1993). The recent launch of so-called ‘next-generation sequencing’ platforms has addressed to this limitation. The first of these technologies to reach the market was the ‘454 Sequencer’ in 2005, developed by 454 Life Sciences, Branford, and acquired by Roche, Basel, under the name of Genome Sequencer™ 20 (Rothberg & Leamon, 2008). The introductory paper by Margulies *et al.* (2005) described little more than 300,000 high-quality reads at 110 bases average length obtained in a single four-hour run. Since then, a number of plant transcriptome studies have been published, including the assembly of full-length cDNA in *Medicago truncatula* (Cheung

et al., 2006), *Arabidopsis thaliana* (Weber *et al.*, 2007), maize (Emrich *et al.*, 2007) and pea (Bräutigam *et al.*, 2008). In contrast to the traditional chain-termination method, the 454 technology utilises a sequencing-by-synthesis approach. In brief, nucleotide species are individually flowed over plates containing clonally amplified cDNA strands. After adding the substrates luciferin and adenosine 5'-phosphosulfate, pyrophosphate is released and a light signal is generated each time a nucleotide is incorporated into the complementary strand (pyrosequencing). This signal is then recorded by a highly sensitive camera and automatically processed into coherent nucleotide sequences (Margulies *et al.*, 2005). More details are given in the materials and methods section.

In terms of read length, the 454 platform outperforms other 'next-generation sequencing' technologies such as Illumina and SOLiD, whereas the latter deliver more reads (Mardis, 2008). For counting-based applications, greater abundance of sequence tags allows for increased depth of analysis. However, for the approach presented for bambara groundnut, only the 454 technology provides sufficient read length. As described for maize by Eveland *et al.* (2008), cDNA populations were digested using a 4bp restriction enzyme and the fragments containing a polyA-tail were isolated for sequencing. Thus, the vast majority of sequence tags originate from the 3'-untranslated regions (3'-UTR) of the transcripts. As this region is highly specific for individual species and no bambara groundnut sequence is deposited in public databases, homology search in related species may be challenging with shorter reads. Secondly, it was intended to design 50-mer oligonucleotide probes from the MPSS-derived sequence tags for validation of the MPSS data and analysis of a further experiment. Read lengths between 30 and 40 bases, as achieved through the Illumina and SOLiD platforms (Mardis, 2008), would be too short for such a purpose.

1.4 Objectives of the work

As part of the BAMLINK project, the objective of this work was to create fundamental genetic information for bambara groundnut using a cost- and time-efficient methodology. In detail, the goals were divided into two parts.

The first part dealt with genetic diversity in bambara groundnut. It was intended a) to develop a DArT array containing at least 300 polymorphic markers for whole-genome profiling and future mapping purposes, b) to genotype a significant proportion (around 20%), ideally representative of geographic distribution and genetic and morphological diversity, of bambara groundnut accessions held at the IITA and landrace individuals used by project partners, c) to estimate genetic diversity within the cultivated subspecies and supply information about its population genetic structure and d) to genotype individual genotypes of the six core landraces to gain insight into intra-landrace variation. Due to limited technical equipment, laboratory work on array development and genotyping was carried out as a sub-contract to Diversity Arrays Technology Pty Ltd (Yarralumla, Australia).

The second part focussed on the molecular genetic investigation of the water-deficit stress response in bambara groundnut. The goals were e) to establish the MPSS technology coupled with 454 pyrosequencing in an under-utilised crop, f) to generate ESTs for bambara groundnut, g) to extract expression profiles for genes under water-deficit stress using differently adapted genotypes in a fully controlled environment, h) to attempt to integrate these profiles into the complex regulatory network, i) to validate MPSS-derived data by means of a small custom-made oligonucleotide microarray, j) to investigate the behaviour of a subset of genes in a time series experiment representing a more moderate degree of drought, k) to identify candidate genes potentially explaining different degrees of drought tolerance in a pair of contrasting landraces and l) to support these molecular data by measurements at the physiological level.

2. Materials and methods

2.1 DArT

2.1.1 Plant materials

Thirty-eight bambara groundnut genotypes from 14 countries were chosen to form the genetic base for the construction of a DArT marker discovery array. Selection of this diversity panel was based on the dendrogram from Singrün & Schenkel (2003), who identified 17 clusters of genetic similarity among 223 landraces and IITA accessions using ten AFLP primer combinations for the enzyme system *EcoRI/MseI* and one SSR marker. Two preferably distinct accessions from every cluster and four landraces were chosen in order to maximise the coverage of genetic diversity. Single seeds were placed between two sheets of moistened filter paper in a home-made germination device.

For array expansion and genotyping, 94 genotypes were utilised. Most of the accessions from the initial diversity panel were included again. Furthermore, the panel was complemented with landraces from the above-mentioned study, landraces obtained from national African germplasm collections and additional IITA accessions from countries not included in the primary set of genotypes. Seeds were sown in 2.5l pots and cultivated in a semi-controlled greenhouse cabin at 28°C/23°C day/night temperature with natural daylength and supplementary lighting until maturity. Accessions and landraces are listed in Table 15 (Appendix). For large-scale genotyping, 342 additional genotypes were raised under the same conditions as for array expansion, with the exception of smaller pots (11 cm diameter).

2.1.2 DNA isolation

Three young leaflets (ca. 0.7g) per genotype were harvested and stored at -20°C. Extraction of total genomic DNA was carried out following the CTAB-based method after Saghai-Marooif *et al.* (1984). In brief, plant material was finely ground under liquid nitrogen using mortar and pestle and transferred to 50ml reaction tubes containing 10ml 1.5x CTAB solution (150mM Tris-HCl pH 7.5, 1.05M NaCl, 15mM EDTA pH 8.0, 1.5% (w/v) CTAB, and 1.5% (v/v) β -mercaptoethanol). Samples were briefly vortexed and incubated for at least one hour at 65°C in a shaking water bath, followed by cooling on ice for five minutes. The suspension was extracted twice by adding 15ml chloroform/isoamyl alcohol (24:1), overhead shaking for 20 minutes and centrifugation for 30 minutes at 2,100g and room temperature. The supernatant (aqueous phase) was then carefully transferred to a fresh reaction tube and RNA digested using 15 μ l RNase A (10mg*ml⁻¹; Qiagen GmbH, Hilden, Germany) for around one hour at room temperature. DNA was precipitated by adding 15ml isopropanol (-20°C), inverting and centrifugation for 15 minutes at 2,100g and 4°C. The pellet was transferred to a 1.5ml

reaction tube filled with 1ml 70% ethanol and washed overnight at 4°C. After brief centrifugation at maximum speed (13,000g), a second washing step was carried out for one hour. Pellets were centrifuged again and the supernatant was decanted. The air-dried pellets were finally resuspended in 50-300µl 1x TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0), depending on pellet size. Concentrations were estimated on ethidium bromide-stained 0.8% agarose gels by visual comparison with bands of known concentration from *HindIII*-digested lambda DNA (Fermentas GmbH, St. Leon-Rot, Germany) and adjusted to 100ng*µl⁻¹.

2.1.3 Marker discovery and scoring

The procedure of generating DArT markers, screening for polymorphisms and genotyping was conducted by Diversity Arrays Pty. Ltd., Yarralumla, Australia, essentially following the methods described in Jaccoud *et al.* (2001) and Yang *et al.* (2006).

In order to find a suitable complexity reduction method, individual genomic DNA samples were treated with a combination of two restriction endonucleases. *PstI* was always used as the rare cutter (restriction site 6bp), while eight enzymes (*AluI*, *BanII*, *BsoBI*, *BstNI*, *MseI*, *RsaI*, *TaqI*, and *Tsp509I*; all enzymes from New England Biolabs Ltd., Pickering, Canada) with a 4bp recognition site were tested as the frequent cutter. Digestion and *PstI* adapter (5'-CAC GAT GGA TCC AGT GCA-3', annealed with 5'-CTG GAT CCA TCG TGC A-3') ligation with T4 DNA ligase (New England Biolabs Ltd.) were carried out in one step. Fragments carrying the *PstI* adapter at both ends were PCR amplified using the primer 5'-GAT GGA TCC AGT GCA G-3', REDTaq® polymerase (Sigma-Aldrich Pty. Ltd., Sydney, Australia) and the following PCR programme: 94°C denaturation for one minute, 30 amplification cycles of 94°C for 20 seconds, 58°C for 40 seconds, 72°C for one minute, and a final extension step at 72°C for seven minutes. Satisfactory results were obtained for the enzyme combinations *PstI/AluI* and *PstI/BanII*, which was visualised on an agarose gel by intense and homogeneous smears without the amplification of individual bands. The *PstI/AluI* method produced slightly shorter fragments and was therefore chosen for creating the initial library for DArT marker discovery.

The PCR amplicons from the 38 samples in the diversity panel were pooled and ligated into the pCR2.1-TOPO® vector using the TOPO cloning kit and transformed into electroporation competent TOP10F' (Invitrogen Pty. Ltd., Mount Waverly, Australia) *E. coli* cells according to the manufacturer's instructions. Blue/white screening for successful transformants was done on medium containing ampicillin and X-gal. A total of 1,536 individual white colonies were picked and inserts were amplified using M13 primers (forward: 5'-ACG ACG TTG TAA AAC GAC GGC CAG-3', reverse: 5'-TTC ACA CAG GAA ACA GCT ATG ACC-3'), REDTaq polymerase and the following PCR programme: 95°C for five minutes and 35 cycles of 94°C for 30 seconds, 52°C for 30 seconds, 72°C for one minute. The amplified inserts were precipitated with one volume of isopropanol, washed with 70% ethanol and air-dried.

The purified DNA fragments were resuspended in spotting buffer (1M sucrose, 50% DMSO) and spotted in triplicates onto polylysine-coated slides using a 16-pin MicroGrid II automated microarrayer (Genomic Solutions Inc., Ann Arbor, USA). DNA was immobilised to the slide surface by baking at 80°C for two hours, followed by denaturation in 92°C hot deionised water for two minutes and drying by centrifugation. These fragments served as molecular probes for the subsequent hybridisation experiment.

In order to test the performance of this array and screen for polymorphisms, targets complementary to the probes were produced for a subset of 32 DNA samples following the same method as for probe preparation. Fragments were fluorescently labelled using Cy3-labelled random decamers (Sigma-Aldrich Pty. Ltd., Sydney, Australia) and the *exo*-Klenow fragment of *E. coli* DNA polymerase I (New England Biolabs Ltd.) following the manufacturer's instructions. Targets were mixed with ExpressHyb hybridisation buffer (Clontech Laboratories Inc., Mountain View, USA) and FAM-labelled polylinker fragment of the pCR2.1-TOPO vector as a reference and hybridised to the slides overnight in a humidified hybridisation chamber at 65°C. Washing was done in three steps using SSC buffers at different concentrations, followed by spin-drying.

Processed slides were scanned using a LS300™ microarray scanner (Tecan Group Ltd., Männedorf, Switzerland) and target and reference images stored as TIFF files. These were automatically analysed by means of the DArTsoft software, which localised spots, computed and normalised hybridisation intensities [$\log(\text{Cy3-target}/\text{FAM-reference})$], calculated the median value for replicate spots and identified polymorphic clones by using a combination of ANOVA and fuzzy K-means clustering. Finally, a score of '1' or '0' (= present vs. absent) was assigned to each marker for each genomic representation (sample).

For DArT array expansion and genotyping, 1,152 clones from the initial *PstI/AluI* library were utilised again. A new *PstI/AluI* library was produced as described above using the panel of 94 genotypes, with the exception of adding *BglIII* as an additional restriction endonuclease in the process of complexity reduction, and 4,992 colonies were picked to amplify fragments. Moreover, a second complexity reduction method was applied replacing the frequently cutting enzyme *AluI* by *TaqI*. From this library, 1,536 *PstI/TaqI* clones were combined with the clones from the two *PstI/AluI* libraries and assembled into a full-size genotyping array containing 7,680 clones. Target preparation for 94 DNA samples and hybridisation were conducted as described above, using the *PstI/AluI* and *PstI/TaqI* enzyme combinations in two separate experiments.

2.1.4 Plasmid isolation and sequencing

Four 384-well plates with the *PstI/AluI* clones in freezing medium from the initial discovery array were received from Diversity Arrays Pty. Ltd.. Colonies were transferred into cell culture tubes filled with 1ml LB_{Amp} medium (10g*l⁻¹ peptone, 5g*l⁻¹ yeast extract, 0.6g*l⁻¹ NaCl, 100mg*l⁻¹ ampicillin, pH 7.0) and grown overnight at 37°C in an Unimax 2010 orbital shaker (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany) at 240rpm. Plasmids were isolated using the High Pure Plasmid Isolation Kit (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturer's instructions. PCR amplification of cloned inserts was conducted with the primer pair pBL2SK flanking the multiple cloning site of the vector (forward: 5'-GAC TGG AAA GCG GGC AGT GAG-3', reverse: 5'-TGC TGC AAG GCG ATT AAG TTG-3') and the following reaction assay: 1.5µl cell culture, 3.0µl 10x PCR buffer (Qiagen), 0.5µl of each primer (10mM), 0.25µl dNTP mix (10mM), 0.05µl Taq DNA polymerase (5U*µl⁻¹; Qiagen) and 24.2µl H₂O_{bidest.} Amplification was carried out in a GeneAmp® PCR System 9600 thermocycler (Perkin Elmer, Waltham, USA) with the programme 95°C for two minutes, 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for two minutes, and 72°C for five minutes. PCR products were purified using MultiScreen PCR Plates (Millipore GmbH, Schwalbach, Germany) following the manufacturer's instructions.

The ABI Prism BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Darmstadt, Germany) was used together with the M13(-20) forward primer (5'-GTA AAA CGA CGG CCA CT-3') for dideoxy chain termination sequencing. 1µl template DNA (approx. 5ng), 2µl Reaction Mix (provided with the kit), 0.5µl primer (10mM) and 1.5µl H₂O_{bidest} were utilised per sequencing PCR, which was conducted in 30 cycles of 96°C for ten seconds, 50°C for five seconds and 60°C for four minutes in the above-mentioned thermocycler. PCR products were precipitated with one volume of isopropanol, resuspended in 2µl FAD buffer (50mg blue dextran per ml formamide) and denatured at 95°C for two minutes.

For size separation of the fragments generated in the sequencing PCR, samples were loaded on a denaturing acylamide gel [6M urea, 1x TBE (89mM tris, 89mM boric acid, 2mM EDTA, pH 8.3), 5% Long Ranger® Gel Solution (Lonza Rockland Inc., Rockland, USA); polymerised using 175µl ammonium persulfate (10%) and 24.5µl tetramethylethylenediamine per 25ml gel solution] and analysed on an ABI Prism® 377 DNA Sequencer (Applied Biosystems, Foster City, USA). Base calling was done by means of the Sequencing Analysis Software version 3.2 (Applied Biosystems). Sequence data were checked and, if necessary, manually edited using the Chromas Lite software version 2.01 (Technelysium Pty. Ltd., Helensvale, Australia). Sequences were aligned against each other with the aid of ClustalW software (<http://www.ebi.ac.uk/Tools/clustalw/>) and compared with published sequences using the NCBI BLAST query form (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) with blastn and tblastx algorithms (Altschul *et al.*, 1997).

2.1.5 Statistics and cluster analysis

Polymorphism information content (PIC), a measure of informativeness of a genetic marker, was calculated for each marker using a simplified formula according to Anderson *et al.* (1993):

$$\text{PIC} = 1 - \sum_{i=1}^n p_i^2$$

where p_i is the frequency of allele i and n is the number of allelic states.

Nei's measure of the average gene diversity within populations per locus H_s (Nei, 1973) was estimated by the formula:

$$H_s = \frac{1}{k} \sum_{s=1}^k [1 - q_s^2 - (1 - q_s)^2]$$

where k is the total number of loci and q_s is the frequency of one of the two alleles at the s th biallelic locus. Only polymorphic markers were regarded for this analysis.

Genetic similarities between samples and visualising dendrograms were computed by means of the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) version 2.20 (Rohlf, 2006) software package. A similarity matrix was calculated from the original binary data matrix using the Jaccard coefficient (Jaccard, 1908). These data were hierarchically clustered using the unweighted pair group method with arithmetic mean (UPGMA; Sokal & Michener, 1958) and a corresponding dendrogram was constructed.

The same procedure was applied in order to identify markers with common scoring patterns.

2.2 Gene expression under water-deficit stress

2.2.1 Water-deficit stress experiments

Two controlled environment (CE) water-deficit stress experiments were carried out. The first one was conducted from October 2006 until March 2007 in order to obtain leaf material for the construction of MPSS libraries containing differentially expressed genes. The second one, carried out from May 2008 until October 2008, served to validate the MPSS data and display the temporal expression kinetics of selected water-deficit stress-relevant genes using the microarray technology.

2.2.1.1 Plant materials

As bambara groundnut only exists in the form of more or less heterogenic landraces, seeds from the original seed lots were selected for the 2006 CE experiment in order to reduce the risk of genotypic variation within landraces. Fifteen seeds of each landrace were chosen on the basis of the most frequently occurring characteristics (size, shape and colour) within the

population. Thus, by selecting for dominating genotypes, it is more likely to obtain robust data representative of the entire landrace. Four landraces potentially differing in their capability to tolerate water-deficit stress were chosen (Table 1) for the construction of MPSS libraries. While landraces are generally regarded as being adapted to their areas of cultivation and thus able to cope with the predominant stresses of the particular environment (Zeven, 1998), there is only little experimental evidence in the literature supporting this assumption for bambara groundnut due to the lack of concerted research efforts so far. Reviewing several glasshouse and field trials (Collinson *et al.*, 1999; Berchie *et al.*, 2002; Mwale *et al.*, 2003; Fleißner, 2006; Mwale *et al.*, 2007) tended to rank the landraces in the following order in terms of pod yield under water-deficit stress: DipC = AS-17 > Swazi Red > LunT. This correlates with the annual rainfall in the respective sites of collection.

Several bambara groundnut landraces are known which achieve acceptable yields under drought conditions by exhibiting escape mechanisms (see 1.2.3). However, as it is not clear to what extent a shortened reproductive cycle contributes to water-deficit stress tolerance in bambara groundnut, interpreting the results may become complicated. Thus, only landraces with similar times to maturity (ca. 120 to 140 days after sowing) were chosen.

Table 1. List of bambara groundnut landraces and their characteristics used in the 2006 CE experiment. N/A: not available.

Landrace	Origin	Annual precipitation	Testa colour	Growth habit
AS-17	South Africa	N/A	cream with little rhomboid spots on both sides of the hilum	bunch
DipC	Diphiri, Botswana	527mm	cream	bunch to semi-bunch
LunT	Lungi, Sierra Leone	3590mm	cream to tan	semi-bunch
Swazi Red	Manzini, Swaziland	1391mm	dark red	bunch

The 2008 CE experiment was conducted with each 18 plants of DipC and LunT. DipC seeds were the progeny of the plant selected for the MPSS library and thus, due to the self-pollinating nature of bambara groundnut, likely to be a pure line. This was not possible for LunT, so that seeds from the original heterogeneous landrace were used. In order to prevent phytosanitary problems influencing the water-deficit stress experiment, seeds were surface-sterilised in 1% NaOCl solution for five minutes and subsequently rinsed with water for ten minutes.

2.2.1.2 Growing conditions

Both experiments were conducted in a growth room (VUZPHI, Heraeus-Vötsch GmbH, Balingen, Germany) with artificial lighting. Cool white light (58W, Osram GmbH, Munich, Germany) and GRO-LUX® (58W, Havells Sylvania GmbH, Erlangen, Germany) fluorescent tubes at a ratio of 2:1 provided around $150\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ photosynthetically active radiation. Daylength was set to 16 hours during the vegetative phase (long-day) and twelve hours from the onset of flowering (short-day). Temperature was set at 30°C (day) and 25°C (night) in the first experiment, the second one was conducted under 28°C/23°C day/night temperatures due to technical reasons. Relative humidity was 50% in both experiments.

The plants were grown in plastic pots of 19cm diameter. These were laid out with water-permeable fleece in order to avoid loss of soil and restrict root growth but to prevent waterlogging. Pots were filled with 3.50kg air-dried and steamed natural sandy soil (pH 4.6, determined using the CaCl_2 method) collected from a sand pit near Amberg (Upper Palatinate, Bavaria). The absence of nitrogen-fixing symbionts was compensated through using Flory® 3 (EUFLOR GmbH, Munich, Germany) compound fertiliser containing 15% N, 10% P, 15% K and 2% micro-nutrients, which was applied three times in two-weekly intervals to reach a total amount of 0.4g N per pot.

Before the water-deficit treatment, irrigation was carried out manually in the 2006 experiment. Each plant received 1l water per week, partitioned into three applications. In the follow-up experiment, automatic flood irrigation was used with tensiometer settings at -90 hPa and a flooding time of five minutes.

2.2.1.3 Experimental set-up and sampling

Water-deficit treatments were initiated when all plants had begun flowering. In the first experiment in 2006, this was the case at 52 days after sowing (DAS). As genetically identical plant materials were not available and it was not intended to include landrace replications in the MPSS libraries, sampling of the control variant was done before the treatment. The two youngest fully developed leaves of each the three phenotypically most similar plants per landrace were cut off at the petioles, immediately frozen in liquid nitrogen and stored at -80°C until further use. Then, the weekly water dosage was reduced to 35% of non-limiting conditions. After seven days of reduced irrigation (59 DAS), the same plants were sampled again. Thereafter, irrigation was restored to non-limiting conditions in order to allow full plant recovery and maximise seed harvest for the next experiment.

In 2008, the treatment was started at 61 DAS. Pots had been filled with excess water before sowing and water-holding capacity was determined when no more water was dripping. The amount of stored water in the pots was averaged and rounded to 750ml. During the stress phase, pots were weighed daily. Watering took place in two-day intervals. Each one half of the plants (nine plants per landrace; complete randomisation) was replenished to 100% pot

water-holding capacity, neglecting plant weight, to serve as control plants. The treated variant was watered to one third of the water-holding capacity (total pot weight 3.75kg, corresponding to 250ml water; Fig. 3). The water-limited treatment was applied for nine days. Calculated to the average water dosage per seven days, this resulted in 994 and 1020ml for DipC and LunT reference plants, respectively, and in 358 and 370ml for stressed DipC and LunT, respectively. Thus, values were comparable to the 2006 experiment (1000 vs. 350ml). From 70 DAS on, all plants were subjected to the non-limiting watering regime. Samples of both variants were collected at six dates, viz. after one, two, four and eight days of reduced irrigation (1RI, 2RI, 4RI and 8RI), and after one and three days of restored full irrigation (1REC and 3REC) in order to investigate gene expression during the recovery from water-deficit stress (Fig. 3). Four leaflets from the youngest fully developed leaves were harvested from each three preselected plants per landrace (biological replications) at every sampling date. Due to limitations in space and seeds, plants had to be sampled twice. Intervals between samplings were maximised in order to keep potential sampling effects as low as possible. Moreover, all samplings were carried out at the same time of day (9:00 a.m.) to avoid unintended influences through the plants' diurnal rhythm.

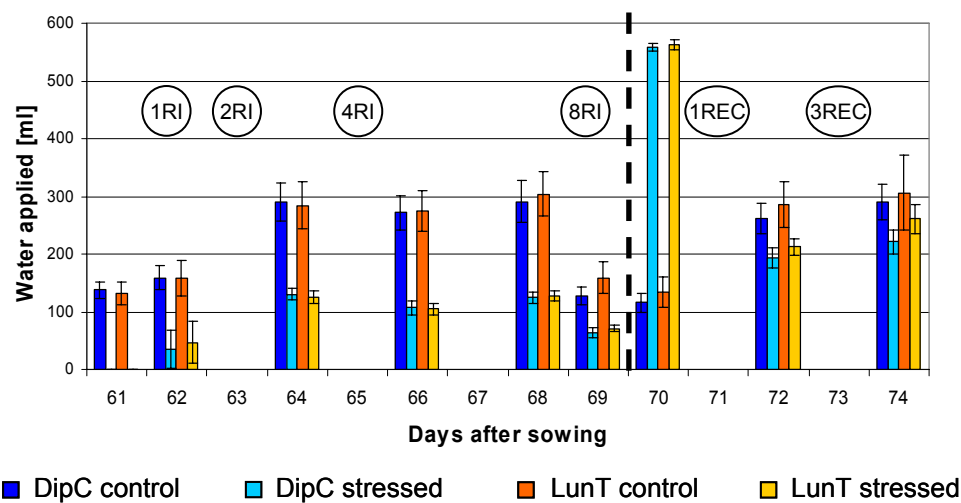


Fig. 3. Water dosages in the 2008 time-course CE water-deficit stress experiment. Error bars represent standard deviations from the mean of each nine individual waterings. Circles indicate sampling dates. The dashed line marks the beginning of the recovery phase.

2.2.1.4 Physiological measurements

A number of physiological measurements were conducted accompanying the second CE water-deficit experiment. At the end of the water-deficit stress phase, i.e. at 69 DAS and 70 DAS, respectively, single light-exposed leaves (two landraces, two treatments, triplicate sampling) were cut (simulating another type of water-deficit stress) and used to measure chlorophyll fluorescence and leaf temperature. One lateral leaflet was fixed in a MINI-PAM

Photosynthesis Yield Analyzer (Heinz Walz GmbH, Etterich, Germany), pulsed measuring light was applied and photosynthesis yield [$Y=(F_m-F_0)/F_m$] was recorded, together with leaf temperature, in one minute intervals under growth room conditions. Measurements were stopped after one hour.

Each two leaflets of the same plants were used for determining osmotic adjustment. Cell sap was pressed using plastic cylinders and 10 μ l were injected in a VAPRO® 5520 vapour pressure osmometer (Kreienbaum Wissenschaftliche Meßsysteme e.K., Langenfeld, Germany).

After the last sampling for gene expression studies, plants were maintained under the automatic flood irrigation regime as described above until maturity between five and six months after sowing. Seeds from eight plants per landrace were harvested, dried at room temperature for at least two weeks, counted, shelled and weighed to quantify yield. Statistical analysis was carried out using Student's t-test in Microsoft® Excel 2003.

2.2.2 RNA isolation

Total RNA from bambara groundnut leaves was isolated following a protocol after Chang *et al.*, 1993.

The frozen plant material was ground to a fine powder under liquid nitrogen using mortar and pestle. 50ml reaction tubes were filled with 13ml extraction buffer [2% (w/v) CTAB, 2% (w/v) polyvinylpyrrolidone, 100mM Tris-HCl pH 8.0, 25mM EDTA and 2M NaCl] and 260 μ l β -mercaptoethanol and 7.2 μ l spermidine and heated to 65°C in a shaking water bath. Around 0.5g of the homogenised leaf material was added to the buffer and vortexed for one minute. The suspension was purified from proteins and polysaccharides with 13ml chloroform/isoamyl alcohol (24:1), followed by centrifugation at 12,000g for ten minutes at room temperature. The supernatant was transferred into a fresh tube and the chloroform purification step was repeated once. RNA was precipitated by adding one third volume 8M LiCl overnight at 4°C. After centrifugation at 12,000g for 20 minutes at 4°C, a pellet formed at the bottom of the tube. The supernatant was discarded and the pellet was dissolved in 500 μ l SSTE buffer (1M NaCl, 0.5% (w/v) SDS, 10mM Tris-HCl pH 8.0 and 1mM EDTA) and transferred to a 2ml reaction tube. Again, the solution was purified with 500 μ l chloroform/isoamyl alcohol. After brief centrifugation at 9,000g, the supernatant was precipitated using 1ml ethanol (100%) for at least two hours at -20°C. Pelleted RNA was obtained by centrifuging at 13,000g for 20 minutes at 4°C. After completely removing the ethanol, total RNA was resuspended in an appropriate volume of DEPC-treated distilled and deionised water (usually 20 μ l) and stored at -80°C.

All glassware, mortars, pestles and spatula were baked at 180°C for at least four hours in order to inactivate ribonucleases degrading the samples. Plastic tubes and solutions were autoclaved at 121°C if possible.

RNA concentrations and purity were determined spectrometrically using a Genesys™ 10 Bio spectral photometer (Thermo Electron Corporation, Madison, USA). Optical density (OD) of a 1:100 solution was measured at 230nm, 260nm and 280nm. The OD₂₆₀ value was multiplied by four to obtain the concentration in $\mu\text{g}\cdot\mu\text{l}^{-1}$. An OD₂₆₀/OD₂₈₀ ratio between 1.8 and 2.1 and an OD₂₆₀/OD₂₃₀ ratio above 2 indicated adequate purity for the subsequent reactions. Furthermore, 1 μl of the undiluted RNA sample was loaded on a 0.8% ethidium bromide-stained agarose gel and run at 60V for ca. 20 minutes. The presence of two distinct and sharp ribosomal bands with the 28S ribosomal subunit RNA fluorescing approximately twice as strong as the 18S subunit showed the integrity of RNA strands. In addition, samples were checked for the absence of high molecular weight nucleic acids (DNA).

2.2.3 cDNA libraries and high-throughput pyrosequencing

cDNA libraries for high-throughput sequencing from eight total RNA populations (four genotypes by two treatments) were prepared by vertis Biotechnologie AG, Freising, Germany. The process is described in detail by Eveland *et al.* (2008).

Total RNA was transcribed to cDNA using a biotinylated T₁₂ primer fused to the 454 sequencing primer B. Purified cDNA was bound to streptavidin-coated beads and digested with *Nla*III to create four-base overhangs (CATG). Adapters containing the 454 sequencing primer A, a three-base multiplex key and a four-base overhang complementary to the restriction site were ligated to the restriction fragments. Unligated adapters and unbound cDNA fragments were removed and the specific 3'-ends were eluted from the beads. After concentration of the desired 5'-A-cDNA-B-3' strands, these were pooled in equal parts for high-throughput sequencing.

454 sequencing was conducted by Eurofins MWG Operon, Ebersberg, Germany using a 454 GS-20 instrument (Roche Diagnostics GmbH, Roche Applied Science, Mannheim, Germany), following the methods in Margulies *et al.*, 2005. Single-stranded template DNA fragments were bound to 28 μm beads under dilution-based conditions that favoured binding one strand per bead. The beads were then captured in droplets of an aqueous PCR reaction mixture within an oil emulsion. Each droplet served as separate PCR microreactor. After PCR amplification, beads carried millions of unique DNA template copies. Thereafter, the emulsion was broken and DNA strands were denatured. The beads were then individually deposited by centrifugation in 44 μm wells of a fibre-optic slide. Smaller beads carrying immobilised ATP sulfurylase and luciferase required for the pyrosequencing reaction were added to the wells. Nucleotides were then cyclically flowed over the microtiter plate. After each nucleotide, wells were washed with a buffer containing apyrase. Nucleotide incorporation into the complementary strand (sequencing-by-synthesis) resulted in the release of inorganic pyrophosphate and the generation of photons. Light signals were recorded by

means of a charged-coupled device camera. Raw signals were background-subtracted, normalised and quality-filtered before base-calling was performed.

2.2.4 MPSS data analysis

The 454 sequencing-derived data in the FASTA format were processed using several text editing commands of the Linux 2.6.18.8 operating system and the software package TGICL (Perteau *et al.*, 2003; downloaded in July 2007 from <http://compbio.dfc.harvard.edu/tgi/software>).

The first step in data analysis comprised partitioning the full library into the eight sub-libraries according to the three-base multiplex keys. Only sequences with the exact match of the nucleotide sequences of the adapters and the *Nla*III restriction site at the 5'-end were retained; residual sequences were rejected. Next, adapters were removed as artificial sequences would influence further clustering and annotation of the sequence tags. Instead, the tags were text-flagged in the header line for later re-partitioning. The sub-libraries were then concatenemered again and cleaned by truncation of polyA-tails and removal of low-quality reads (tags with more than 3% undetermined bases) using the SeqClean script with default settings provided with the TGICL package.

Clustering the cleaned full library into groups with stringent similarity was conducted with the aid of the TGICL programme which uses a modified version of megablast (Zhang *et al.*, 2000) to perform pairwise sequence alignments. These are then filtered and used to build subsets of sequences by a transitive closure approach. This was found to work best with setting parameters to at least 40 overlapping bases with at least 95% identity of the overlapping region and at most one base overlap distance from the sequence end.

The number of transcript tags within every sub-library was then determined by counting the genotype-and-treatment-specific text flags using the Linux 'grep' command. Data were normalised by dividing the number of tags in a cluster by the number of total tags in the respective sub-library. These values were expressed as transcripts per million (TPM) through multiplication by 10^6 . Consensus sequences for each cluster were computed using the cap3 assembly programme (Huang & Madan, 1999) with stringency of the overlapping region set to 97%. These contigs reduced the number of sequence tags per cluster to around 10% representative sequences, which were then used to validate the purity and to annotate the clusters by comparisons with sequences in the NCBI 'nucleotide collection (nr/nt)' and non-human/non-mouse EST (est_others)' public databases. This was done by loading the sequences into the BLAST client programme netblast-2.2.16-ia32 for Linux (downloaded in August 2007 from <ftp://ftp.ncbi.nih.gov/blast/executables/LATEST>) and applying the blastn algorithm with a threshold E-value of $1e^{-4}$. In case BLAST searches resulted only in unannotated ESTs, these were taken as queries and searched against the 'nucleotide collection' database with a cut-off E-value of $1e^{-10}$.

Functional classification of the clusters was done manually according to the functional catalogue of the Munich Information Center for Protein Sequences (Ruepp *et al.*, 2004) by means of bibliographic searches.

Expression factors were calculated by dividing the greater value by the lower value in each case in order to achieve values symmetric about zero for both up- and downregulated genes. For comparison with microarray data, normalised tag numbers were \log_2 transformed and the \log_2 ratio of the stressed and non-stressed variants was calculated.

2.2.5 Selection of sequences for microarray analysis

Based on the data analysis of the MPSS libraries, a set of 132 tentative unigenes (clusters) was chosen for validating and further investigating the water-deficit stress response in bambara groundnut using the microarray technology. Selection criteria were a) a broad coverage of functional categories, b) clear profiles of differential expression, c) qualitative or quantitative differences between the two landraces used for the 2008 CE experiment, d) significance for water-deficit stress tolerance according to the literature, e) potential novelty (no BLAST homology) and f) adequate 454 read length for the design of 50-mer oligonucleotides.

The abundance of individual sequences per cluster, potential sequencing errors and the occurrence of sequence variation between the four genotypes made it necessary to further purify the clusters to one correct read. Thus, all transcript tags within a cluster were loaded into the nrcl ('non-redundification clustering') programme (included in the TGICL package) which creates sub-clusters covered by representative 'parent' sequences. By setting stringency to 100% and comparing the number and origin of contained reads, this utility revealed incorrect bases as well as SNPs and InDels, which were located using ClustalW. Consequently, where this was possible, only the conserved region was retained for the design of oligonucleotides.

2.2.6 Oligonucleotide and microarray design

The design of 50-mer oligonucleotides was conducted by Ocimum Biosolutions Ltd., Hyderabad, India. Based on BLAST and Smith-Waterman-Analysis, the risk of unwanted cross-hybridisation was minimised. Quality parameters like GC content, melting point and self annealing score were calculated and optimised. A total of 132 oligonucleotides (see Table 16, Appendix) were then synthesised and printed in duplicates onto epoxy-coated glass slides. Furthermore, three oligos complementary to the SpotReport® Alien® spike-in mRNAs 2, 7 and 8 (Stratagene, La Jolla, USA) were included as internal references for array normalisation. The array was designed with four sub-arrays each consisting of an 8x9 grid with a spot distance of 400µm.

2.2.7 cDNA synthesis and hybridisation

cDNA synthesis was performed according to a direct labelling of single-stranded cDNA protocol provided by the array manufacturer. Every sample was labelled with one fluorescent dye only and hybridised to a single array. Dye-swap, i.e. multiplexing of two or more dyes as often conducted in microarray experiments, was not carried out.

Prior to cDNA synthesis, a low-C dNTP mastermix was prepared consisting of each 5mM dATP, dGTP and dTTP and 2mM dCTP (Carl Roth GmbH & Co. KG, Karlsruhe, Germany). Furthermore, defined amounts of reference mRNAs were included in the transcription assay. Therefore, spike-in controls were diluted to 500pg* μl^{-1} (Alien 2), 50pg* μl^{-1} (Alien 7) and 5pg* μl^{-1} (Alien 8) and mixed in equal volumes.

For each reaction, 40 μg total RNA was diluted in DEPC-treated water to a volume of 13.5 μl . 3 μl spike-in mix were added. Selective reverse transcription of polyA-mRNA was achieved by utilising 1 μg oligo(dT)₁₂₋₁₈ primer (Invitrogen GmbH, Karlsruhe, Germany). The primer was annealed in a heating block at 65°C for ten minutes, followed by ten minutes incubation at room temperature and two minutes chilling on ice. Afterwards, 300U Superscript® II Reverse Transcriptase (Invitrogen), 8 μl 5x first-strand buffer, 4 μl 100mM DTT (both provided with the enzyme), 4 μl dNTP mix and 4 μl 1mM Cy3-dCTP (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) were added. First strand synthesis was conducted at 39°C for two hours.

The reaction was terminated and template RNA was degraded by adding 10 μl 1M NaOH with ten minutes incubation at 65°C. The reactions were neutralised with 10 μl 1M HCl and 200 μl 1x TE buffer (pH 7.5). Primers, RNA fragments, unincorporated dNTPs and fluorescent dyes were removed using QIAquick® PCR Purification Kits (Qiagen) following the manufacturer's instructions. cDNA was eluted two times with 40 μl DEPC-treated water each. cDNA concentration (1) and incorporation of the fluorescent dye (2) were determined using a photo-meter and the following formulae:

$$(1) \beta_{\text{cDNA}}[\text{ng} \cdot \mu\text{l}^{-1}] = \text{OD}_{260} \cdot 37 \text{ng} \cdot \mu\text{l}^{-1}; (2) n_{\text{Cy3}} [\text{pmol}] = (\text{OD}_{550} - \text{OD}_{650}) \cdot V[\mu\text{l}] \cdot 0.15^{-1}$$

At least 400ng of each labelled cDNA were evaporated to a volume less than 5 μl in a vacuum centrifuge at 45°C. Care was taken not to completely dehydrate the cDNA samples.

Hybridisation frames (provided by the manufacturer) were stuck to microarray slides in a way that the spotted area was surrounded. Concentrated cDNA samples were resuspended in 120 μl preheated (42°C) salt-based hybridisation buffer (provided by the manufacturer) and denatured at 95°C for three minutes, followed by subsequent chilling on ice for two minutes. The hybridisation solution was then pipetted onto the slides and covered with polyester cover slips. Slides were placed in a IHC1 hybridisation chamber (Quantifoil Instruments GmbH, Jena, Germany), which was humidified with 3 μl DEPC-treated water, and incubated in a gently shaking hybridisation oven at 42°C for 16-20 hours. Cover slips and hybridisation frames were then removed and slides were washed in three steps with washing buffers of

different stringencies [washing buffer 1: 2x SSC (0.3M NaCl, 30mM trisodium citrate dehydrate, pH 7) and 0.1% (w/v) SDS; washing buffer 2: 1x SSC; washing buffer 3: 0.5x SSC]. All washing steps were carried out in an orbital shaker for five minutes at 30°C and 200rpm. Finally, washed slides were individually placed in 50µl tubes and dried by centrifugation at 500g for two minutes.

The fluorophore and labelled cDNA samples were protected from light at all times by using opaque reaction tubes and wrapping equipment in aluminium foil to avoid photobleaching.

2.2.8 Microarray scanning and data analysis

Processed slides were scanned on a GenePix® 400A Microarray Scanner (Molecular Devices, Sunnyvale, USA). Fluorescence of the Cy3 dye was excited by a green laser at a wavelength of 532nm. The sensitivity of the photomultiplier tube was adjusted individually for each experiment. Images were captured with the GenePix Pro 6.1 software saved as TIFF files. Quantification of spot intensities was done using the ArrayVision™ software version 8.0 (Imaging Research Inc., St. Catharines, Canada). Stored images were converted into greyscale using Adobe Photoshop® 8.0.1 (Adobe Systems Inc., San Jose, USA). A grid was defined according to the microarray layout and spots were labelled by means of a corresponding annotation file. Spots were aligned semi-automatically; local background was calculated from the corners between individual spots. Results (background-corrected signal intensities) were exported to Microsoft Excel tables.

Arithmetic means were calculated from duplicated spots on the microarray. These values were divided by the signal intensities of Alien 2 mRNA in order to normalise slides. Expression data were \log_2 transformed and arithmetic means from each three replicated experiments were calculated. \log_2 ratios [$\log_2(\text{stressed}/\text{non-stressed})$ and $\log_2(\text{DipC}/\text{LunT})$, respectively] were calculated for corresponding data points. Hierarchical clustering of gene expression profiles was carried out using the Cluster 2.1 programme (Eisen *et al.*, 1998). Average linkage clustering was computed from uncentred correlation metrics. These data were visualised in the form of dendrograms using TreeView 1.60.

k-Means clustering of gene expression profiles according to Soukas *et al.* (2000) was performed in the TIGR MultiExperiment Viewer 4.3 software package (Saeed *et al.*, 2003). The number of clusters was determined by empirically testing various values. Euclidean Distance was used to calculate the distance metrics and the iterations parameter was set at 50.

For the comparison of single gene expression data between treatments or landraces, respectively, Student's t-tests (Microsoft Excel 2003) were employed on logarithmised raw data. For calculating expression factors, delogarithmised data were used and the greater value was divided by the lower value in each case.

3. Results

3.1 Genetic diversity

3.1.1 Initial discovery array

For the construction of an initial DArT marker discovery library, DNA samples from 38 diverse bambara groundnut genotypes were digested using the *PstI/AluI* restriction enzyme combination as complexity reduction method. Fragments carrying *PstI* ends on both sides were selectively amplified, pooled and cloned. Out of 1,536 colonies picked from this library, an estimated 90% of the inserts were amplified successfully and spotted in a microarray format to form the initial marker discovery array. In order to identify polymorphic clones, targets complementary to the spotted probes were prepared from a subset of 32 DNA samples and individually hybridised to the array. Scanned images were automatically converted to a series of binary scores (present vs. absent), hereafter referred to as scoring profiles or segregation signatures. Seventy-six fragments (5.5%) were assigned with at least one different score within the population of 32 genotypes and were thus termed polymorphic. A UPGMA dendrogram based on these markers showed the majority of genotypes clustering together while three samples (TVsu927 from Zambia, 1691/2 from Namibia and DipC from Botswana) were clearly separate from the rest (Fig. 4a).

A closer investigation of the 76 polymorphic marker segregation signatures using a UPGMA tree from a transposed similarity matrix suggested that 13 clones exhibited unique scoring profiles while six groups of clones emerged that consisted of identical patterns. One such group was strikingly large and contained 51 clones, the others included two or three clones with equal scores each. To gain more insight into this phenomenon, all polymorphic clones were sequenced. Comparing the sequences by alignment among each other and to public databases confirmed the uniqueness of the 13 clones. Three groups with each two identical segregation signatures were each resolved to singular clones (Table 2). In contrast, two groups of three clones each turned out to consist of each three identical or highly similar sequences. Consequently, only one respective clone was retained for further analysis. The large group of 51 clones broke down into one cluster with 44 redundancies and one with three repeated clones. Three clones without similarity to other fragments were identified. One clone failed in sequence analysis.

Table 2. Assumed and actual redundancy of DArT markers from the initial discovery array as revealed by sequence analysis. *One clone failed in sequence analysis.

	No. of polymorphic clones	No. of identical sequences	No. of unique sequences	No. of clones used for diversity analysis
Clones with unique segregation signatures	13	0	13	13
	51	44+3*	3	5
	3	3	0	1
Clones with repeated segregation signatures	3	3	0	1
	2	0	2	2
	2	0	2	2
	2	0	2	2
Total	76			26

Finally, 50 DArT markers were excluded due to redundancy and 26 markers of unique sequence information were retained. The discriminatory power of the DArT markers was expressed as polymorphism information content (PIC). Values ranged from 0.12 to 0.50 with an average of 0.32 (standard deviation, SD=0.14). The level of polymorphism in the initial non-redundant *PstI/AluI* library prepared from 38 diverse genotypes and screened with a subset of 32 samples was thus estimated at 1.9%. The set of 26 markers allowed the unambiguous discrimination of the germplasm used (Fig. 4b).

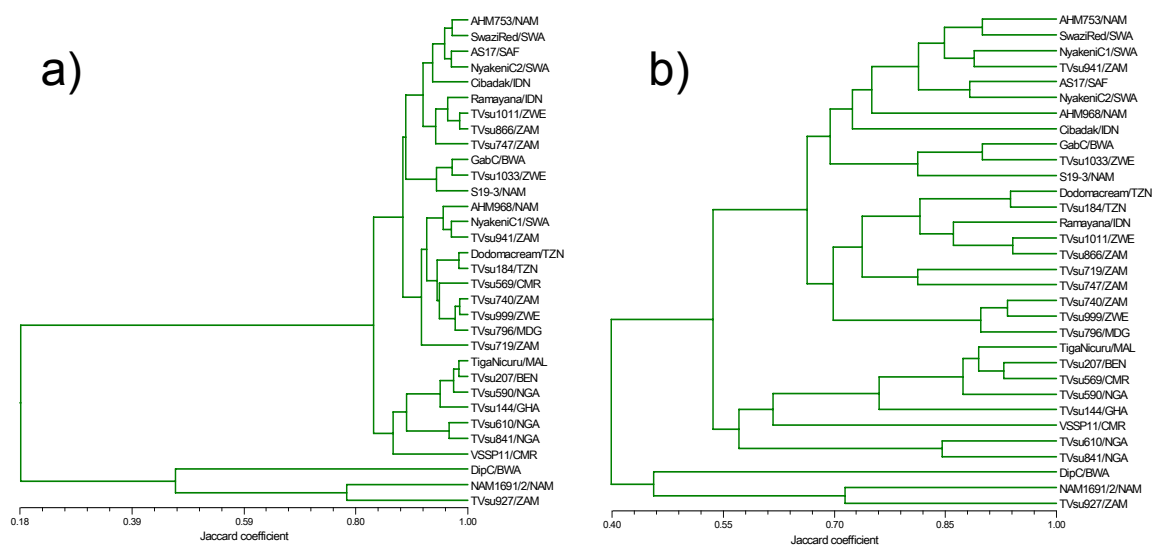


Fig. 4. UPGMA dendrograms representing 32 bambara groundnut genotypes based on the similarity matrix of a) 76 polymorphic DArT markers (with redundancy) and b) 26 unique DArT markers from the initial *PstI/AluI* array.

3.1.2 Development of full-size array

The low number of unique polymorphic DArT markers obtained from the initial discovery array necessitated further development in order to construct a viable genotyping array. Several actions were taken to increase the yield of polymorphic clones.

- a) It was attempted to widen the genetic base of the discovery panel. Therefore, the number of genotypes was increased to 94. Accessions from eight countries not represented in the initial discovery array were included.
- b) From these genotypes, another discovery library was produced according to the initial one, with an additional restriction digest using *Bgl*III to circumvent repeated enrichment of highly redundant clones. Thirteen plates (4,992 clones) were included in the full-size array.
- c) Three plates (1,152 clones) of probes from the initial discovery library were also included to expand the *Pst*I/*Alu*I(*Bgl*III) array to the fourfold of the initial array (6,144 clones).
- d) A second complexity reduction method replacing the frequently cutting restriction endonuclease *Alu*I by *Taq*I was tested. Four plates (1,536 clones) from this genomic representation were assembled with the *Pst*I/*Alu*I(*Bgl*III) clones to a full-size DArT array of 7,680 clones.

Hybridising *Pst*I/*Alu*I-digested genomic representations from the 94 genotypes to the full-size array resulted in 337 polymorphic fragments. The numbers of probes per enzyme combination contributing to total polymorphism are given in Table 3. The highest level of polymorphic clones (8.7%) was observed in the clones from the initial *Pst*I/*Alu*I library, which was derived from 38 genotypes. The triple digested library, containing clones from the extended set of genotypes, showed a polymorphism rate of 4.1%. Additionally, probes prepared using *Taq*I as an alternative frequent cutter accounted for 34 polymorphic clones. Of the 337 polymorphic clones, 180 showed redundant segregation signatures within the set of markers discovered with targets prepared from *Pst*I/*Alu*I. Seventy-five of these displayed identical scoring profiles which were highly similar to those of the highly abundant clone in the initial dataset. Comparison of 28 accessions used in both experiments yielded 26 identical scores and two differences for landraces, where it is possible that different genotypes were used. Seventy-three of these clones could be attributed to the initial *Pst*I/*Alu*I library, whereas only two were found within the sub-array containing fragments additionally digested with *Bgl*III. Thus, the latter complexity reduction method turned out to be superior to the original one.

The second complexity reduction method with targets prepared using the *Pst*I/*Taq*I enzyme combination yielded 321 polymorphisms. Again, the highest frequency of polymorphic clones was achieved in clones prepared from the same enzyme combination (6.1%). The polymorphism content of triple digested clones was almost as high as with *Pst*I/*Alu*I targets and the initial discovery library contributed another 25 polymorphic clones. After redundancy cleaning, 168 unique markers remained. The share of repeated scoring profiles with using

TaqI as the codigesting enzyme for preparing both probes and targets was 32% and thus the smallest of all combinations investigated.

Both datasets were merged and checked for equal scoring profiles again. Forty-three clones displayed the same results in both target libraries. On the other hand, 35 clones were polymorphic using both genomic representations but yielded independent scores. The conservative approach of removing all repeated discrimination patterns has led to a final dataset of 296 singleton polymorphic DArT markers for bambara groundnut. The non-redundant polymorphism rate for targets prepared using *PstI/AluI* and *PstI/TaqI* was 2.0% and 2.2%, respectively. Both complexity reductions combined yielded 3.9% polymorphic markers. PIC values ranged between 0.04 and 0.50 with a mean of 0.35 (SD=0.14). Classification according to PIC values showed that almost half of the markers (47.3%) were present in the highest class with PIC between 0.4 and 0.5. Frequency steadily declined with decreasing PIC, so that 5.7% of the DArT markers grouped in the class with lowest discriminatory power (PIC between 0 and 0.1).

Table 3. Fractions of polymorphic clones (total and after removing repetitive discrimination patterns) in the full-size DArT array depending on enzyme combinations used for producing clones (probes) and marker scoring (targets).

Enzyme combination		Total polymorphic clones		Unique scoring profiles	
Targets	Probes (from no. of genotypes)	No. of polymorphic clones	Frequency of polymorphic clones	No. of polymorphic clones	Frequency of polymorphic clones
<i>PstI/AluI</i>	<i>PstI/AluI</i> (38)	100	8.7%	16	1.4%
	<i>PstI/AluI/BglII</i> (94)	203	4.1%	128	2.6%
	<i>PstI/TaqI</i> (94)	34	2.2%	13	0.8%
	Total	337	4.4%	157	2.0%
<i>PstI/TaqI</i>	<i>PstI/AluI</i> (38)	25	2.2%	13	1.1%
	<i>PstI/AluI/BglII</i> (94)	202	4.0%	91	1.8%
	<i>PstI/TaqI</i> (94)	94	6.1%	64	4.2%
	Total	321	4.2%	168	2.2%
Merged datasets	<i>PstI/AluI</i> (38)	125	10.9%	29	2.5%
	<i>PstI/AluI/BglII</i> (94)	405	8.1%	194	3.9%
	<i>PstI/TaqI</i> (94)	128	8.3%	73	4.8%
	Total	658	8.6%	296	3.9%

3.1.3 Genotyping and cluster analysis

As both hybridisation experiments were not free from technical failure in a few genotypes, only 87 (out of 94) accessions could be used for constructing an UPGMA dendrogram. Setting the genetic similarity threshold at a Jaccard coefficient of 0.51 led to the emergence of four clusters (Fig. 6). The majority of genotypes contained in cluster I (47) originated from a south-eastern African stretch between Kenya and the Republic of South Africa. Two samples

from West Africa (TVsu841/Nigeria and TVsu155/Ghana) were also included. Furthermore, the three landraces from Indonesia formed their own sub-cluster (i) embedded in a group of East African accessions (Tanzania, Zambia, Zimbabwe and Malawi). Cluster II consists of eight samples from the inner parts of northern sub-Saharan Africa (Sudan, Central African Republic, Cameroon, Mali, Burkina Faso), while cluster III with 25 entries tended to reflect coastal proveniences (Ghana, Nigeria, amongst others) of West Africa. However, there are several overlaps between these two clusters. LunT from Sierra Leone formed a separate cluster (IV) together with an accession from Gambia (TVsu246). Both countries are located in the outmost west of Africa. Figure 5 shows the geographic location of clustered African samples.

Within the African bambara groundnut materials, no clear division based on the country of origin was obvious. The spontaneous form VSSP6 (*V.s. var. spontanea*) was not outgrouped from cultivated landraces. Regarding the dendrogram on the whole (Fig. 6), the separation of cluster I from the rest of accessions is apparent. Genetic similarity indices are generally higher than within and between clusters II, III and IV. The average allele diversity H_S within cluster I is 0.21 with 55 of the 296 loci investigated being monomorphic, whereas the combined clusters II to IV had an H_S of 0.33 and eleven monomorphic signatures.

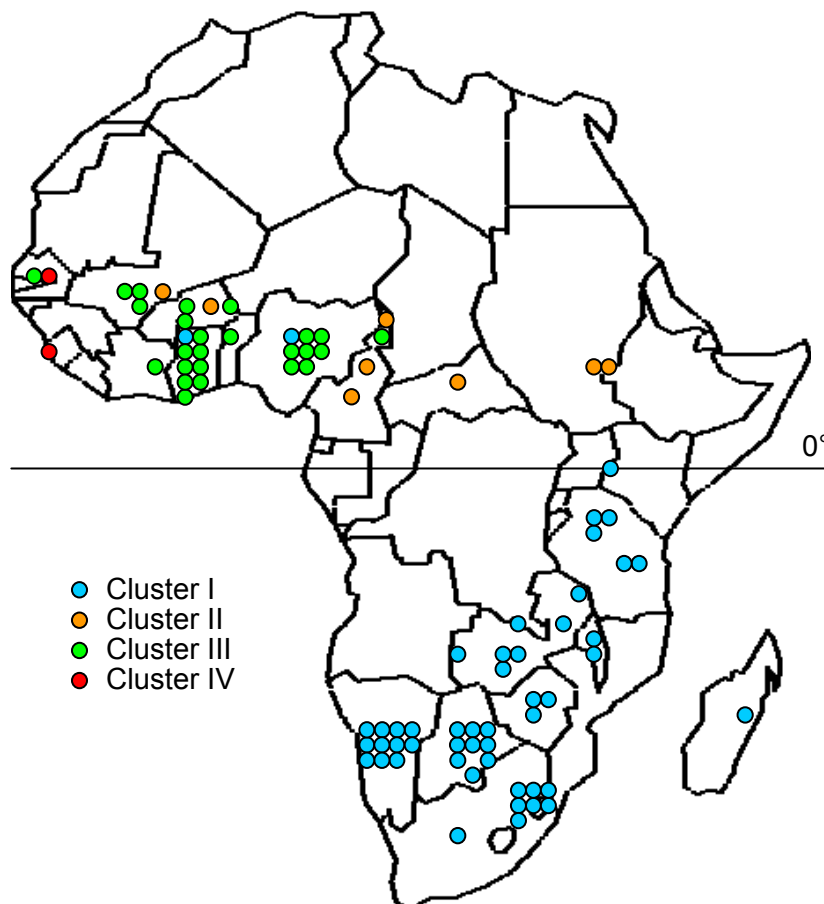


Fig. 5. Geographic distribution of 84 African bambara groundnut genotypes according to UPGMA clustering based on 296 DArT markers. In the case of missing geographic coordinates, accessions were placed in the centre of the respective country.

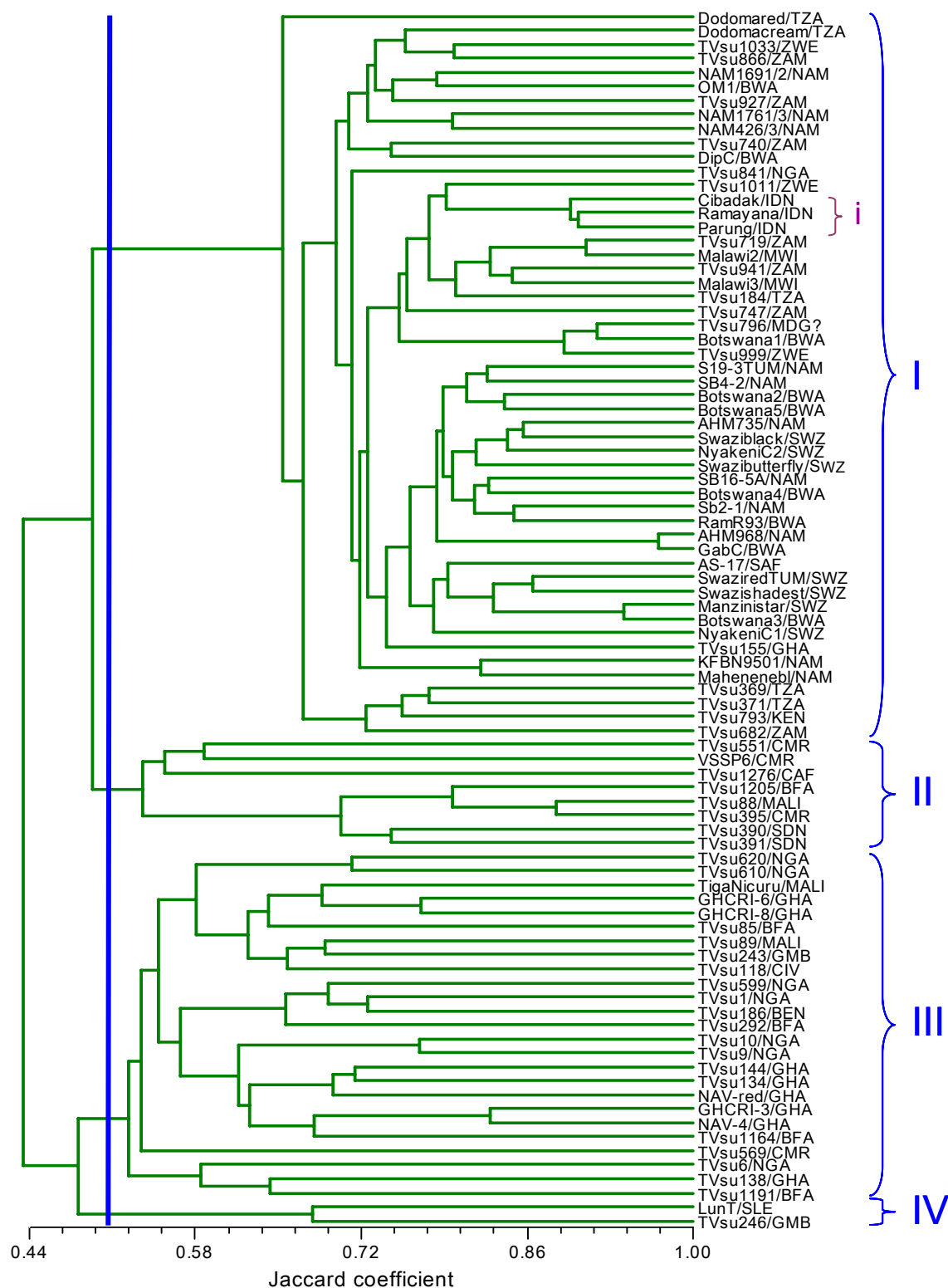


Fig. 6. UPGMA dendrogram representing genetic similarity of 87 bambara groundnut genotypes based on the similarity matrix of 296 unique polymorphic DArT markers from the full-size *PstI/AluI*(*Bg*II) and *PstI/TaqI* array using targets prepared with both complexity reduction methods. Cluster designations are explained in the text. Abbreviations for countries of collection are explained in Table 4.

3.1.4 Large-scale genotyping

After the proof of concept and the discovery of 296 singleton polymorphic markers, the DArT marker system was used to fingerprint another 342 bambara groundnut genotypes. This time, 635 polymorphic clones were discovered, of which 460 were identical to clones in the above-named analysis. Rejecting repeated discrimination patterns yielded 201 markers that allowed co-analysis of both experiments. Most of the genotypes were individuals of accessions held at the IITA and landraces from project partners. UPGMA clustering of newly genotyped accessions alone as well as of the fused dataset (not shown for lack of space) confirmed the level of genetic diversity and phylogenetic relationships of bambara groundnut germplasm illustrated above with three reasonably diverse and one genetically narrow cluster. Analogous to Figure 5, accessions were found to cluster according to their place of collection (Table 4). In particular, cluster IV, including only two members in the previous experiment, proved stable and was extended by three more genotypes. Figure 7 shows a PCoA plot produced from the same similarity matrix with genotypes colour-coded according to their geographic origin north or south of the equator. Again, two reasonably unstructured pools of germplasm differing in their genetic diversity became apparent. However, five and six genotypes, respectively, did not fit into the model of strict geographic separation of bambara groundnut germplasm.

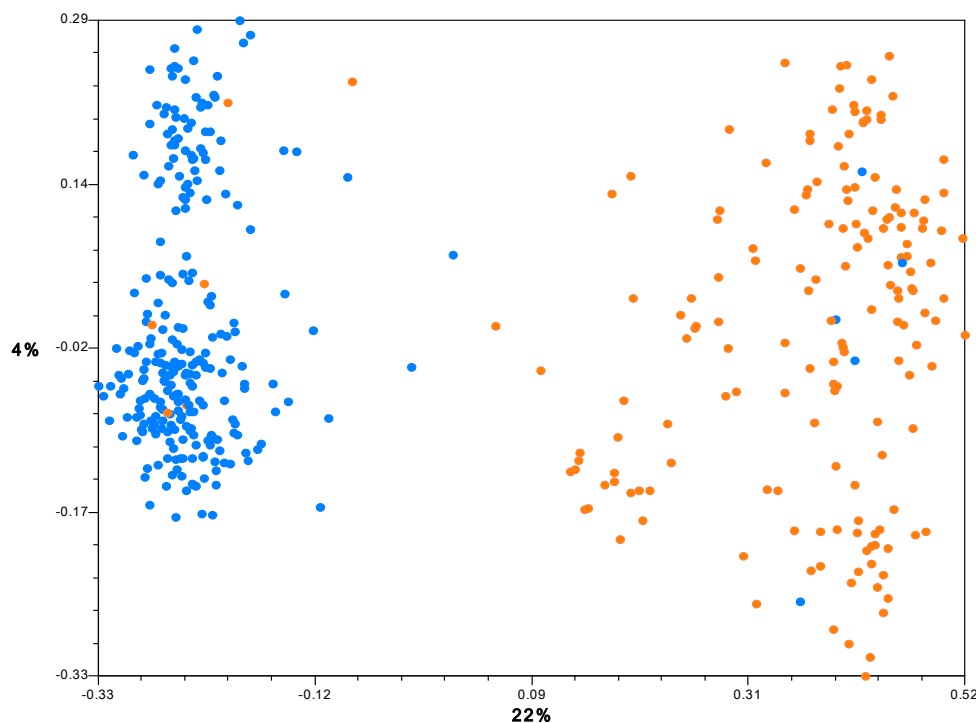


Fig. 7. Fused principal coordinate analysis of 429 bambara groundnut genotypes based on 201 DArT markers. The first two principal coordinates, accounting for 22% and 4% of total genetic variation, respectively, were plotted. Orange and blue colours indicate origins from north and south (including Indonesian germplasm) of the equator, respectively.

Table 4. Geographic distribution of 347 African bambara groundnut genotypes according to UPGMA clustering based on 201 DArT markers commonly identified in two separate analyses. Landraces represented by more than one genotype were only counted once.

Country	Number of accessions			
	Cluster I	Cluster II	Cluster III	Cluster IV
Countries north of the equator (from west to east)				
Senegal (SEN)	-	-	1	1
Gambia (GMB)	-	-	1	1
Sierra Leone (SLE)	-	-	-	2
Mali	-	1	3	-
Cote d'Ivoire (CIV)	-	-	1	-
Burkina Faso (BFA)	-	1	11	-
Ghana (GHA)	3	-	30	-
Togo (TGO)	-	-	13	-
Benin (BEN)	-	-	5	-
Nigeria (NGA)	1	1	45	1
Cameroon (CMR)	2	16	8	-
Central African Republic (CAF)	-	9	-	-
Sudan (SDN)	-	2	-	-
Countries south of the equator (from north to south)				
Kenya (KEN)	1	-	-	-
Tanzania (TZA)	25	-	3	-
Zambia (ZAM)	56	-	-	-
Malawi (MWI)	2	-	2	-
Zimbabwe (ZWE)	38	-	-	-
Botswana (BWA)	10	-	-	-
Namibia (NAM)	42	-	-	-
Swaziland (SWZ)	8	-	-	-
Republic of South Africa (SAF)	1	-	-	-

3.1.5 Intra-landrace diversity and 'exotic' germplasm

Additionally, individual genotypes of the six BAMLINK core landraces DipC (originating from Botswana; 19 genotypes), Swazi Red/Uniswa Red (Swaziland; 12), S19-3 (Namibia; 13), Ramayana (Indonesia; 10), Dodoma Red (Tanzania; 10) and Tiga Nicuru (Mali; 10) were investigated in terms of genetic diversity. The resulting dendrogram (Fig. 8) revealed considerable structural differences between landrace populations. The genetically narrowest landrace is S19-3. A total of 53 polymorphic loci was discovered which corresponds to 26% of all common singleton markers polymorphic within 429 genotypes or 0.7% of all clones screened. Average allele diversity H_S (referring to 201 markers) is 0.04. However, among the 13 independent samples of S19-3, only one duplicate was observed. Ten individual plants from the landrace Ramayana were co-analysed with single genotypes from three other Indonesian landraces – Cibadak, Parung and Gresik. UPGMA clustering displayed the latter completely merging into Ramayana instead of genetic separation of Indonesian materials.

Diversity of the Indonesian landraces was similar to S19-3 with 55 polymorphic loci and an H_S of 0.04. Dodoma Red also formed a confined cluster. Polymorphisms were detected at 90 loci. Separation of dendrogram clades occurred at a similarity coefficient of 0.77 and average allele diversity was 0.12. Tiga Nicuru presented a similar picture with 78 polymorphic markers and H_S of 0.11. The clear separation from the other landraces fits with the observation of geographic segregation in bambara groundnut germplasm (see 3.1.3). Swazi Red and Uniswa Red are regarded as synonyms for the same landrace. This assumption was confirmed through cluster analysis. Seventy-one polymorphic loci and H_S of 0.09 implied medium genetic intra-landrace diversity. However, the existence of two discrete subpopulations breaking up at a similarity coefficient of 0.81 was obvious. As genetic diversity is generally low in accessions from the region south of the equator, the example of Swazi Red partly exceeds differences between landraces. This situation was even more extreme in the case of DipC. A total of 113 polymorphisms and average allele diversity of 0.14 led to DipC individuals or sub-clusters spanning almost the entire sub-equatorial cluster. In addition to Indonesian materials, five IITA accessions from Madagascar were analysed as a second example of bambara groundnut in a geographically isolated place. Four accessions clustered together, but segregated at a similarity coefficient of 0.84. One accession (TVsu810) was clearly outgrouped from the other Madagascan materials. All accessions were found in the sub-equatorial cluster.



Fig. 8. UPGMA dendrogram representing genetic similarity of individual genotypes from six bambara groundnut landraces and five Madagascan accessions based on the similarity matrix of 201 unique polymorphic DArT markers from the full-size *PstI/AluI*(*Bg*/II) and *PstI/TaqI* array. Genotypes belonging to the same landrace are highlighted in the same colour.

3.2 Gene expression under water-deficit stress

3.2.1 MPSS expression profiling

3.2.1.1 Data assembly and analysis

The 454 sequencing run resulted in a total of 216,022 sequence tags. The average read length was 108 nucleotides. After the raw data were cleaned by truncation of polyA-tails and 454 adapters and removal of reads that did not perfectly match the envisaged 5'-end sequence (multiplex identifier plus *Nla*III restriction site) and those with a number of unidentified bases exceeding 3% and, 197,400 partial transcripts with an average read length of 84 nucleotides remained. Partitioning the full dataset into the eight genotype and treatment-specific sub-libraries yielded an average of 24,675 (SD=2,640) sequences.

Clustering resulted in 10,583 homogenous groups ranging from two to 3,026 transcripts. Clusters were numbered consecutively by their size, starting with cluster 1 including the most abundant sequence tags relating to the full dataset. A total of 34,427 singleton sequences were not assigned to one of the clusters.

In a few cases, transcripts were not present in one of the four genotypes while there was significant expression in the others. At the same time, clusters with mRNAs in only one genotype were detected. When BLAST searches identified two clusters being derived from the same transcript by displaying identical matches to published sequences, clusters were merged and treated as one. However, this was only possible if significant homologies to mRNAs or cDNAs in the NCBI databases were obtained. An example of such a situation is depicted in Figure 9. Thus it appears that this kind of cluster split must have occurred due to sequence variation in the restriction site. Furthermore, duplication of several clusters which included transcripts from all genotypes was observed, suggesting that the restriction digest was not complete during the preparation of samples for 454 sequencing. Here, too, corresponding clusters were analysed as one. By these actions, it was made sure to the maximum possible extent that one cluster represented only transcripts derived from individual genes homologous between the four genotypes used for MPSS expression profiling.

It often appeared that direct search for homologies to annotated transcripts in the 'non-redundant' NCBI databases did not yield a significant match but search in the 'est_others' databases was successful. In such cases, sequence homologies of ESTs were used for functional classification of the respective clusters (Fig. 10).

Setting the threshold at ten transcripts per cluster, 2,425 clusters remained. Averaging the expression changes of all four genotypes yielded at least twofold upregulation for 656 clusters (27.1%) and at least twofold downregulation of 661 clusters (27.3%). That means that more than half of the genes (54.3%) were differentially expressed upon the water-deficit stress treatment.

For a more detailed analysis, only clusters containing a minimum of 40 sequence tags were considered as expression profiles became less clear when transcript abundance became low for less strongly expressed genes. Furthermore, the probability of obtaining biological information through finding homologies to annotated genes dropped in genes with lower expression levels. While there were 70 significant matches in the 100 largest clusters, clusters 501 to 600 contained only 29 annotated genes. Out of the 570 clusters with at least 40 transcripts, nine were merged due to redundancy. Among the 561 unique clusters, 270 (48.1%) could be assigned to the non-redundant BLAST database and classified in a functional category. For 138 clusters (24.6%), classification was either unclear or only ESTs without significant homologies to annotated genes were obtained. 153 clusters (27.3%) did not match any published sequence at a significance level (E-value) less or equal to e^{-4} . Annotations and expression changes of the 561 largest clusters are summarised in Table 17 (Appendix).

```

vrCipLhcb2 CATTGTCACT GGCAAAGGCC CTATTCAGAA CCTTTACGAC CATGTTGCTG 800
Cluster32 -----
Cluster176 -----
*****

vrCipLhcb2 ACCCTGTTGC CAACAATGCT TGGGCTTATG CCACCAACTT CGTCCCTGGA 850
Cluster32 ACCCTGTTGC CAACAATGCT TGGGCTTATG CCACCAACTT TGTCCTGGA 60
Cluster176 -----
*****

vrCipLhcb2 CAATGAGCAT GTCAACTTGT GCCTTCC-AG ACGCAATGCA ATGCAATCTT 899
Cluster32 CAATGAGCAT ATTAACTTGT GCCATCTTAG ----- 90
Cluster176 -----CAT GTTAACTTGT GCCATCTTAG ATGCAATGCA ATGCAATCTT 43
***** * ***** ** * *****

vrCipLhcb2 TCTGCTCATA TTTGCAGTTT TTCCCTTGTG TTTAACTCAA ATTT-CACTG 948
Cluster32 -----
Cluster176 -CTGCTCGTG TTTGCAGTTT TTCCCTTGTG TTTAACTCCA ATTTTCAA-- 90
***** * *****

vrCipLhcb2 ATGTAACTAC ACAGCGTGTA GCAGAAGAAT CTGTATGAGA ACAGTTATAT 998
Cluster32 -----
Cluster176 -----

vrCipLhcb2 ATTATAAATT TGGATCTCTG AAGTATCAAA AAAAAAA 1033
Cluster32 -----
Cluster176 -----

```

Fig. 9. ClustalW alignment of MPSS cluster 32 (with tags from DipC, AS-17 and Swazi Red) and cluster 176 (tags from LunT only) with the 3'-end of the mRNA coding for *Vigna radiata* LHCII type II chlorophyll a/b-binding protein (vrCipLhcb2). In position 861 of the *V.r.* transcript, a SNP appears between both bambara groundnut clusters, leading to a *Nla*III restriction site in LunT (highlighted in grey). Note: Although both MPSS tags overlap by 23bp, they were not assembled due to matters of stringency.

Table 5. Clusters with differential expression upon water-deficit treatment grouped under the ‘metabolism’ functional category and their putative functions.

MPSS cluster no.	Homology	Function
Upregulated		
34	Endo-1,4- β -mannanase	Carbohydrate metabolism
78	Phosphatase	Phosphate metabolism
102	CPRD14/cinnamoyl alcohol dehydrogenase	Secondary metabolism/ lignin biosynthesis
106	Asparagine synthetase	Amino acid metabolism
112	Low temperature and salt responsive protein/o-methyltransferase	Secondary metabolism
114	Chloroplast-targeted β -amylase	Carbohydrate metabolism
152	<i>myo</i> -Inositol oxygenase	Carbohydrate metabolism
166	Invertase inhibitor	Carbohydrate metabolism
182	Cytidine or deoxycytidylate deaminase	Nucleoside metabolism
224	β -Amylase 1	Carbohydrate metabolism
252	Glutaredoxin	Secondary metabolism
432	ATP sulfurylase	Sulfur metabolism
563	Glucosyltransferase	Carbohydrate metabolism
Downregulated		
9	RuBisCO small subunit	Calvin cycle
16	RuBisCO activase	Calvin cycle
27	Thioredoxin F	Calvin cycle
54	Carbonic anhydrase	Calvin cycle
99	Caffeic acid methyltransferase	Secondary metabolism/ lignin biosynthesis
105	Phosphoribulokinase	Calvin cycle
125	Acid phosphatase	Phosphate metabolism
142	Thioredoxin H	Calvin cycle
149	Alanine aminotransferase 2	Amino acid metabolism
185	1-Aminocyclopropane 1-carboxylic acid oxidase	Ethylene biosynthesis
188	Glutamine synthetase	Amino acid metabolism
359	Acetyl-CoA carboxylase	Fatty acid metabolism
380	Pyridoxine biosynthetic enzyme	Vitamin b6 biosynthesis
400	Geranyl geranyl hydrogenase	Chlorophyll biosynthesis
453	Inorganic pyrophosphatase	Phosphate metabolism
457	<i>myo</i> -Inositol 1-phosphate synthase	Carbohydrate metabolism
502	Sedoheptulose 1,7-bisphosphatase	Calvin cycle
526	Ribulose 5-phosphate 3-epimerase	Ascorbate metabolism

Fifty-six clusters belonging to the ‘energy’ functional category were identified. Among these, 54 were downregulated and mostly represent proteins of photosystems I and II, but also play a role in respiratory processes such as glycolysis (MPSS clusters 162 – phosphoglycerate kinase, 19/309/329 – fructose 1,6-bisphosphate aldolases and 49/398/411 – glyceraldehyde 3-phosphate dehydrogenase subunits), the Krebs cycle (clusters 320 – nucleoside diphosphate

kinase and 349 – malate dehydrogenase) and photorespiration (clusters 144 – serine hydroxymethyltransferase, 187 – H-protein of glycine decarboxylase and 242 – glycine cleavage complex P protein). Only two energy-related genes with differing expression profiles were found. Cluster 268, coding for a 14kDa protein of the ubiquinol-cytochrome c reductase complex, was downregulated 2.3-fold in DipC and non-significantly altered in AS-17, whereas its expression was induced in Swazi Red and LunT by the factors 2.0 and 2.2, respectively. Cluster 426, a putative subunit of a mitochondrial F1-ATPase, was significantly downregulated in Swazi Red and remained relatively constant in the other genotypes.

Although nine of 15 clusters involved in 'transcription' were termed as generally up- or downregulated, few clear expression patterns common to all four genotypes were observed.

With respect to 'protein synthesis', 13 clusters coding for ribosomal proteins were repressed in their expression under water-deficit. Furthermore, one transcript cluster corresponding to a translation initiation factor (cluster 339) showed downregulation. However, cluster 92, another putative translation initiation factor, was induced upon water-deficit as well as one transcript for a ribosomal protein (cluster 433).

Among the 22 genes functioning in the modification or degradation of proteins ('protein fate'), eight were identified as differentially expressed upon the water-deficit treatment. Seven (clusters 131 – cysteine protease, 238 – vacuolar processing enzyme, 300 – polyubiquitin, 413 – protease inhibitor, 472 – protease precursor, 493 – zinc-dependent protease and 570 – ubiquitin carrier) were upregulated. One ubiquitin-like protein mRNA (cluster 550) was strongly repressed in DipC, constant in LunT and slightly induced in AS-17 and Swazi Red.

The proportion of differentially expressed proteins with 'binding function' (eight out of 19) also was by far lower than the two thirds of affected medium and highly expressed genes. Downregulated were: Clusters 117 – RNA-binding protein, 301 – chloroplast RNA-binding protein, 325 – lectin, 476 – heat shock factor-binding protein and 486 – 14-3-3 related protein, while mRNAs for an ISCA-like protein (cluster 306), for another RNA-binding protein (cluster 377) and for a selenium-binding protein (cluster 551) were more strongly expressed in the challenged variant.

Eleven clusters were presumed to play a role in 'cellular communication/signal transduction'. Upregulated were a nodule-enhanced protein phosphatase (cluster 269) and a serine/threonine kinase (cluster 471), and clusters 143, 167 and 233, coding for a CDPK-related protein kinase, a UDP-glucuronosyltransferase and a CBS domain containing protein, respectively, were downregulated.

Thirty-five clusters were ascribed to the functional category 'cell rescue, defence and virulence' (Table 6). Among the 20 induced clusters, five different transcripts coding for LEA proteins were detected. Furthermore, four heat shock proteins and three lipid transfer proteins were identified. Six clusters were repressed after the water-deficit treatment.

Table 6. Clusters with differential expression upon water-deficit treatment grouped under the ‘cell rescue, defence and virulence’ functional category and their putative functions.

MPSS cluster no.	Homology	Function
Upregulated		
2	CPRD22/dehydrin 1	LEA protein
4	CPRD86/KS-type dehydrin	LEA protein
5	Lipid transfer protein I	Cuticle development
12	Type 2 metallothionein	Metal detoxification
24	Lipid transfer protein II	Cuticle development
36	LEA 5 protein	LEA protein
52	18kDa LEA protein (putative)	LEA protein
71	Dehydrin	LEA protein
158	Heat shock protein 22	Chaperone
209	Heat shock protein associated protein-like	Chaperone
228	Glutathione S-transferase	ROS detoxification
239	Alcohol dehydrogenase 1	Alcohol/aldehyde detoxification
243	Snakin-like cysteine rich protein	Pathogen defence
250	Lipid transfer protein (putative)	Cuticle development
254	Aldo/keto reductase	Aldehyde detoxification
261	Ferritin	Metal detoxification
283	Cytosolic ascorbate peroxidase 2	ROS detoxification
322	Fibre protein/universal stress protein	Unknown
404	Heat shock protein 70	Chaperone
451	Heat shock protein (<i>Hevea brasiliensis</i>)	Chaperone
Downregulated		
13	Type 1 metallothionein	Metal detoxification
20	Stored cotyledon mRNA/defensin	Pathogen defence
93	Pathogenesis-related protein 4.2	Pathogen defence
312	Small heat shock protein	Chaperone
356	DnaJ heat shock protein	Chaperone
454	Cu/Zn-superoxide dismutase	ROS detoxification

Four clusters were presumed to function in the ‘biogenesis of cellular components’. While cluster 534 was particularly upregulated in LunT, clusters 1 and 172, coding for cell wall proteins, were slightly and highly downregulated, respectively.

The analysis of MPSS-based transcription profiling yielded a high proportion of transcripts without available biological information. The number of up- and downregulated clusters with homologies to proteins of unknown or unclear function or unannotated ESTs was nearly equal. However, induced clusters without any significant BLAST match were 1.7-times more abundant than repressed ones, although in total, more downregulated genes were found among the 561 largest clusters.

Overall gene expression changes within the functional categories identified are summarised in Figure 11.

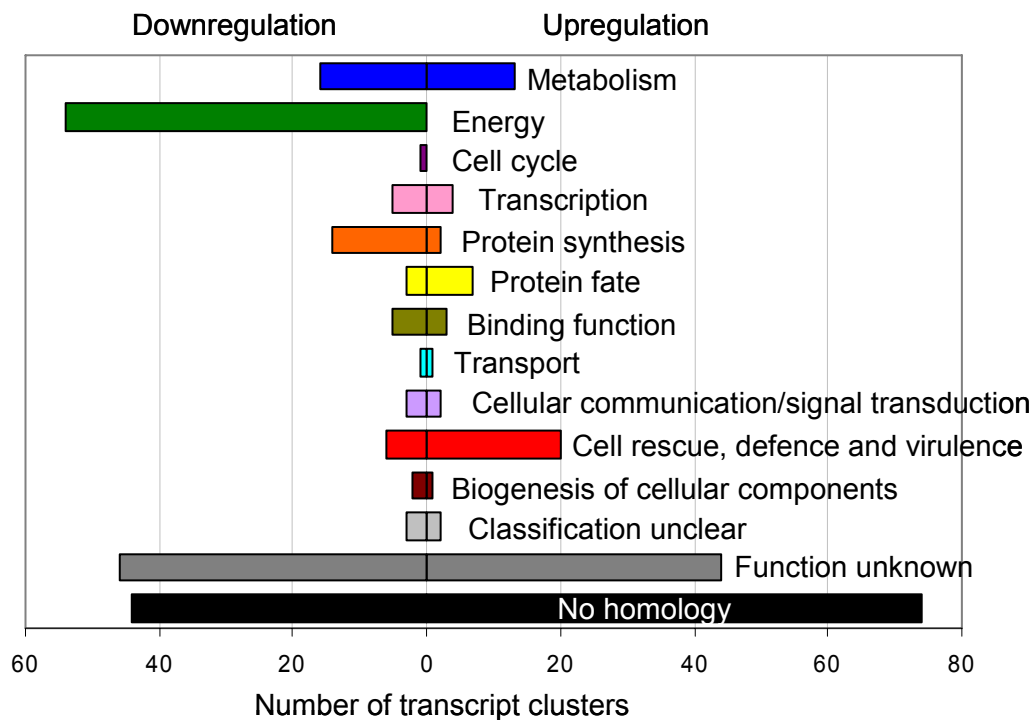


Fig. 11. Number of transcript clusters with at least twofold up- or downregulation after seven days of irrigation reduced to one third of non-limiting conditions broken down into functional categories. Numbers are based on the averaged induction/repression factors of four bambara groundnut genotypes and the 561 largest clusters.

3.2.1.3 Genotypic differences

Concerning the four genotypes taken individually, 188 clusters with a different expression profile in at least one genotype were identified at the 561-cluster level (Fig. 12a). The majority of classified genes (28) function in protein synthesis, which corresponds to 58% of the total number of genes in this category. High proportions of genotype-specific differences were also observed in the categories ‘cell cycle and DNA processing’ (66.7%), ‘transcription’ (58.8%) and ‘cellular transport’ (57.1%). The ‘energy’ and ‘cell rescue, defence and virulence’ classes were clearly underrepresented with 12.7% and 17.6%, respectively.

When comparing the two genotypes used for the follow-up experiment, DipC and LunT, the number of genes with differing expression profiles was reduced to 48 (Fig. 12b). Protein synthesis-related clusters still account for most of the classified genes. However, no more cell ‘cycle genes’ were present and only two clusters functioning in transcription remained. Whereas the percentages of all other groups strongly declined, the number of stress- and defence-related genes remained relatively constant with only one genotypic difference breaking away when the differences between DipC and LunT were related to all clusters with diverse expression profiles.

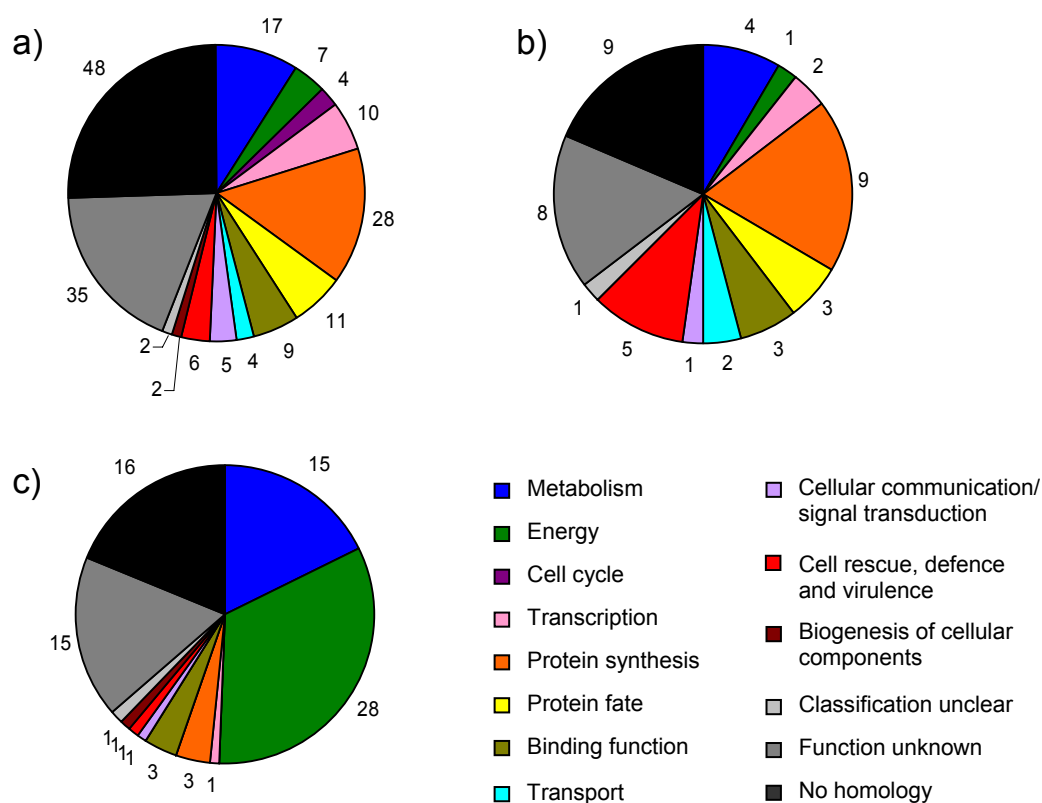


Fig. 12. Functional classification of bambara groundnut genes with a) differential expression profiles in at least one of the four genotypes; b) differential expression profiles between DipC and LunT; and c) quantitative expression differences in repressed genes between putatively drought tolerant and sensitive genotypes after water-deficit treatment.

In order to further narrow down the number of genes relevant for drought tolerance, the averaged induction/repression factors of the putatively tolerant landraces DipC and AS-17 were compared with those of the landraces adapted to humid environments (Swazi Red and LunT). Nine clusters displayed differential expression profiles between the two genotype classes and are shown in Figure 13. The expression of six of these genes is repressed or constant in the genotypes from drought-prone areas and induced in Swazi Red and LunT. With respect to the significance of genes in actively counteracting the effects of water-deficit, three clusters with (slight) upregulation in drought-adapted genotypes and strong downregulation in non-adapted genotypes were found. These show significant homologies to a valine transfer-RNA gene from *Cyanea pilosa*, a photosystem II core complex protein psbY from maize and the MYB transcription factor 123 from soybean.

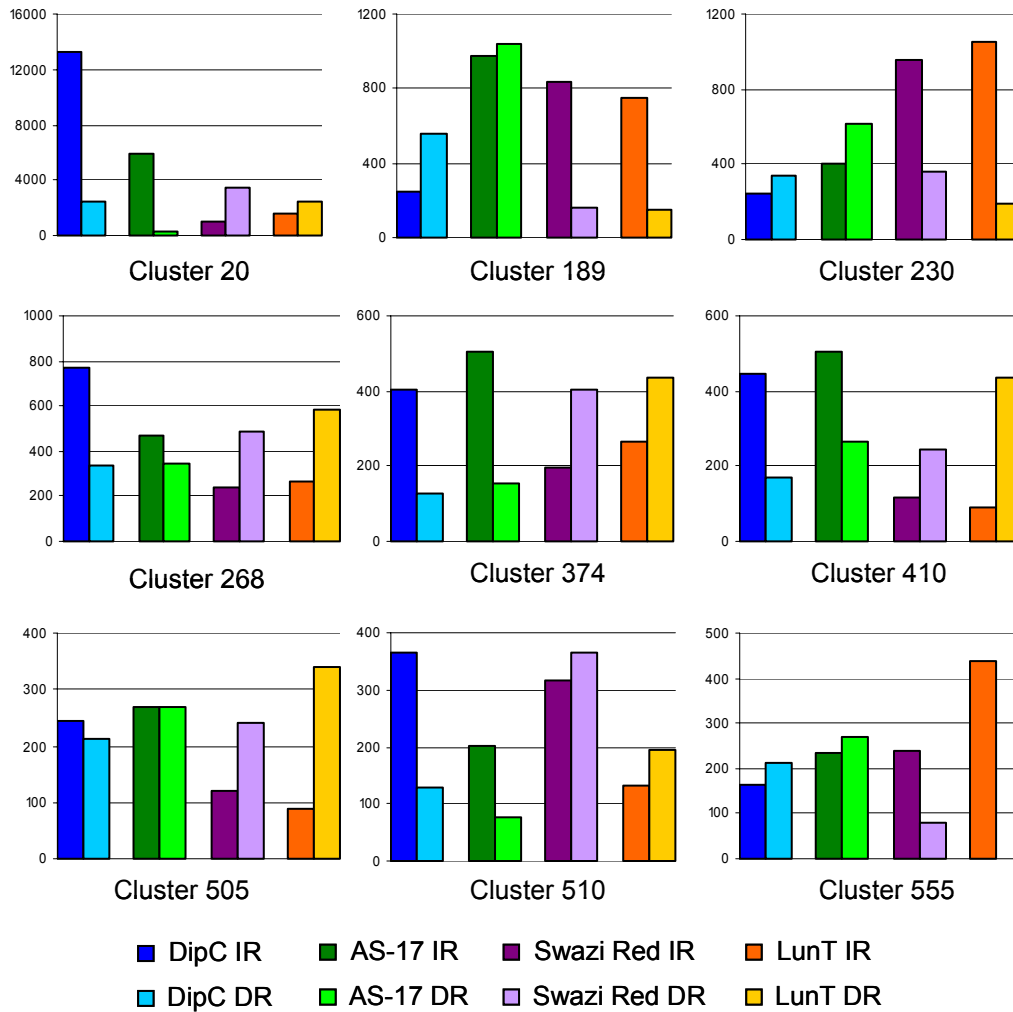


Fig. 13. Bambara groundnut genes with differential expression profiles between drought-adapted (DipC, AS-17) and non-adapted (Swazi Red, LunT) genotypes after water-deficit treatment. Values on the y-axis indicate the number of transcripts expressed as TPM. Cluster 20: stored cotyledon mRNA/defensin; cluster 189: tRNA-Val gene; cluster 230: PSII protein psbY; cluster 268: protein of ubiquinol-cytochrome c reductase complex; cluster 374: ribosomal protein S17; cluster 410: heat shock protein 23.9; cluster 505: unknown function; cluster 510: unknown function; cluster 555: transcription factor MYB123. IR: gene expression under non-limiting irrigation, DR: gene expression after one week of water-deficit.

In addition to qualitative differences in gene expression, i.e. induction versus repression, MPSS data were also analysed for quantitative differences. Searching for genes generally being induced upon water-deficit stress and showing at least twice the expression in DipC and AS-17 compared to the non-adapted pair resulted in two clusters. Cluster 433, showing homology to a ribosomal protein from *Arabidopsis*, was expressed 2.3 and 1.7-times more in the challenged DipC and AS-17 variants, respectively, compared to Swazi Red and LunT while induction factors were comparable. The other transcript tag was similar to an EST from soybean without annotation and its expression in stressed DipC and AS-17 was 2.3 and 5.8-times higher than in LunT. More quantitative differences were obvious in repressed clusters.

Out of 85 genes with at least double expression in the drought-adapted genotypes, 15 are metabolism-related (usually connected with CO₂ fixation) and 28 fall into the ‘energy’ category (Fig. 12c). The averaged repression factors for all repressed energy-related genes reveal clear differences. In AS-17, these genes were downregulated 5.3-fold, followed by DipC (7.1-fold) and Swazi Red (13.4-fold). Repression in LunT was most pronounced with 22.4-fold. A similar result was observed in the upregulated defence-related clusters. While the average induction in DipC and AS-17 was relatively moderate (12.9% and 16.0%, respectively), factors increased to 62.2% for Swazi Red and climaxed in LunT with 92.4%.

3.2.2 Microarrays

3.2.2.1 Signal linearity

Defined quantities (5, 50 and 500pg) of three different mRNA spike-in controls were added to each total RNA population prior to the cDNA transcription and labelling reactions. Arithmetic means of 60 hybridisation experiments showed a regression line with an R^2 value of 1 and moderate standard deviations (coefficients of variance less than 50%; Fig. 14). Due to the linear correlation between transcript quantity and hybridisation signals, values of Alien spike-in mRNA 2 were used to normalise data of all genes.

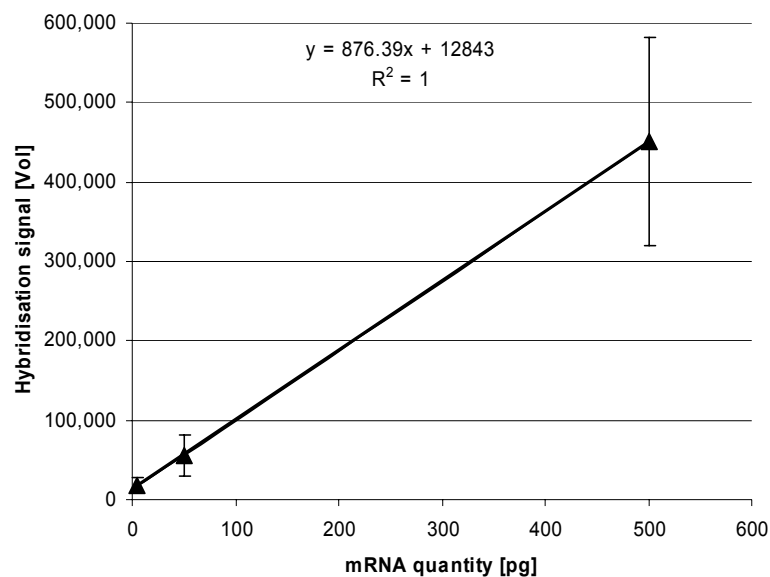


Fig. 14. Regression of three spike-in control mRNA quantities against hybridisation signals. Vertical bars represent standard deviations from 60 hybridisation experiments.

3.2.2.2 Validation of MPSS

Validation of a subset of genes identified through MPSS was carried out by means of a small custom-made microarray. In contrast to MPSS expression profiling, which provides digital representations of transcript abundance, microarrays rely on differences in hybridisation intensity between a pair of samples, in this case bambara groundnut leaves grown under non-limiting irrigation against water-deficit stress. Oligonucleotide probes were synthesised for 132 genes based on consensus sequences of the respective MPSS clusters. Beside technical feasibility, i.e. adequate read length, selection criteria included a broad coverage of the functional categories identified, interesting expression profiles and ideally landrace-specific differences, and evidence of stress relation in the literature as well as potentially novel genes without annotation.

As the original mRNA populations, from which the MPSS cDNA libraries had been prepared, were not available in quantities sufficient for microarray analysis, leaves from different but phenotypically highly similar plants from the same CE experiment were utilised. Replication of hybridisation experiments was not carried out for the same reasons.

In general, comparison of \log_2 ratios for both technologies (Table 18, Appendix) showed good correlation between MPSS and microarray data for highly to medium expressed genes. When data points are classified into three groups – upregulated at a \log_2 ratio ≥ 2 , downregulated at a ratio ≤ -2 and insignificantly altered in between – the 89 most highly expressed genes on the array, i.e. up to MPSS cluster number 195, showed agreement for 304 out of 354 data points (85.9%) between the two technologies. Fifty of these genes produced matches for all four genotypes investigated. Discrepancies were mainly of quantitative nature. For example, according to MPSS, three out of four genotypes displayed significant induction of cluster 4. Using microarrays, values were shifted downwards so that only LunT remained with significant upregulation. Similar parallel changes, upwards or downwards, occurred in clusters 12, 63, 133 or 168. However, there also were between-genotype differences. While according to MPSS results, cluster 33 was strongly downregulated in DipC only, no significant landrace disparity was evident in the hybridisation-based approach. Clusters 111 and 177 exhibited MPSS \log_2 ratios in a similar range for all genotypes whereas microarrays suggested a far greater induction in the genotypes not adapted to drought. A clear discrepancy of technologies was only observed with cluster 110. Expression profiles were inconsistent for the four genotypes using MPSS, but microarrays showed significant downregulation throughout the panel.

Data agreement drastically decreased in clusters with weaker expression. Cluster 209 is composed of 111 transcript tags within the total cleaned MPSS library. This corresponds to 562 transcripts per million. From this point on, technical accordance was only given for 62.5% of the data points investigated. While the set of genes chosen from the MPSS analysis was enriched for putative genotypic differences, these often disappeared using microarrays. For example, MPSS revealed slight induction in cluster 230 for DipC and AS-17 and repression in Swazi Red and LunT (Fig. 13). Microarrays, however, did not detect any differences beyond the significance threshold. Cluster 323 was presumed to be upregulated in DipC and slightly repressed in the other genotypes. The microarray approach indicated downregulation in DipC (just under the significance threshold) and strong repression in AS-17, Swazi Red and LunT. In general, regarding genes of weaker expression, microarrays tended to show higher consistency between genotypes or their presumed water-deficit tolerance level (DipC/AS-17 vs. Swazi Red/LunT) than the 454 sequencing-based technology. Six oligonucleotide probes were designed for clusters where no expression was detected in LunT. For five of these, microarray analysis revealed clear hybridisation signals (Fig. 15). This indicates that detection through MPSS was inhibited by sequence variation in the *NotI* restriction site. However, the absence of cluster 124, for which no significant homology was found, was confirmed through missing hybridisation (Fig. 15).

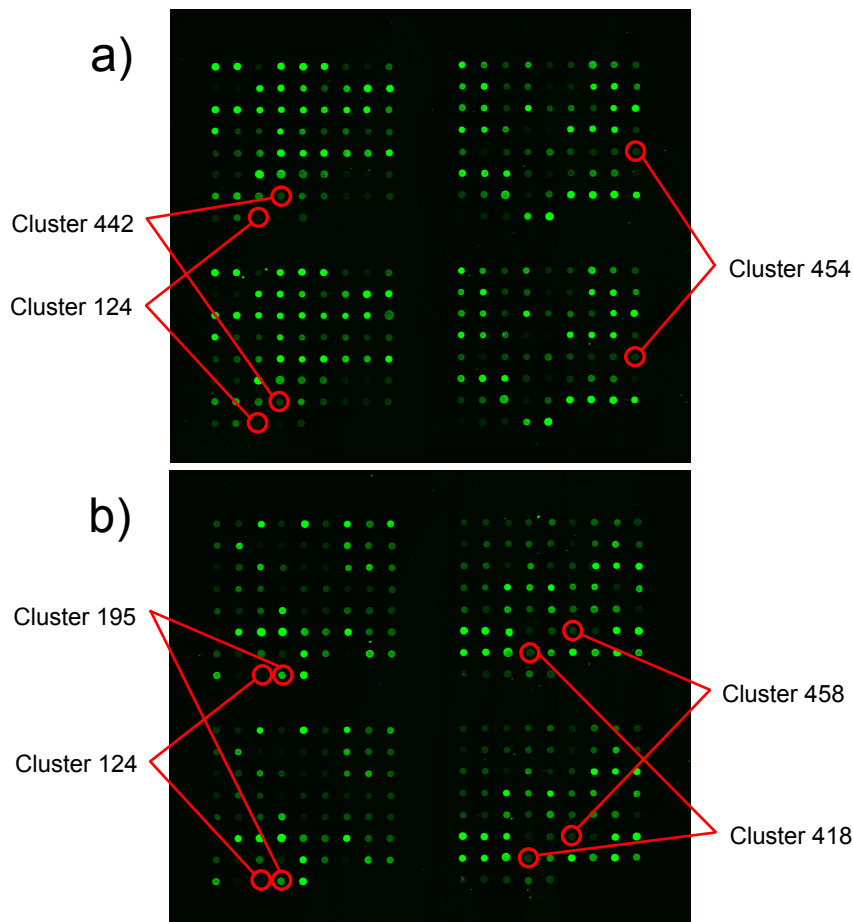


Fig. 15. Microarray hybridisations for LunT under non-water limiting conditions (a) and after seven days of water-deficit treatment (b). Genes not detected through MPSS are marked by circles. Clusters 195, 418, 442, 454 and 458 show a hybridisation signal in at least one variant, whereas cluster 124 does not.

3.2.3 Time course experiment

3.2.3.1 Physiological studies

In order to assess the significance of the second CE experiment, which was conducted at a temperature lower by 2°C than for the MPSS gene expression profiling experiment, and also to support the gene expression data obtained from it, various physiological data were collected.

Differences between both treatments and landraces already began to show on a visual level. After two days under reduced irrigation, DipC leaves tended to change their orientation towards the light source. Leaf angles slightly steepened possibly resulting in decreased light interception, a phenomenon described as paraheliotropism. Previous observation in the first CE experiment had shown that AS-17, for example, exhibits this phenomenon to a much

greater extent. LunT did not display any signs of paraheliotropism (Fig. 16). Instead, LunT leaves decreased in turgescence, which was not apparent in DipC.

One day later, after three days under limiting conditions, older LunT leaves showed first signs of necrotic lesions in the leaf edges (Fig. 17). Young leaves were not affected. DipC plants were not damaged at all. This effect continued until the end of the stress phase with the three or four oldest leaves dying off in LunT, whereas DipC leaves remained free from damage.



Fig. 16. Slight evidence of paraheliotropism in DipC after reducing irrigation for two days (left).



Fig. 17. Wilted old leaves in LunT after reducing irrigation for eight days.

At the end of the water-deficit treatment, i.e. after eight and nine days under water-limiting conditions, single light-exposed leaflets were detached and allowed to dry under the same experimental conditions. Developing photosynthesis yields and leaf temperatures were recorded simultaneously for one hour. Regarding photosynthesis yield (Fig. 18), both landraces and treatments displayed similar curves. Freshly cut leaflets had Y values close to 0.8. LunT leaves in the drought-adapted state were characterised by an increase in chlorophyll fluorescence for the first five minutes until values of the other variants were reached. Thereafter, all curves showed a constant decrease until measurements were terminated. After around 18 minutes, treatments diverged with the untreated variants losing in Y more rapidly than plants which had been subjected to water-deficit before. While there was no landrace-specific difference between stressed leaves, unstressed DipC tended to maintain a slightly higher chlorophyll fluorescence level than LunT towards the end of the experiment.

Differences between the landraces were also reflected in leaf temperatures (Fig. 18). In the fully irrigated controls of both landraces, leaf temperatures decreased immediately upon cutting. After around five minutes, this effect inverted with drastically increasing temperatures until around 20 minutes after cutting and low slopes thereafter. In contrast, drought-challenged LunT leaves did not show such a curve. Temperatures constantly rose throughout the study. Values for stressed DipC developed between these two types. The effect of declining temperature early upon cutting was clearly visible, however to a lower extent as in unstressed reference leaves.

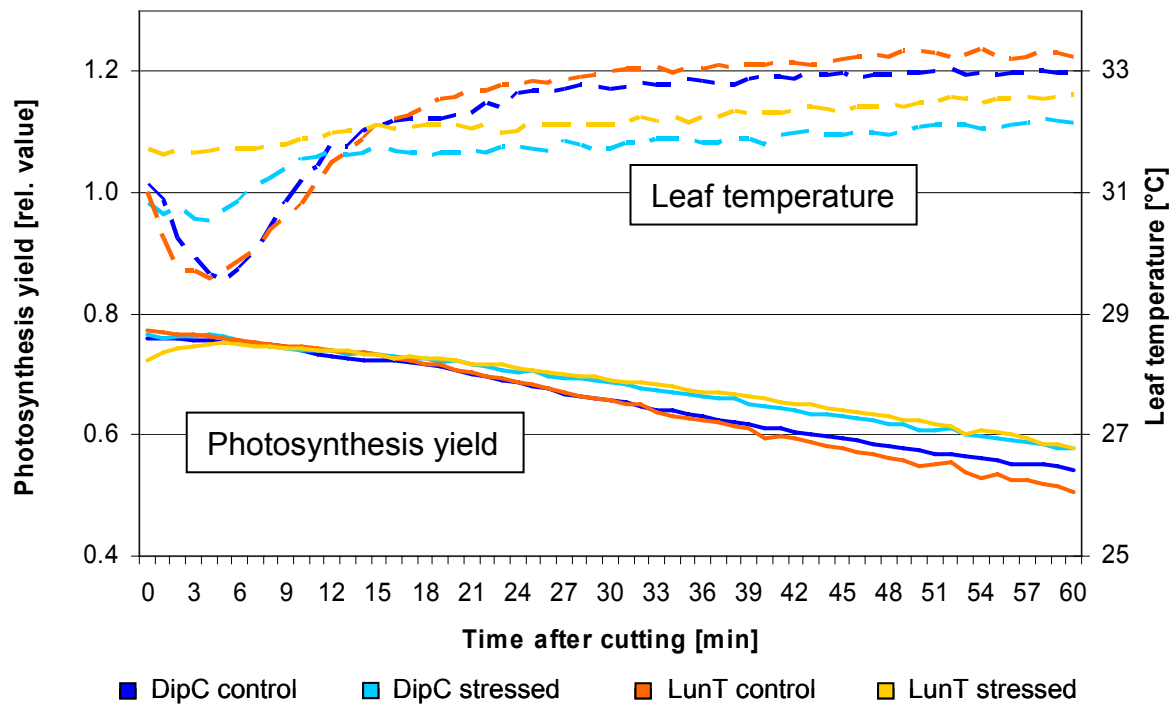


Fig. 18. Photosynthesis yields (solid lines) and leaf temperatures (dashed lines) in drying bambara groundnut leaflets from plants grown for eight or nine days under water-limiting and non-limiting conditions. Lines represent the means of three replicate measurements each.

Nine days after initiating the water-deficit treatment, leaves were also analysed for osmotic adjustment. The data (expressed as osmolality; Table 7) suggest a slight but insignificant increase in the concentration of osmotically active solutes in both landraces under water-limiting conditions and comparable levels between them. Assuming lower relative water content in stressed leaves, which was not determined, it seems obvious that osmotic adjustment has not occurred under the above-mentioned conditions.

Regarding the number of mature seeds harvested (Table 7), a wide difference in yield potentials between the landraces is striking. Under full irrigation, DipC produced 2.6-times the seeds of LunT ($p=0.00002$). Upon water-deficit treatment, the average seed number was reduced by 40% in LunT. However, these changes did not stand the statistical test. Two plants produced ten and twelve seeds, respectively, while the other plants yielded between one and six seeds. Without these outliers, the yield reduction is statistically significant ($p=0.0036$). While in DipC, seed number tended to slightly increase upon water-deficit treatment, a significant reduction in seed weight by 9% was recorded. No change was observed for LunT.

Table 7. Osmolality (mean of three replications \pm standard deviations; determined after nine days of reduced irrigation), seed number and seed weight (mean of eight replications \pm standard deviations; determined at maturity) of the two bambara groundnut landraces in the second CE water-deficit experiment). ns: not significant, *: $p \leq 0.05$ (significant)

Landrace	Treatment	Osmolality [mmol*kg ⁻¹]		Seeds per plant		Seed weight [g]	
DipC	Control	427 \pm 16	ns	23.6 \pm 4.95	ns	0.46 \pm 0.048	*
	Stress	487 \pm 67		27.1 \pm 4.51		0.42 \pm 0.019	
LunT	Control	422 \pm 80	ns	9.1 \pm 3.24	ns	0.31 \pm 0.040	ns
	Stress	500 \pm 59		5.5 \pm 3.74		0.32 \pm 0.044	

3.2.3.2 Expression kinetics of selected genes

From the second CE water-deficit stress experiment, samples at four time points during the treatment, namely one day, two days, four days and eight days after initiating the reduced irrigation regime (1RI, 2RI, 4RI and 8RI), and two time points marking stress recovery, i.e. one day and three days after re-watering to non-limiting conditions (1REC and 3REC), were collected and subjected to gene expression analysis by means of the microarray presented above (3.2.2.1 and 3.2.2.2). For the treated variant, all samples (three replications) were analysed using independent microarray hybridisations. The same was done for the untreated variant at stages 1RI, 4RI and 1REC. For time points 2RI, 8RI and 3REC, only one sample was tested each. As no major differences to surrounding time points were obvious, data from nearest time points with three biological replications analysed, i.e. 1RI for 2RI (one day difference) and 1REC for 8RI and 3REC (two days difference each), were used as references for reason of statistical evaluation. In the following two chapters, expression patterns of 132 genes and the two bambara groundnut landraces DipC and LunT are presented as log₂ ratios of the challenged variants against their respective controls.

3.2.3.2.1 Hierarchical cluster analysis

Figure 19 shows the results of gene expression analysis obtained from the 2008 CE water-deficit experiment in the form of a hierarchically clustered heatmap. Here, data from both landraces were concatenated to produce gene expression curves including different genotypes and thus investigate patterns of gene expression with a focus on functional grouping. The intensities of green and red colours symbolise the degrees of down- or upregulation, respectively.

Cluster A is characterised by relatively strong repression starting two days after beginning the stress treatment. In DipC, this effect climaxed at 4RI, declined thereafter and disappeared in the recovery stages. Only ferritin was still repressed in 1REC. Expression profiles in LunT were less clear. Maximum downregulation of genes varied between 2RI and 8RI. However, the data point towards continuation of differential gene expression in the first recovery stage investigated. Beside two genes with unknown function, member of this cluster play a role in defence (Cu/Zn-superoxide dismutase and ferritin), ethylene biosynthesis (1-aminocyclopropane 1-carboxylic acid oxidase), lignin biosynthesis (caffeic acid methyltransferase), signalling (UDP-glucuronosyltransferase) and RNA-binding.

A second downregulated group (cluster B) was identified with genes being less repressed than those in cluster A. DipC tended to show an earlier response to water-deficit than LunT. For a sub-group comprising nine genes, expression levels returned to the values of unstressed control plants after the second stress stage while LunT displayed prolonged downregulation until 4RI. Most of the genes in cluster B (12 of 17) are related to the processes of photosynthesis and carbon fixation. Other genes with similar expression kinetics encode for a cinnamyl alcohol dehydrogenase-like protein, which is involved in lignin biosynthesis, the CPRD86 protein with significant homology to a soybean KS-type dehydrin (e^{-64}), a glutamine synthetase, a CDPK-related protein kinase and an EST from *Phaseolus vulgaris* homologous to a suppressor of CONSTANS mRNA ($4e^{-141}$).

Genes in cluster C showed early, short-term induction upon reduced irrigation in both landraces. After 2RI however, expression levels fell below those of plants grown under non-limiting conditions, which was still visible in the recovery stages. While this occurred rapidly in DipC with maximum repression at stage 3RI, LunT reacted more slowly and had least expression values one day after full rewatering. Many genes in this cluster are involved in the synthesis, modification or degradation of proteins (translation initiation factor, protein phosphatase, serine hydroxymethyltransferase, ubiquitin-like protein, cysteine proteinase and peptidase). Another group of genes is related to defending mechanisms, such as pathogenesis-related protein 4.2, type 1 metallothionein, heat shock protein-associated protein, glutathione peroxidase and catalase.

Cluster D comprises genes which were strongly upregulated in response to water-deficit. While in LunT, these genes were induced throughout the stress phase followed by rapid adjustment to values of the unstressed variant during recovery, DipC tended to show maximum upregulation not before two days after initiating the treatment and dropped in overexpression in the course of stress acclimatisation, i.e. at stage 8RI. Among functionally classified transcripts, those belonging to the “cell rescue, defence and virulence” group are dominating. One cluster showed significant sequence homology to alcohol dehydrogenase 1, one lipid transfer protein mRNA was identified, and two clusters are highly similar to cowpea LEA proteins (LEA5 and CPRD22, which represents a dehydrin). Two heat shock protein mRNAs are outgrouped due to a lower degree of upregulation, but expression patterns basically follow those of cluster D.

Another cluster with induced gene expression emerged (cluster E). In contrast to cluster D, induction factors decreased after two days under water-deficit conditions. Again, the reaction is more pronounced in the putatively drought-sensitive landrace which also displayed clear downregulation in a sub-group of genes after restoring full irrigation. Classified genes can be divided into three functional classes: Four transcript clusters are assumed to have direct dehydration-protective function (Snakin-like cysteine-rich protein, LEA4 protein, *Phaseolus vulgaris* dehydrin and DnaJ-like protein) and four genes are assigned to protein synthesis, modification or degradation (ribosomal protein S27, polyubiquitin 2, cysteine proteinase and cyclophilin 1). Five genes are metabolism-related, with three genes being involved in sugar metabolism (invertase inhibitor, β -amylase and endo-1,4- β -mannanase) and two genes possibly functioning in secondary metabolism (*o*-methyltransferase and CPRD14, which is homologous to a cinnamyl alcohol dehydrogenase mRNA from *Malus x domestica* at an E-value of e^{-178}).

3.2.3.2.2 k-Means cluster analysis

While hierarchical cluster analysis was conducted with a merged dataset in order to give a general overview of time-dependent gene expression in the 2008 CE experiment and identify common groups of differentially expressed genes, the k-means algorithm was individually applied to the two landraces. Thus, it was intended to more precisely contrast expression profiles of DipC and LunT particularly in terms of the degree of up- and downregulation and also to gain insight into regulatory genes potentially underlying structural gene action through investigation of co-expression. Furthermore, the genes affiliated to corresponding clusters were compared in order to work out possible systematic differences in the response to water-deficit and re-irrigation, i.e. the influence of different regulatory factors.

As explained in chapter 3.2.2.2, MPSS cluster 124 was shown not to be expressed in the LunT plant used in the first CE experiment. However, possibly due to the heterogeneity of the germplasm used, one plant in the 2008 experiment displayed a hybridisation signal distinctly above the baseline signal of around 8,000 sVol. Consequently, the data for cluster 124 in LunT were omitted in k-means clustering in order not to influence results by an inconsistent expression pattern.

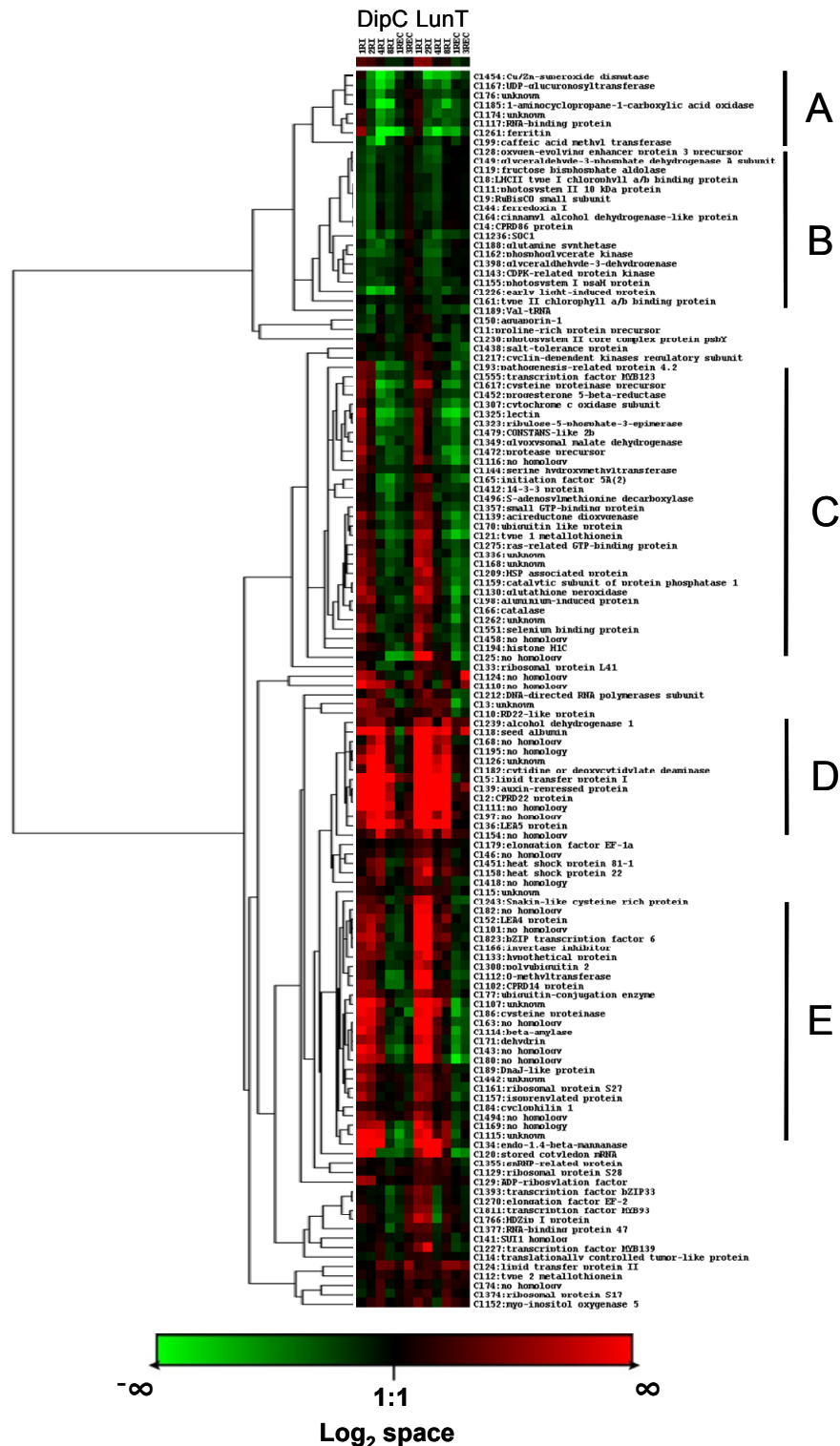


Fig. 19. Hierarchical clustering of 132 differential gene expression patterns in the second CE water-deficit stress experiment. Columns represent sampling dates in the following order (from left to right): 1RI, 2RI, 4RI, 8RI, 1REC and 3REC. The six columns on the left and right hand side reflect the landraces DipC (adapted to drought-prone areas) and LunT (adapted to a humid environment), respectively. The colour intensity of each individual data spot indicates the magnitude of the \log_2 ratio with red colour signifying higher and green colour lower transcript levels in the stressed variant in relation to the corresponding reference. Cluster designations (A-E) are explained in the text.

Log₂ ratios of all six sampling dates investigated were grouped into seven clusters (Fig. 20). According to the number of genes overlapping between the landraces DipC and LunT, clusters were compared with each other and combined, if necessary. Clusters I showed the highest congruency between the landraces with only one gene (MPSS cluster 195) being additionally included in LunT. Entries were basically the same as in the hierarchical cluster D. In DipC, these genes displayed steadily increasing expression until four days after imposing water-deficit conditions, reduced upregulation at stage 8RI and approximation to the level of reference plants one day after resuming non-limiting irrigation. Genes seem to differ in their degree of upregulation at stage 1RI. At least two genes, encoding for a putative cytidine or deoxycytidylate deaminase and a LEA5 protein, were found to be delayed in their response to dehydration. In contrast, all genes of cluster I were significantly induced (log₂ ratio greater than 2) after one day under water-deficit in LunT. Maximum upregulation was observed one day later. At stage 4RI, however, a slump was registered and a second phase of transcript accumulation occurred after eight days under stress.

Clusters II also showed good consistency between both landraces. Eleven common genes were identified out of 15 genes in DipC and 17 in LunT and mainly comprise those of hierarchical cluster E. Genotypic differences in expression trends are not as obvious as in clusters I. The primary discrepancy is quantitative in nature. While the average log₂ ratio is around 2 (corresponding to an induction factor of 4) in DipC during the first two stress stages, LunT featured values of around 3 or eightfold induction, respectively. Out of the four genes only present in DipC cluster II, two genes follow the expression pattern of LunT k-means cluster IIIb, and MPSS cluster 195 (without significant homology) is grouped under LunT cluster I. As there is a fine line between the three mentioned k-means clusters, only one gene (cluster 110) with an expression profile distinct between the two landraces remains. One gene with putative regulatory activity due to its homology (E-value $3e^{-38}$, 97% sequence identity over 91 nucleotides) to the bZIP transcription factor 6 mRNA from *Phaseolus vulgaris* was found to be co-expressed with genes of clusters II.

Twenty-three genes were grouped in DipC k-means cluster III. Twenty of these were found in two clusters in LunT, either cluster IIIa or IIIb. While the average expression profile in DipC was characterised by only slight, if at all, upregulation during the first three stress stages, these genes split into one group showing a similar marginal stress response (IIIa) and another cluster with expression patterns like those of LunT cluster I but at a lower level (IIIb). The three outliers were again assigned to clusters of similar trends. Figure 21 summarises eight genes common between DipC k-means cluster III and LunT cluster IIIb, i.e. with differential expression upon water-deficit being marginal in DipC but considerable induction in LunT. Both transcription factors in DipC cluster III, MYB139 and a transcript homologous to the soybean HDZip I protein mRNA (E-value $8e^{-69}$, 76% sequence identity over 355 nucleotides of a corresponding *Phaseolus vulgaris* EST) were co-expressed with the genes of LunT cluster IIIb.

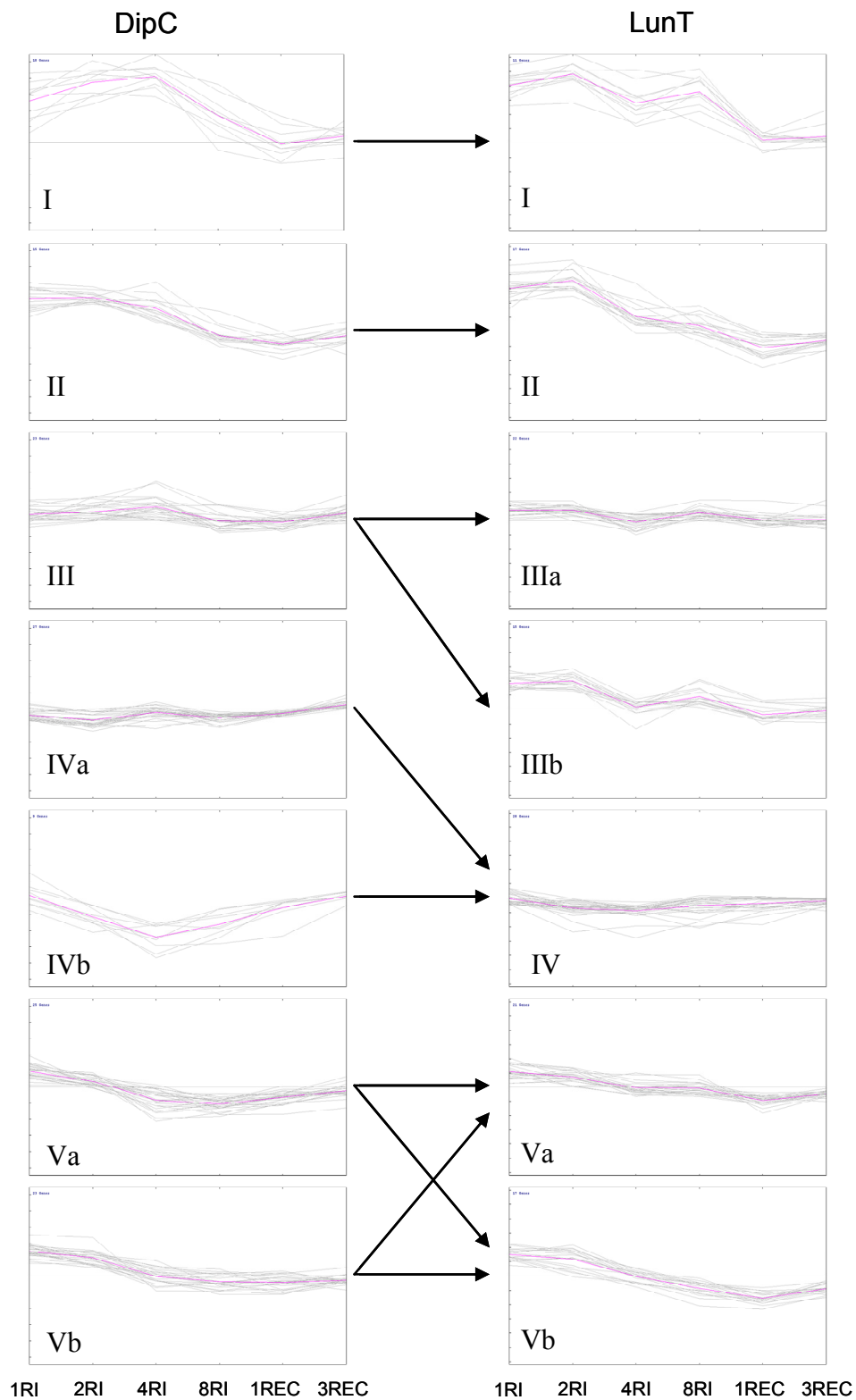


Fig. 20. k-Means clustering of 132 (DipC) and 131 (LunT) genes according to their expression kinetics over six sampling dates during the second CE water-deficit stress experiment. Arrows mark corresponding clusters. Units on the y-axis reflect \log_2 ratios of gene expression, ranging from 5 to -5 in DipC and 6 to -6 in LunT, respectively, against corresponding reference plants (horizontal line). The pink line represents the mean expression pattern of each cluster.

Genes that were more or less distinctly downregulated throughout the stress phase are grouped in clusters IV. Again, expression profiles split between the landraces. Eighteen out of 27 genes in DipC cluster IVa, which is indicated by marginal repression, and all nine genes of IVb, with downregulation being particularly pronounced at stage 4RI, are reflected in a single k-means cluster in LunT. As already observed in the hierarchical cluster analysis (Fig. 19), at least two independent regulatory systems seem to exist for the downregulation of genes in bambara groundnut. One predominantly comprehends genes of energy-related processes and showed similar expression kinetics in both landraces. Cluster IVb involves genes of different biological functions (see 3.2.3.2.1). In LunT, their response upon reduced irrigation was not distinctive enough to form an own cluster like in DipC. The suppressor of CONSTANS transcript was co-expressed with the marginally downregulated genes. Two other genes potentially acting as transcription factors were detected in DipC cluster IVa. However, the mRNA coding for transcription factor bZIP33 was grouped under cluster IIIa in LunT, and the soybean salt-tolerance protein mRNA, which is homologous to a zinc finger B-box protein mRNA from *Solanum sogarandinum* (E-value e^{-83} , 72% sequence identity over 565 nucleotides) was found in LunT cluster Va, thus indicating slight upregulation for both genes. Instead of transcription factors, cluster IVb contains genes that may function in signal transduction networks, such as UDP-glucuronosyltransferase or ethylene-mediated signalling, or posttranscriptional regulation (RNA-binding protein).

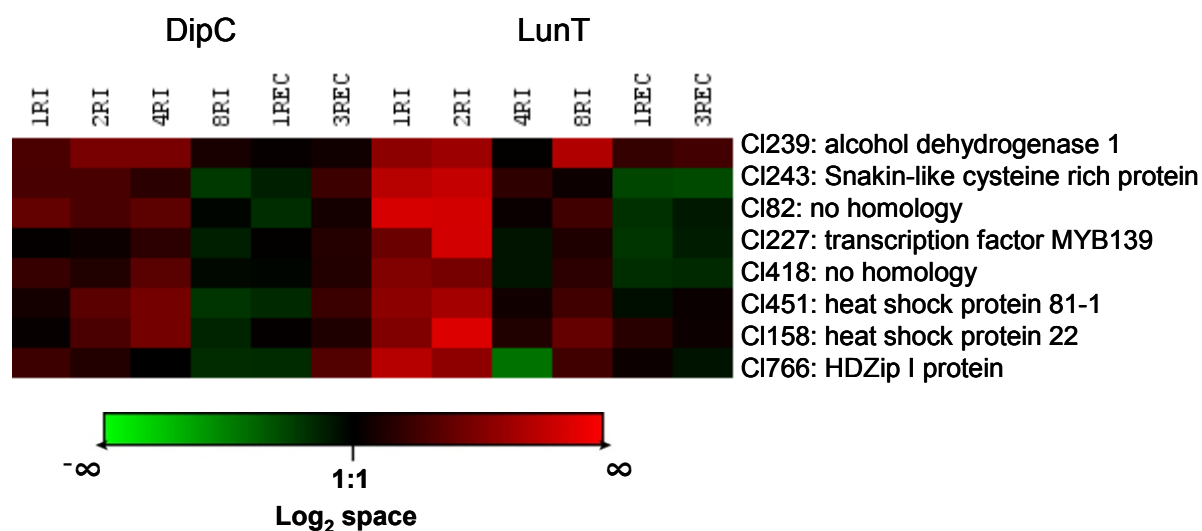


Fig. 21. Heatmap of eight genes with induction being marginal in DipC and pronounced in LunT upon reduced irrigation as revealed by k-means cluster analysis.

Clusters Va and Vb overlap with each other and do not exhibit major differences within and between the landraces. Expression profiles are characterised by moderate upregulation in the early stress stages (1RI and 2RI). Thereafter, transcript abundances in the stressed variants drop to the values of control plants or slightly fall behind. The high numbers of genes following these curves (48 in DipC and 38 in LunT) as well as their manifold functions

complicate the definition of these clusters. Two transcription factors (CONSTANS-like 2b and MYB123 are present in cluster Va both in DipC and LunT. A third one, homologous to soybean MYB93 (E-value $4e-10$, 91% sequence identity over 61 nucleotides) was clustered under DipC Va and LunT IIIa.

3.2.3.3 Within-time point analysis

3.2.3.3.1 Differential gene expression between treatments

For each time point and gene printed on the microarrays, mean expression differences of \log_2 transformed raw data were compared using Student's t-tests. Data obtained from all three replications of the water-deficit treated variants and corresponding reference plants entered the analysis. Differences were regarded significant if p-values were ≤ 0.05 (significant) or ≤ 0.01 (highly significant). The following tables list genes that featured at least one significant delogarithmised induction or repression factor for either landrace above 2 or below -2, respectively, in the course of the second CE water-deficit stress experiment.

Twelve differentially expressed genes are metabolism-related (Table 8). All three genes assigned to sugar metabolism were upregulated following reduced irrigation which is congruent with the results obtained through MPSS expression profiling after one week of more severe dehydration stress. The highest induction factors were observed for the endo-1,4- β -mannanase gene at two days after initiating the treatment in both landraces. Strong transcript accumulation was also present at stages 1RI and 4RI, however not significant in LunT. The mRNA coding for a β -amylase was significantly upregulated in both landraces at the first two stress stages investigated while the induction of invertase inhibitor mRNA was only significant in LunT at stage 2RI.

Two genes playing a role in the lignin biosynthetic pathway responded differently to the treatment. Caffeic acid methyltransferase was significantly downregulated in DipC at stage 4RI, while the expression of CPRD14 (cinnamyl alcohol dehydrogenase) was slightly induced in DipC at the first stress stage and slightly repressed after eight days under water-deficit. LunT showed stronger upregulation, however only significant at stage 2RI.

One gene was differentially expressed after the full irrigation regime had been resumed. Ribulose 5-phosphate 3-epimerase was significantly downregulated in LunT at 1REC, while its expression tended to be induced early in the experiment.

Table 8. Metabolism-related genes with statistically significant differential expression and minimum twofold change (delogarithmised induction or repression factors ≥ 2 or ≤ -2 , respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
Endo-1,4- β -mannanase (34)	DipC	8.3*	33.4**	17.1*	-1.4	-2.4	-1.9
	LunT	7.3	46.7**	9.7	2.5	-1.4	-1.2
Glyceraldehyde 3-phosphate dehydrogenase (49)	DipC	-1.5	-2.1*	-1.3	-1.4	-1.2	1.1
	LunT	-1.3	-1.7	-1.9	-1.1	-1.0	-1.0
Caffeic acid methyltransferase (99)	DipC	-1.3	-2.9	-5.9**	-1.6	-1.1	1.3
	LunT	-1.0	-1.5	-1.6	-1.0	-1.6	1.0
CPRD14 (cinnamyl alcohol dehydrogenase) (102)	DipC	2.9*	2.9	-1.2	-2.2*	-2.2	-1.2
	LunT	11.8	20.9**	1.4	1.5	-1.5	-1.3
<i>o</i> -Methyltransferase (112)	DipC	2.1	1.7	-1.1	-2.4	-2.2	1.2
	LunT	6.9	9.1**	-1.1	-1.1	-1.9	-1.7
β -Amylase (114)	DipC	6.8*	4.1*	1.5	-1.2	-1.8	-1.1
	LunT	6.3**	7.5**	2.1	1.2	-3.2	-1.7
Invertase inhibitor (166)	DipC	2.3	3.5	1.9	1.4	-1.4	-1.3
	LunT	4.9	8.4**	1.4	1.6	-1.3	-1.1
Cytidine or deoxycytidylate deaminase (182)	DipC	1.5	7.5*	16.4**	2.3	-1.3	1.2
	LunT	8.5	21.5**	5.1**	10.8**	1.4**	1.2
1-Aminocyclopropane 1-carboxylic acid oxidase (185)	DipC	-1.0	-2.3	-12.7**	-5.4*	-1.2	1.1
	LunT	1.5	-1.7	-2.1	-4.3*	-1.9	-1.3
Glutamine synthetase (188)	DipC	-1.2	-1.8	-2.3**	-1.4*	-1.0	1.2
	LunT	-1.1	-1.8	-1.8	-1.2*	1.1	-1.0
Ribulose 5-phosphate 3-epimerase (323)	DipC	2.8	1.2	-3.1*	-2.5	-1.6	-1.3
	LunT	2.2	1.2	-1.4	-2.4	-3.3*	-1.4
S-adenosylmethionine decarboxylase (496)	DipC	1.2	1.1	-1.4	-2.5*	-1.5	-1.0
	LunT	2.1	1.9	-1.6	-1.1	-1.3	-1.3

A high proportion (seven out of twelve genes analysed) of differentially expressed genes was also detected among the ‘transcription’ functional category (Table 9). Most transcription factors (MYB139, CONSTANS-like 2b, HDZip I, MYB93, bZIP6) displayed similar expression characteristics. Significant upregulation was only observed in LunT one and/or two days after reducing the water dose. bZIP6 was the most strongly induced gene with a 14.1-fold change at stage 2RI in LunT, whereas in DipC, upregulation was not significant. The suppressor of CONSTANS 1 gene was the only transcript downregulated during the stress phase. Although to a low degree, repression was significant at later stages of the stress experiment (4RI and 8RI in DipC, 4RI in LunT).

Table 9. Transcription-related genes with statistically significant differential expression and minimum twofold change (delogarithmised induction or repression factors ≥ 2 or ≤ -2 , respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
DNA-directed RNA polymerases subunit (212)	DipC	1.1	1.7	2.1*	-1.7	-1.5	-1.0
	LunT	1.7*	2.5**	1.2	1.3	-1.9*	-1.6*
Transcription factor MYB139 (227)	DipC	1.1	1.2	1.5	-1.3	1.0	1.4
	LunT	3.0*	5.6**	-1.2	1.3	-1.5	-1.3
CONSTANS-like 2b (479)	DipC	2.7	1.3	-1.8	-1.6	-1.6	-1.4
	LunT	2.7*	1.0	-1.5	-1.6	-2.3	-1.5
HDZip I protein (766)	DipC	1.7	1.3	1.0	-1.4	-1.4	2.0
	LunT	4.0	3.2*	-2.6	1.7	1.1	-1.1
Transcription factor MYB93 (811)	DipC	1.7	1.3	-1.1	-1.7	-1.3	1.6
	LunT	2.7	2.6*	-2.0	1.2	1.3	-1.2
bZIP transcription factor 6 (823)	DipC	3.6	4.0	2.5	-1.2	-1.7	1.1
	LunT	7.0*	14.1**	2.0	2.3	-1.2	-1.3
Suppressor of CONSTANS 1 (1236)	DipC	-1.6	-2.5	-1.5**	-2.2*	-1.3	-1.4
	LunT	1.4	-1.6	-2.3*	-1.5	-1.6	-1.3

Among 18 differentially expressed genes functioning in ‘cell rescue, defence and virulence’ (Table 10), one transcript, lipid transfer protein I, was identified which was significantly upregulated at all data points during the stress phase. The same applies for CPRD22 with one exception at stage 4RI in LunT. In general, most significances were observed after two days under water-limiting conditions with 19 of 36 data points displaying a p-value below 0.05. Thirteen genes were significantly differentially expressed in LunT at this stage, and induction or repression factors were generally higher than in DipC. Four genes were significantly repressed in LunT during the recovery phase, either at 1REC only (pathogenesis-related protein 4.2 and ferritin) or both at 1REC and 3REC (type 1 metallothionein and catalase) while no significant differences between treatments were evident in DipC. Qualitative changes in gene expression (induction or repression) were widely consistent with the MPSS-derived data in the first CE water-deficit stress experiment. However, two differences between the experiments were apparent. In the 2006 experiment, MPSS cluster 20 (stored cotyledon mRNA/defensin) was downregulated 5.5-fold in DipC and upregulated 1.5-fold in LunT after one week of water-deficit stress, while the second CE experiment resulted in significant induction in DipC at the first two stress stages and no significant differences afterwards and strong induction in LunT at stage 2RI. Discrepancy was even more prominent with MPSS cluster 261 (ferritin) where more severe stress led to transcript accumulation by factors of 5.2 and 29.9 for DipC and LunT, respectively. The 2008 experiment, however, yielded significantly repressed expression of this gene in the later stress stages in both landraces.

Table 10. Defence-related genes with statistically significant differential expression and minimum twofold change (delogarithmised induction or repression factors ≥ 2 or ≤ -2 , respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
CPRD22 protein (dehydrin) (2)	DipC	9.7*	14.3*	19.3*	3.7**	-1.1	1.4
	LunT	19.4*	45.7**	9.5	19.8**	1.4	1.3
Lipid transfer protein I (5)	DipC	19.6**	22.1**	24.0**	12.2**	2.2	1.7
	LunT	20.6**	33.7**	35.1**	22.7**	-1.6	1.3
Stored cotyledon mRNA (defensin) (20)	DipC	5.8*	5.2*	-1.9	-1.9	-2.1	-1.5
	LunT	2.7	28.0*	10.3	1.3	-3.2	-2.6
Type 1 metallothionein (21)	DipC	2.2	1.7	-1.7	-2.3	-2.0	-1.5
	LunT	3.1	3.9	1.1	-2.3	-4.2*	-2.8*
LEA5 protein (36)	DipC	2.5	16.8	44.2**	9.1**	3.1	1.4
	LunT	6.5	30.2**	6.0	9.6*	1.0	1.0
LEA4 protein (52)	DipC	2.7	3.3	2.4	-1.4	-1.6	1.1
	LunT	12.8	20.4**	1.4	2.8*	-1.0	-1.2
Catalase (66)	DipC	1.7	1.6	-1.0	-1.3	-1.5	-1.3
	LunT	2.0	1.9	-1.0	-1.2	-2.2**	-1.5*
Dehydrin (71)	DipC	3.0	3.5*	1.6	-1.5	-1.7	1.0
	LunT	4.5**	5.7**	1.6	-1.2	-2.7	-1.7
DnaJ-like protein (89)	DipC	2.4*	2.1*	1.1	1.1	-1.1	-1.1
	LunT	2.0	2.0	1.1	1.7	-1.2	-1.2
Pathogenesis-related protein 4.2 (93)	DipC	2.2	1.9	-2.9	-2.1	-2.2	-1.3
	LunT	1.9	1.1	1.2	1.0	-2.9**	-1.6
Glutathione peroxidase (130)	DipC	3.7*	2.2*	-1.2	-1.4	-1.3	-1.2
	LunT	4.2*	3.4*	1.4	-1.5	-3.7	-2.0
Heat shock protein 22 (158)	DipC	1.1	1.9	2.6**	-1.3	1.1	1.3
	LunT	2.5	6.2**	1.3	2.3	1.4	1.1
HSP associated protein (209)	DipC	2.4	2.0	-1.2	-1.9	-1.4	-1.3
	LunT	2.6	4.5*	1.3	-1.6	-2.0	-1.9
Alcohol dehydrogenase (239)	DipC	1.9	2.7	2.7	1.2	1.1	1.2
	LunT	3.1*	3.6*	1.0	4.3*	1.5	1.8
Snakin-like cysteine rich protein (243)	DipC	1.8	1.8	1.4	-1.6	-1.3	-1.7
	LunT	3.9*	5.0*	1.5	1.1	-1.8	-1.8
Ferritin (261)	DipC	3.0	-1.6	-7.5*	-6.9*	-5.0	-1.4
	LunT	1.1	-3.2	-7.0	-3.2*	-3.6*	-1.4
Heat shock protein 81-1 (451)	DipC	1.2	2.1	2.6*	-1.5	-1.4	1.6
	LunT	3.1*	3.9**	1.2	1.7	-1.1	1.1
Cu/Zn-superoxide dismutase (454)	DipC	1.3	-2.8	-10.9**	-4.3	-1.8	-1.0
	LunT	-1.1	-5.1*	-3.8	-3.9	-2.3	-1.5

In terms of significant expression differences, functional categories other than metabolism, transcription and cell rescue, defence and virulence were clearly underrepresented (Table 11). Only two energy-related genes were found to be differentially expressed. The early light-induced protein gene was downregulated at stages 2RI and 4RI in DipC and at stage 8RI in LunT. The glyoxysomal malate dehydrogenase gene was slightly induced at the earliest stress stage investigated in LunT and showed significant downregulation after rehydration. No significant change was observed for DipC in both experiments. The other nine genes of the

'energy' functional category chosen for microarray analysis were not differentially expressed beyond the significance level at any sampling date.

Four genes showed significant but slight upregulation in the early stress stages in LunT only. These encoded for histone H1C, cyclin-dependent kinases regulatory subunit, both grouped under 'cell cycle and DNA processing', polyubiquitin 2 (protein fate) and ras-related GTP-binding protein (transport). Major changes were observed for the cysteine proteinase gene with significant water-deficit-induced expression in both landraces (stages 1RI and 2RI in DipC and 1RI until 4RI in LunT) and repression in the first recovery stage in LunT. The ADP-ribosylation factor gene, classified as signalling-related, was upregulated in DipC but not in LunT, thus tending to confirm the data from the first water-deficit experiment.

Downregulation of genes not functioning in photosynthesis and carbon fixation was more pronounced in the drought-adapted landrace DipC. This became apparent with the RNA-binding protein and lectin genes, both exhibiting binding function, and the UDP-glucuronosyl-transferase gene with putative signalling function. Four days after imposing water-limiting conditions, repression of these genes was only significant in DipC.

Similarly to the data mentioned above, most significant changes upon water-deficit stress were found in the two early stages of the experiment in the set of unclassified genes (Table 12). While both landraces showed comparable numbers of differentially expressed genes at these time points (16 vs. 13 and 17 vs. 22 at stages 1RI and 2RI in DipC and LunT, respectively), the intensities of upregulation were generally higher in the non-adapted landrace. Divergence in the number of significances was, however, observed from stage 4RI until 1REC. At four days under water-limiting conditions, 17 genes were found to be induced in DipC whereas this was the case for only three LunT genes. This situation is also reflected in LunT k-means clusters I and IIIb (see 3.2.3.2.2) where a downward bend is visible at this stage. The last stress stage investigated showed an opposite reaction. Only one gene was upregulated in DipC and seven in LunT. At the first rehydration stage, no significant difference to the control treatment was found in DipC in contrast to ten genes in LunT. One gene (MPSS cluster 3) was induced while no change was detected during the stress phase. On the other hand, eight genes were significantly downregulated while these were upregulated at least under water-deficit stress, usually during stages 1RI and 2RI.

In most cases, the microarray data of the second CE water-deficit experiment confirmed the results obtained from the initial one. MPSS cluster 25 is an exception. Strong induction was detected for both landraces in the first experiment, whereas the succeeding study led to upregulation in LunT only. Cluster 63 showed downregulation in DipC and upregulation in LunT in the 2006 experiment but no landrace-specific differences in 2008. Cluster 336 was also suspected to be expressed differentially between landraces according to MPSS. However, neither microarray analysis of both experiments supported this assumption.

Table 11. Genes of underrepresented functional categories with statistically significant differential expression and minimum twofold change (delogarithmised induction or repression factors ≥ 2 or ≤ -2 , respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
Energy							
Early light-induced protein (226)	DipC	-1.7	-4.3*	-3.1*	-3.6	-1.2	1.3
	LunT	-1.2	-1.6	-1.5	-2.0*	-1.3	-1.3
Glyoxysomal malate dehydrogenase (349)	DipC	2.5	1.2	-2.6	-1.8	-1.2	-1.4
	LunT	2.3*	1.0	-1.5	-1.5	-2.3*	-1.4
Cell cycle and DNA processing							
Histone H1C (194)	DipC	1.9	1.4	1.2	-1.4	-1.9	-1.2
	LunT	3.2*	1.3	-1.1	1.2	-2.5	-1.5
Cyclin-dependent kinases regulatory subunit (217)	DipC	1.1	-1.3	-1.2	-1.6	-1.1	1.2
	LunT	2.2*	1.5	-1.3	-1.3	-1.6	-1.8
Protein fate							
Ubiquitin-conjugation enzyme (77)	DipC	2.7*	2.5*	1.6	1.0	-1.3	-1.4
	LunT	2.9*	2.7	1.4	1.5	-1.4	1.0
Cysteine proteinase (86)	DipC	6.0**	5.2**	2.2	-1.3	-1.1	-2.6
	LunT	4.4*	14.2**	3.8*	-1.1	-3.7*	-1.9
Polyubiquitin 2 (300)	DipC	2.1	1.7	1.0	-1.4	-1.4	1.1
	LunT	3.8	3.3**	-1.0	1.4	-1.6	-1.4
Protein with binding function							
RNA-binding protein (117)	DipC	1.4	-1.4	-3.0**	-2.0	-1.2	1.1
	LunT	1.4	-1.8	-1.5	-2.0	-1.7*	-1.1
Lectin (325)	DipC	3.8	1.2	-3.8*	-3.6	-2.0	-1.7
	LunT	2.0	1.2	-1.7	-4.3	-5.0	-2.6
Selenium binding protein (551)	DipC	3.4*	2.1	-1.6	-1.3	-1.0	-1.2
	LunT	3.4*	2.3*	-1.4	-1.2	-1.8	-1.2
Transport							
Ras-related GTP-binding protein (275)	DipC	1.7	1.5	1.1	-1.9	-1.5	1.0
	LunT	3.0*	3.0**	-1.0	-1.5	-1.8	-1.3
Cellular communication and signal transduction mechanism							
ADP-ribosylation factor (29)	DipC	3.2*	2.9*	1.0	1.2	-1.0	1.0
	LunT	1.5	1.6	1.3	1.2	1.0	1.2
UDP-glucuronosyl-transferase (167)	DipC	-1.3	-2.6	-5.2**	-3.4*	-1.2	1.1
	LunT	-1.4	-2.3	-1.9	-2.9*	-1.1	-1.3

Table 12. Unclassified genes with statistically significant differential expression and minimum twofold change (delogarithmised induction or repression factors ≥ 2 or ≤ -2 , respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
Unknown (3)	DipC	3.1*	2.7**	1.5	1.4	-1.8	-1.3
	LunT	1.8	2.0	1.9	1.9	3.6*	-1.6
RD22-like protein (10)	DipC	2.3**	2.5**	1.9	1.9	1.1	1.3
	LunT	1.2	1.4	2.0	1.5	-2.0	-1.2
Seed albumin (18)	DipC	7.6	8.4*	7.2*	2.1	-2.3	2.5
	LunT	33.5*	33.8**	5.0	25.6**	1.6**	4.9
No homology (25)	DipC	1.1	1.1	-1.1	-3.3	-3.3	-2.5
	LunT	7.5*	8.3*	1.3	1.2	-3.5*	-1.8*

Table 12 (continued)

Gene annotation (MPSS cluster)	Land-race	Differential expression (induction/repression factors)					
		1RI	2RI	4RI	8RI	1REC	3REC
Auxin-repressed protein (39)	DipC	13.4*	18.6*	15.9**	3.3	-1.3	-1.1
	LunT	13.4**	22.0**	8.8**	20.9**	1.4	2.6
No homology (43)	DipC	6.3*	4.9*	1.9	-1.9	-2.6	-1.5
	LunT	7.9**	11.8**	1.6	-1.1	-3.0*	-1.6
No homology (63)	DipC	4.9**	5.0**	2.3	-1.2	-1.6	-1.2
	LunT	7.8	8.0*	1.8	-1.1	-1.9*	-1.7
No homology (68)	DipC	-1.0	2.7	5.1*	1.4	-1.6	1.1
	LunT	6.1	13.1**	2.9	3.6**	-1.3	-1.2
Unknown (76)	DipC	-1.0	-2.1	-3.3**	-1.6	-1.2	1.3
	LunT	1.1	-1.6	-1.5	-1.9	-1.4	-1.1
No homology (80)	DipC	4.5	3.7*	2.9	-1.4	-1.8	-1.2
	LunT	6.1	9.2**	2.0*	-1.6	-5.7**	-2.5
No homology (82)	DipC	2.3	1.9	2.1	-1.1	-1.5	1.2
	LunT	4.7	5.6**	1.1	1.7	-1.5	-1.2
No homology (97)	DipC	2.9	5.3*	12.4**	3.2	1.5	1.9
	LunT	5.7*	7.0**	2.6	11.4	1.7	1.1
Aluminium-induced protein (98)	DipC	2.6	3.0	1.1	1.5	-2.0	-1.3
	LunT	2.9	3.7*	1.2	-1.1	-3.1*	-1.3
No homology (101)	DipC	2.5*	3.9*	1.7**	-1.5	-1.8	1.1
	LunT	17.4	32.2*	1.8	1.4	-1.2	-1.1
Unknown (107)	DipC	7.9*	6.8*	3.5	1.3	-1.7	-1.6
	LunT	8.8*	10.8**	4.7*	2.1	-3.7*	-1.5
No homology (110)	DipC	6.4*	4.0*	3.7*	2.4	-1.0	1.5
	LunT	3.9	-1.0	-1.7	2.3	-1.0	2.7
No homology (111)	DipC	8.2**	20.3**	25.3**	5.1**	-1.0	1.5
	LunT	30.3	74.7**	22.5	36.7*	1.0	1.2
Unknown (115)	DipC	6.6*	6.1*	5.2*	-1.7	-3.3	-1.7
	LunT	5.4*	7.2*	1.7	2.8	-2.1*	1.0
No homology (116)	DipC	2.4	1.1	-1.3	-1.9	-1.6	-1.2
	LunT	2.3*	1.7	-1.5	-2.0	-3.7	-2.2
No homology (124)	DipC	3.6*	3.0*	1.3	-1.3	-1.3	-1.8
	LunT	1.8	1.5	-1.2	1.4	1.1	8.0
Unknown (126)	DipC	3.8*	7.1*	10.6*	1.3	-1.6	1.2
	LunT	12.5	26.2**	3.9	6.4*	1.3	1.2
Hypothetical protein (133)	DipC	2.6*	2.4	2.0	-1.6	-1.2	-1.2
	LunT	3.3	8.5*	1.9	1.1	-1.7	-1.2
No homology (154)	DipC	1.3	1.7	5.5**	1.9	1.4	1.8
	LunT	2.4	4.7**	1.3	4.0*	1.4	1.4
No homology (169)	DipC	3.1*	3.5*	3.8*	-1.3	-2.2	-1.6
	LunT	6.2*	3.9**	1.3	2.7	-1.6	1.1
No homology (195)	DipC	1.9	4.4	8.3*	1.4	-1.3	1.2
	LunT	14.5	23.9**	2.3	4.7**	1.1	1.1
Unknown (262)	DipC	2.6	1.6	-1.6	-1.4	-1.3	-1.1
	LunT	2.5*	1.5	1.1	-1.4	-3.3**	-1.4
Unknown (336)	DipC	2.2	1.7	-1.5	-1.8	-1.8	-1.1
	LunT	2.7*	2.6*	-1.3	-1.1	-1.8	-1.5
No homology (418)	DipC	1.6	1.3	2.1*	-1.1	-1.1	1.3
	LunT	2.6*	2.7*	-1.2	1.5	-1.5	-1.4
Unknown (442)	DipC	3.8	2.5*	1.3	1.1	1.2	-1.3
	LunT	2.5*	2.5*	1.3	1.4	-1.5	-1.1
No homology (458)	DipC	1.6	1.1	-1.1	-1.5	-1.3	1.3
	LunT	3.5*	1.5	-1.3	-1.4	-2.1	-1.3
No homology (494)	DipC	2.6	2.3	1.4	1.2	1.1	1.4
	LunT	3.5*	2.3	1.0	1.2	-1.6	-1.2

3.2.3.3.2 Landrace-specific differences

Statistical analysis of differential gene expression levels between landraces was carried out analogously to between-treatments analysis (see 3.2.3.3.1). In total, 20 of the 132 genes used for microarray analysis displayed at least twofold stronger expression and a p-value ≤ 0.05 in either landrace at at least one of the six time points investigated.

Nine genes turned out to be more strongly expressed in the landrace adapted to humid conditions, LunT. All of these differences became visible in the stressed state while reference plants grown under non-limiting conditions displayed no disparity to DipC. With the exception of the β -amylase gene, which also showed 2.3-fold higher, but not statistically significant, transcript abundance than DipC, all genes differed at the latest stress stage after eight days under water-deficit conditions. Additionally, three genes were also more strongly expressed at stage 2RI. Two genes (CPRD22/dehydrin and LEA4 protein) function as LEA proteins and were both induced by reduced irrigation. The genes encoding for caffeic acid methyltransferase, cinnamyl alcohol dehydrogenase and *o*-methyltransferase also displayed higher expression in LunT. The first two, possibly all three genes are involved in lignin biosynthesis. Cinnamyl alcohol dehydrogenase and *o*-methyltransferase were similar in their expression profiles (Table 13) with upregulation early upon water-deficit and a decrease at later stages. Caffeic acid methyltransferase however tended to be repressed.

Regarding genes with higher expression in DipC, differences were also observed in the unstressed variant. Seven genes constitutively accumulated more transcripts than in LunT. However, this phenomenon only occurred at single stages and did not continue all over the experiment. Of the five genes significantly exceeding LunT in their expression levels in the challenged variant, three were only different at single stages. MPSS cluster 20, encoding for a defensin, was 6.8-fold more expressed in DipC 24 hours after initiating the stress treatment. Thereafter, differences were not significant but expression tended to be higher in LunT. A LEA5 protein was found to be significantly overexpressed in the first recovery stage and also tended to be six times as abundant as in LunT after four days under water-deficit conditions. At the same time, the *myo*-inositol oxygenase 5 transcript accumulated slightly stronger. While the former two genes were generally upregulated upon water-deficit stress, the latter did not exhibit treatment-specific differences in the second CE experiment. Two genes, both without significant homologies to published sequences, were by far more strongly expressed in DipC than in LunT. Cluster 110 was significantly upregulated in DipC at the first three stress stages, while no significant difference was observed in LunT. This also became apparent in the between-landrace analysis. As also explained above, cluster 124 was not detected to be expressed in LunT. Thus, all DipC data points (except for 3REC) are highly significant over LunT.

Table 13. Genes with statistically significant differential expression between landraces (delogarithmised expression factors ≥ 2 or ≤ -2 , where positive and negative values stand for higher expression in DipC and LunT, respectively) at at least one sampling date in the 2008 CE water-deficit stress experiment. *: $p \leq 0.05$ (significant); **: $p \leq 0.01$ (highly significant)

Gene annotation (MPSS cluster)	Treatment	Expression factors					
		1RI	2RI	4RI	8RI	1REC	3REC
Higher expression in DipC							
Proline-rich protein precursor (1)	Control	2.6*	-	-1.0	-	1.2	-
	Stress	1.3	1.0	1.2	-1.1	1.4	1.3
Seed albumin (18)	Control	5.8	-	1.2	-	3.3*	-
	Stress	1.3	1.5	1.7	-3.7	1.1	1.9
Stored cotyledon mRNA (defensin) (20)	Control	3.1	-	2.5	-	-1.1	-
	Stress	6.8*	-1.7	-7.6	-2.6	1.4	1.6*
Endo-1,4- β -mannanase (34)	Control	-1.2	-	-2.1	-	2.9*	-
	Stress	-1.0	-1.6	-1.2	-1.2	1.7	1.8
LEA5 protein (36)	Control	2.1	-	-1.2	-	1.5	-
	Stress	-1.3	1.2	6.0	1.4	4.5*	2.1
Auxin-repressed protein (39)	Control	1.0	-	1.1	-	4.8*	-
	Stress	-1.0	-1.2	2.1	-1.3	2.7	1.8
No homology (110)	Control	2.4	-	1.1	-	2.1	-
	Stress	3.9*	9.7**	6.9**	2.2	2.1	1.2
No homology (111)	Control	2.5*	-	1.4	-	1.5*	-
	Stress	-1.5	-1.5	1.6	-4.8	1.5	1.8
No homology (124)	Control	19.9**	-	20.4**	-	32.3**	-
	Stress	39.4**	41.3**	29.5**	17.6**	22.5**	2.2
Hypothetical protein (133)	Control	2.2*	-	1.2	-	1.6	-
	Stress	1.7	-1.6	1.3	-1.1	2.1	1.6
<i>myo</i> -Inositol oxygenase 5 (152)	Control	1.7	-	-1.2	-	1.6*	-
	Stress	-1.2	1.8	2.1*	-1.3	1.2	1.8
Higher expression in LunT							
CPRD22 protein (dehydrin) (2)	Control	2.0	-	-1.2	-	1.7*	-
	Stress	-1.0	-1.6	1.6	-3.2*	1.1	1.9
LEA4 protein (52)	Control	1.6	-	-1.1	-	1.4	-
	Stress	-3.0	-3.9*	1.5	-2.9*	-1.2	1.8
No homology (68)	Control	2.6	-	-1.5	-	-1.0	-
	Stress	-2.5	-1.9	1.2	-2.6*	-1.2	1.3
Caffeic acid methyl-transferase (99)	Control	1.4	-	1.1	-	-1.4	-
	Stress	1.1	-1.3	-3.5	-2.2*	1.1	-1.1
CPRD14 (cinnamyl alcohol dehydrogenase) (102)	Control	1.8	-	1.0	-	1.1	-
	Stress	-2.3	-4.2	-1.6	-3.0*	-1.3	1.2
<i>o</i> -Methyltransferase (112)	Control	1.4	-	-1.2	-	-1.2	-
	Stress	-2.4	-3.9*	-1.2	-2.7*	-1.4	1.6
β -Amylase (114)	Control	-1.4	-	1.0	-	-1.6	-
	Stress	-1.3	-2.6*	-1.3	-2.3	1.1	1.0
Early light-induced protein (226)	Control	1.2	-	-1.0	-	-1.9	-
	Stress	-1.3	-2.3	-2.1	-3.4*	-1.7	-1.2
Polyubiquitin 2 (300)	Control	1.2	-	-1.0	-	1.2	-
	Stress	-1.6	-1.6	1.0	-2.4*	-1.1	1.2

4. Discussion

4.1 Genetic diversity

4.1.1 Experimental approach

The initial DArT marker discovery array was constructed from 38 landrace individuals using the restriction endonucleases combination *PstI/AluI*. Seventy-six clones proved to be polymorphic within 32 geographically distinct genotypes which corresponds to a proportion of 5.5% of approximately 1,380 successfully amplified clones. This is similar to the findings of Yang *et al.* (2006) who discovered maximum 5.9% - depending on the method of complexity reduction - polymorphic clones in a set of 48 cultivated pigeonpea accessions, another “orphan” legume with similar genome structure and breeding system. Other crops, to which DArT had been applied in a comparable way, featured higher frequencies of polymorphic clones. Polymorphism levels of 2.9 to 10.4% were found in barley discovery arrays which were prepared from only two cultivars (Wenzl *et al.*, 2004). Wittenberg *et al.* (2005) reported 7.3% by building the marker library from only one *Arabidopsis* ecotype and using a mapping population to score polymorphisms. Values of 4.2 to 9.4%, derived from 13 cultivars, were obtained in wheat (Akbari *et al.*, 2006). In cassava, polymorphism frequencies reached between 9.0 and 17.2% (Xia *et al.*, 2005). However, related species were also included in the latter study.

Closer investigation of the 76 bambara groundnut DArT markers obtained in the initial screening experiment revealed a high proportion of clones with identical segregation signatures. Through sequence analysis, six markers with a repeated scoring profile were shown to be unique. This suggests these markers represent separate loci that are closely linked in the bambara groundnut genome. Integrating these markers into a genetic linkage map, which is being done by the University of Nottingham (Mayes *et al.*, 2009), will further address to this question. The vast majority of clones, however, proved to be redundant as sequences were identical or highly similar. One group of truly repetitive clones consisted of 44 members, thus accounting for 58% of all informative markers discovered. Such an enrichment of clones is usually associated with the occurrence of a prominent individual band when investigating the PCR amplified genomic representation on an agarose gel. It remains unclear why this was not observed during the optimisation of the complexity reduction method, where homogenous smears indicated even amplification of fragments. A similar case of overrepresented clones was reported in oat by Tinker *et al.* (2009), who identified a set of around 100 redundant clones among 6,144 clones arrayed which were characterised by varying numbers of tandemly-repeated 171bp elements sharing similarity with telomere-associated sequences. In contrast to the latter study, the highly redundant bambara groundnut clones were present at far higher frequencies and their consensus sequence shows clear

homology to an expressed sequence from cowpea (GenBank accession number FG879525.1, E-value $2e^{-108}$, nucleotide identity 244/258). It is not surprising that these clones do not seem to originate from predominantly repetitive genomic regions which are highly methylated (Rabinowicz *et al.*, 1999). Genomic representations were generated by selective amplification of overhangs of *Pst*I, which is sensitive to most methylation events tested so far (Roberts *et al.*, 2003). However, the special case observed in bambara groundnut deserves further attention.

Taking redundancy into account, the level of polymorphic clones was reduced to 1.9% in the initial *Pst*I/*Alu*I library.

Based on these findings, an expanded array from three different complexity reduction methods was constructed. The inclusion of 1,152 clones from the initial *Pst*I/*Alu*I library allowed for direct comparison of both screening panels. In the first experiment, 32 genotypes were used to discover polymorphism. Among 5.5% polymorphic clones, 19 showed unique segregation signatures (disregarding clones with unique sequences but redundant scoring profiles), which corresponds to 1.4% of the successfully amplified clones. When these clones were screened with a panel of 94 *Pst*I/*Alu*I-digested accessions, total polymorphism rate rose to 8.6%. This suggests that the higher number of target libraries added to a certain degree of polymorphism. However, this increase was only based on the augmented occurrence of the highly redundant clones. Sorting redundancy out from the data again resulted in 1.4% unique scoring profiles. As the initial probe library had been prepared from a smaller set of germplasm, it did not contain alleles specific for additional genotypes. Nevertheless, it would have been expected to find additional clones scoring '0' while being attributed a '1' in all previously analysed genotypes, which was not the case. Consequently, the initial library prepared from 38 diverse genotypes according to Singrün & Schenkel's (2003) AFLP study seemed to sufficiently cover the greatest portion of genetic diversity within (for the most part) cultivated bambara groundnut.

Beside the generally low genetic diversity detected through the *Pst*I/*Alu*I method, the second factor limiting the performance of this approach was the overrepresentation of identical or highly similar clones. The consensus sequence (Fig. 22) contains a *Bgl*III recognition site. Thus, this restriction endonuclease was used to prevent further PCR amplification of such fragments in preparing another *Pst*I/*Alu*I library from the 94-genotype discovery panel. Hybridising *Pst*I/*Alu*I targets to this sub-array resulted in a drastic decrease to only two clones with the respective segregation signature out of 4,992 clones investigated compared to 3.7% and 6.3% in probes prepared without the additional digest.

```

GATGGATCCAGTGCAGCAACAGAGTGAGATGTCTTATTTGAAGCCAGGTTACCGTAAATGAC
AACGAGTAGATGTCTAAAAGTGCCCATTTGATTCAATAAATAAACAAACTACCATTTTATGGC
TTTGAAATAAAATCCAAGAAGTTTGTGTGTCAGAAAACAACAACATTCAGTACAAGTATGGCT
ACGGGGTCATTGATTAGTAAGATTGGTTATGGGTTTGGTAGATTGAGGTAAATAATCAAATA
AATGGGGTTTGGATATAAGATCTATGTAAACGAGCCTTGGAGGAACAAACTGCAAGCAATAA
TACCATAAAAATTTTGCTAACTCCATCTTTGAGACATAGTTAAAGTCTGTTCATACTGCAAC
TCTACTGAAATCGCCACTAAAATGGACGCTAGTTCTGCACTAAGATAAGATTATTTAATATC
ATAGATACACATGCCATTCTCTTCCAAGTGCTATGCTTTCTGCCGGTACTGTCCACCAATTGA
AGTGACATCGAACTAACTATATATTTCTTGTCAATTTTGCGGCCTGCACTGGATCCATC

```

Figure 22. Consensus sequence of the highly abundant DArT clones in the initial *PstI/AluI* library. *PstI* adapters, the *BgIII* recognition site and the *TaqI* recognition site are highlighted in grey, blue and red, respectively.

The frequency of total polymorphic clones was reduced to 4.1% in this combination of targets and probes, which confirmed the low level of diversity in cultivated bambara groundnut as concluded from the initial results. The fraction of clones with unique scoring profiles was however increased to 2.6% after eliminating highly repetitive markers without the occurrence of prominent clusters. Marker redundancy of 1.59-fold is in good agreement with results published for *Arabidopsis* (Wittenberg *et al.*, 2005; 1.78-fold) and oat (Tinker *et al.*, 2009; 1.51-fold) and lower than in barley (Wenzl *et al.*, 2004; approximately 2.5-fold). While in the former two studies redundancy was revealed through sequence analysis of clones, mere investigation of scoring profiles in the bambara groundnut full-size array excluded markers at closely linked loci. Thus, the actual number of unique markers is likely to be a little higher. Nevertheless, including clustered markers in a diversity study would not add new information but lead to biases of particular genetic regions. Technical problems of the *PstI/AluI* complexity reduction method were hence overcome by implementing an additional digest with *BgIII* in the restriction-ligation reaction.

In order to evaluate the above-mentioned results, another complexity reduction method replacing the frequently cutting enzyme *AluI* by *TaqI* was tested. Regarding the fraction of markers obtained with both targets and probes prepared using the *PstI/TaqI* combination, polymorphism levels were 6.1% and 4.2% for total and uniquely scoring polymorphic clones, respectively. Segregation signatures of the well-known highly abundant clones were not observed for any of these markers. This is explained by looking at the sequence (Fig. 22) which contains a restriction site for *TaqI*. Accordingly, such fragments were not amplified in the PCR reaction. In terms of both the frequency of polymorphic markers and redundancy (1.47-fold), the *PstI/TaqI* sub-array outperformed the *AluI*-based methods. Differences among restriction endonucleases in their ability to detect DNA polymorphisms were observed in several studies on DArT. The use of *TaqI* as codigesting enzyme in combination with *PstI* resulted in the highest levels of polymorphism in wheat (Akbari *et al.*, 2006) and in barley, where *TaqI* yielded 10.4% polymorphic clones compared to 7.0% for *AluI* (Wenzl *et al.*, 2004). In cassava, *TaqI* performed second to *BstNI*, but significantly better than *ApoI* (Xia *et*

al., 2005). Kilian *et al.* (2005) hypothesised that this phenomenon may be attributed to different methylation sensitivities. While *TaqI* is able to cleave methylated CpG, one of the predominant sites of methylation in higher plants (Gruenbaum *et al.*, 1981), *AluI* is sensitive to most methylation events (Roberts *et al.*, 2003). However, as the methylation-insensitive enzyme detects more polymorphisms, the capture of additional methylation polymorphisms cannot be considered as an explanation.

In practise, targets were prepared using both the *PstI/AluI* and the *PstI/TaqI* complexity reduction methods and individually hybridised to the full-size DArT array containing clones from three different libraries. Hybridisation and the discovery of polymorphic clones did not only occur between probes and targets prepared from the same enzyme combinations but also between the respective others. This compatibility appears logical as both methods rely on the selective amplification of restriction fragments with *PstI* overhangs on both sides. Thus, cloned fragments are potentially of the same origin. Likewise, the molecular basis of polymorphisms detected through DArT is similar. SNPs, InDels and/or methylation polymorphism within the *PstI* recognition sites or InDels within the amplified fragments are potential features of variation common to both enzyme treatments. Evidence is given by the fact that, when the two datasets were merged, 43 clones produced identical scores for differently prepared targets. The presence or absence of recognition sites for the codigesting endonucleases thus remains the only difference. Levels of polymorphism were always higher when both probes and targets were prepared using the same enzyme combination (Table 3). This may be due to the generally lower number of fragments complementary between probes and targets as a part of potentially hybridising probes is filtered by digestion with the other frequent cutter. However, regarding *PstI/AluI* probes, the level of polymorphism revealed by screening with *PstI/TaqI* targets was comparable to the fraction obtained with congeneric targets, basically only differing with respect to the highly repetitive clones. This again points out the particular usefulness of *TaqI* as codigesting enzyme in this approach. With respect to methylation, the sensitivity of *AluI* may have resulted in the PCR amplification of a greater share of *PstI-PstI* fragments and thus only marginally decreased the number of clones complementary to *PstI/TaqI* targets. Sequencing the polymorphic clones from the initial library indeed revealed the presence of *AluI* recognition sites in several clones, which partly supports this hypothesis. In addition, 35 clones were found to hybridise with targets prepared from both enzyme combinations, but gave independent scores. Accordingly, such clones must contain neither (accessible) *AluI* nor *TaqI* recognition sites whereas targets are polymorphic in terms of cleavage through codigesting endonucleases.

In summary, the degree of genetic diversity in cultivated bambara groundnut revealed through DArT was lower than in most other species investigated so far. However, results were similar to those for cultivated pigeonpea (Yang *et al.*, 2006), the closest relative in terms of biology and “breeding” history with DArT data available. Isozyme analysis also indicated low diversity within domesticated bambara groundnut, although the technique was capable of detecting higher diversity for wild populations (Pasquet *et al.*, 1999). A reason for this is the

self-pollinating nature of bambara groundnut as indicated by very low levels of heterozygosity (Basu *et al.*, 2007b). Inbreeding is expected to decrease diversity, which has been shown for many organisms (Charlesworth, 2003). On the other hand, AFLP and RAPD markers suggest far higher levels of genetic polymorphism (Amadou *et al.*, 2001; Massawe *et al.*, 2002; Massawe *et al.*, 2003; Singrün & Schenkel, 2003; Ntundu *et al.*, 2004). However, dendrograms from these analyses displayed Jaccard similarity coefficients of a similar range, which may be attributed to a higher discriminatory power (PIC values) of DArT markers.

4.1.2 Cluster analysis

Despite the low number of unique DArT markers discovered in the initial experiment, high average PIC values of 0.32 (the maximum for a biallelic marker system is 0.5) allowed unambiguous discrimination of 32 distinct genotypes. UPGMA clusters tended to form two groups of accessions at a similarity coefficient of 0.55 (Fig. 4b). One cluster exclusively contained germplasm originating from Southern Africa (plus two landrace individuals from Indonesia); the other one consisted of materials from Western Africa. Three outliers were found, but these arose due to the inclusion of redundant scoring profiles in spite of unique sequence data.

The UPGMA dendrogram obtained from 296 DArT markers with unique segregation signatures and 87 genotypes confirmed the preliminary results of separation of Southern African genotypes and further dissected bambara groundnut germplasm collected north of the equator into three groups (Fig. 6). Apart from the higher number of markers, a reason for this is certainly the fingerprinting of a larger set of genotypes. While the preliminary dendrogram included only eight samples from the northern hemisphere, 37 were utilised in the full-size experiment. Plotting bambara groundnut accessions and landraces against their proveniences (Fig. 5) showed good correlation between genetic relationship and geographic origin. Similar findings were reported by Amadou *et al.* (2001). Cluster IV consisted of only two genotypes in the dendrogram presented, but was confirmed in the subsequent large-scale genotyping, which added three more genotypes (Table 4). With the exception of one Nigerian accession, these genotypes originated from countries in the outmost west of Africa (Sierra Leone, Gambia and Senegal). Cluster III mainly included accessions from countries along the southern coast of West Africa, i.e. from a stretch between Cote d'Ivoire and Nigeria, and countries north of it (Mali and Burkina Faso). The origins from accessions in cluster II tended to be shifted further towards the east of the North African savannah belt with most accessions hailing from Cameroon and the Central African Republic. This region fits reasonably well with the area in which wild bambara groundnut was collected (Hepper, 1963). One spontaneous form (VSSP6) clustered with these accessions. Thus, it seems possible that cluster II reflects genotypes that did not completely undergo the process of domestication. However, regarding the coefficients of similarity, there was no clear separation between

clusters II to IV and drawing the discrimination line at 0.51 seems to be a little arbitrary. This was also apparent in the PCoA plot from 429 genotypes (Fig. 7), where northern accessions formed a rather unstructured cloud. Consequently, there were several overlaps between clusters in terms of countries included. For example, 16 accessions from Cameroon were represented in cluster II while 8 were grouped under cluster III. It is not surprising that no pure country-specific clusters were observed. The spread of bambara groundnut out of its natural habitat is expected to have occurred long before political borders were drawn in the course of colonisation in the middle of the 19th century. Instead, clusters seem to be linked to ethnolinguistic groups (Murdock, 1959). Western Bantoid languages prevail in the area of cluster IV, cluster III matches well with Central and Eastern Bantoid and Guinean languages and Central and Eastern Sudanese languages are spoken in the places of accessions in cluster II.

A clear distinction between the above-named germplasm and cluster I was obvious. Geographic assignment showed this separation, with few exceptions, being in accordance with the equator as the line of discrimination. Comparing H_S values indicated a significant loss of genetic diversity in cluster I in the range of one third. Such genetic bottlenecks are usually the result of human selection in the processes of domestication and introduction to foreign places as shown for many crops like cowpea (Parella & Gepts, 1992), cotton (Iqbal *et al.*, 2001) and soybean (Hyten *et al.*, 2006). Direct evidence for a domestication bottleneck between wild and cultivated material was not obvious in the data presented. Two wild relatives were included in the final large-scale DArT analysis. These were not found to be outgrouped from IITA accessions or landraces. A reason for this may be that the genotyping array was mainly developed using DNA from domesticated germplasm and alleles specific for spontaneous forms were highly “diluted” by common alleles (Mace *et al.*, 2008). However, the work of Pasquet *et al.* (1999) has shown genetic erosion during the domestication of bambara groundnut. Consequently, a second bottleneck must have occurred when germplasm was moved from the centre of domestication towards Southern Africa. As cluster I was more closely linked to cluster II than to the others, it is possible that these materials directly evolved from the proposed “wildish” cluster. Tanzania would then function as a geographical bridge pier. The average allele diversity of 0.19 within Tanzania accessions was close to the total diversity of cluster I (0.21). This narrowing genetic base may lead to the hypothesis of bambara groundnut moving from Eastern Africa towards the south. Including accessions from Gabon, Congo and Angola will provide further insight to this question.

It is generally difficult to interpret clusters in a context other than the geographical or ethnological one. A number of morphological and agricultural descriptors are available for bambara groundnut (IPGRI, IITA, BAMNET; 2000). However, the evaluation of accessions is time consuming and expensive, which is a particular constraint in “orphan” crops. Thus, the data available are far from being complete. Furthermore, it has not been achieved to group domesticated bambara groundnut by means of morphological characteristics except for internode length, a trait which is likely to be under monogenic control (Basu *et al.*, 2007a) and

can vary even within landraces. Moreover, this and other phenological traits are strongly influenced by environmental conditions, especially in the light of different degrees of photoperiod response (Linnemann, 1991; Harris & Azam-Ali, 1993; Brink, 1999). Consequently, weak correlation between molecular and phenotypic data has been reported (Ntundu *et al.*, 2004).

4.1.3 Intra-landrace diversity and “exotic” germplasm

In addition to the analysis of single individuals from germplasm collections, each ten or more individual plants from the six BAMLINK core landraces were investigated with respect to genetic diversity. No landrace was found to consist of a single genotype which confirmed the assumption that bambara groundnut landraces are true landraces and not pure lines (Zeven, 1998). Nevertheless, very few duplicates were detected. This and the conservative marker quality parameters applied point out that the DArT technique did not overestimate intra-landrace variation. AFLP and RAPD marker studies (Massawe *et al.*, 2003; Singrün & Schenkel, 2003) as well revealed diversity among landraces. However, while these authors concluded that intra-landrace diversity is generally lower than between landraces, DArT data suggested considerable differences among landraces. Two landraces, S19-3 and Ramayana, indeed formed narrow sub-clusters. In contrast, the landraces Swazi Red and DipC displayed such a high genetic dissimilarity so that landrace individuals occasionally showed more similarity to other genotypes than to other landrace members when they were co-analysed (not shown). DipC even exhibited average allele diversity which was two thirds of the entire cluster I covering half of the African continent. Due to the fact that Dodoma Red was somewhat isolated within cluster I and Tiga Nicuru belonged to cluster III, where diversity is generally higher, clusters did not break apart for these landraces when co-analysed with the full set of genotypes although their allelic diversity was greater than the one of Swazi Red (0.11 and 0.12 for Dodoma Red and Tiga Nicuru, respectively, vs. 0.09 for Swazi Red). Initial data using microsatellite markers revealed similar difference among landraces (Mayes *et al.*, 2009). While three of 14 SSRs produced ten alleles in Uniswa Red (=Swazi Red), no polymorphism was found in S19-3. Pasquet *et al.* (1999) also observed high genetic diversity within domesticated populations in relation to the generally low diversity of cultivated bambara groundnut.

Furthermore, accessions and landraces from Indonesia and Madagascar were used to study the genetic diversity of bambara groundnut in places isolated from its origin. Madagascan accessions showed a fairly high degree of diversity ($H_S=0.12$). All genotypes clustered with sub-equatorial germplasm, and five accessions clustered together. Their closest relatives were two landrace individuals from Namibia and Botswana. Another accession was placed in a separate clade in the dendrogram and was highly similar to a Zimbabwean accession. Thus, UPGMA clustering suggested that bambara groundnut was introduced to Madagascar from

South-Eastern Africa, which is plausible due to the small geographic distance, either as a diverse mixture of seeds or more likely, in the course of several independent incursions. Results were clearer with Indonesian materials. In addition to the limited genetic base of landrace Ramayana, individuals from three other Indonesian landraces were found within this sub-cluster. Two accessions from Tanzania displayed high genetic similarity according to the UPGMA dendrogram of 429 genotypes (not shown). Thus it appears that all landraces from Indonesia can be traced back to sparsely diverse germplasm from Tanzania, provided the materials investigated are representative of all Indonesian bambara groundnut landraces. The narrow genetic base of bambara groundnut then is the result of at least three genetic bottlenecks. The first one occurred during the domestication of bambara groundnut, the second one is associated with the crop moving out of its centre of domestication, and the third one happened when it was introduced to Indonesia, possibly via other South and Southeast Asian countries.

4.2 Gene expression under water-deficit-stress

4.2.1 MPSS expression profiling

Massively parallel signature sequencing (MPSS) was carried out using the 454 pyrosequencing technology. One run created a total of 23 million bases and, after trimming tags by removal of adapters and polyA-tails, 17 million bases of informative, high-quality *de novo* sequence for bambara groundnut. Clustering 197,400 sequence tags yielded 10,583 groups of homologous sequences and 34,427 singleton transcript tags. Random checks were performed to assure the accuracy of clustering by BLASTing consensus sequences which were “assembled” using stringent parameters. Tags did not assemble into longer contigs as the 3′ proximate *Nla*III restriction sites fixed the position of 5′-ends and the 454 protocol limited the length of 3′-ends of the sequences. This resulted in a cleaned average read length of 87 nucleotides. Sequence alignments using the *blastn* algorithm yielded 408 (72.7%) matches to published mRNA sequences among the 561 largest transcript clusters at a significance threshold (E-value) of $1e-4$. As the frequency of discovering annotated genes strongly declined with increasing depth of analysis, i.e. investigation of clusters containing a diminishing tag quantity, comparative statistics are only presented up to a cluster size of at least 40 transcripts. It is unlikely that the decreasing homologies to annotated genes or ESTs are due to the short tag lengths. A more likely reason is the position of sequence tags in the 3′-UTR of mRNAs. In comparison with coding regions, the 3′-UTR is known to be richer in SNPs and particularly InDels (Ching *et al.*, 2002; Zhu *et al.*, 2003). Thus, this region is highly specific for individual genomes and less conserved between different species. While this feature greatly facilitates the discrimination of closely related genes (Eveland *et al.*, 2008), cross-species alignment may become difficult with increasing evolutionary distance. As for bambara groundnut, there are no cDNA sequences available in public databases, highest sequence homologies were obtained with ESTs or annotated genes from nearest relatives. Out of the 561 largest transcript clusters, 191 produced best matches to sequences from *Phaseolus* species (*Ph. vulgaris*, *Ph. coccineus* and *Ph. acutifolius*) and 100 to sequences from *Glycine* species (*G. max* and *G. soja*). Both genera are grouped together with *Vigna* in the *Phaseolae* tribe. Taxonomic delimitation between *Vigna* and *Phaseolus* has been an issue among botanists for a long time (Maréchal, 1982), and the genera are estimated to have separated only about eight million years ago (Lavin *et al.*, 2005). Search in *Vigna* species itself was less effective (each 16 hits with *V. unguiculata* and *V. radiata*) due to limiting published resources. A significant number of cowpea ESTs was recently submitted to the NCBI database (Varshney *et al.*, 2009), and re-BLASTing would probably have shifted the statistics towards cowpea. However, all unmatched sequences were aligned again in March 2009, and hardly any new hits were obtained. Other species subject to extensive EST projects such as maize, *Arabidopsis*, rice or even the legume *Medicago truncatula*, which shows common

ancestry with bambara groundnut very early in the course of legume diversification (around 54 million years ago; Lavin *et al.*, 2005), did not much contribute to annotating additional genes.

The proportion of transcript tags without sequence homology was 27.3% in highly expressed genes, but exceeded 70% with increasing depth of analysis. While 454-based transcriptome sequencing is able to discover a lot of novel transcripts even in the model species *Arabidopsis* (3.5%; Weber *et al.*, 2007), maize (22 and 30%; Emrich *et al.*, 2007; Eveland *et al.*, 2008) and *Medicago* (29%; Cheung *et al.*, 2006), this high fraction in bambara groundnut is noticeable as soybean proved to be suited for BLAST search and the number of available ESTs is only little lower than in *Arabidopsis*. The short average read length of public soybean ESTs (474 nucleotides as of December 2003; Mooney & Thelen, 2004) may provide an explanation why sequence alignment of 3'-anchored transcript tags was constrained. A similar problem became apparent when transcripts were functionally annotated. Of 561 MPSS clusters, 125 could be directly assigned to a gene in the 'non-redundant' NCBI database. The remainder was first aligned to the 'est_others' database, and if significant homologies to ESTs were obtained, these were used to query the 'non-redundant' database. In 184 cases, significant hits were produced. As exemplified in Figure 10, this was usually due to the shortness or absence of the 3'-UTR in annotated genes. Eighty-seven ESTs homologous to highly expressed bambara groundnut genes did not match a 'non-redundant' gene which indicates that, independent from possible technical limitations, a comprehensive image of transcriptomes in legume crops is still not achieved.

A general technical constraint of the approach presented is the requirement of an *Nla*III recognition site. Torres *et al.* (2008) estimated that 4% of the *Drosophila melanogaster* transcripts lack such a 4bp site. Secondly, these restriction sites may be too far away from the polyA-tail in some transcripts (estimated at 6% by the same authors) so that these would also be lost if cDNA transcription primed with oligo(dT) is incomplete. Thus, it is possible that around 10% of the bambara groundnut transcriptome are missing in the MPSS libraries. Generating additional libraries using a different 4bp restriction endonuclease would likely increase transcriptome coverage. Furthermore, when libraries are prepared from different genotypes, it is possible that sequence polymorphism is present in restriction sites which results in a gene breaking up into different clusters due the varying position of the sequenced tag. This was demonstrated in two cases by means of alignment to a common homologous sequence (Fig. 9). Within the set of 561 MPSS clusters, the situation of transcript tags missing in one genotype occurred another eleven times, but the reason could not be resolved due to lacking BLAST homologies. In all cases, transcript tags were absent in LunT, which is consistent with the DArT diversity analyses, where DipC, AS-17 and Swazi Red showed high similarity among each other and maximum diversity to LunT (Fig. 5).

Although the analysis of MPSS data was constrained by limiting functional annotation, further investigation of anonymous cDNAs was carried out up to a depth of 10 transcripts per cluster. As the cleaned dataset of nearly 200,000 454 sequencing reads evenly dispersed into

eight sub-libraries, it became increasingly difficult to obtain significant expression profiles for weakly expressed genes. Therefore, the fraction of genes differentially expressed decreased from two thirds to little more than 50% comparing the 561 and 2,425 largest clusters. As inherent to non-normalised cDNA libraries (Soares *et al.*, 1994), results were strongly biased towards highly expressed genes. The eleven largest clusters alone contained 10% of all transcript tags, and one fourth of sequence information was allocated to only 65 genes. Nevertheless, it was still possible to detect clear expression profiles and, in case of successful functional annotation, some of the less abundant cDNAs were selected for the validation experiment. Continuous improvement of the 454 technology or different high-throughput sequencing platforms will allow an increased depth of analysis due to higher numbers of transcript tags and sequence alignments may also benefit from read extension. The latest advancement in 454 sequencing has led to the generation of more than one million reads per run with an average length of 400 nucleotides (Rothberg & Leamon, 2008).

4.2.2 Validation of MPSS through microarrays

From a subset of 132 consensus sequence tags, 50-mer oligonucleotide probes were designed and printed on microarray slides. Plant material from the 2006 water-deficit stress experiment was used to validate the gene expression data obtained through the tag-based method against the hybridisation-based microarray technology. However, direct comparison proved difficult as the original RNA preparations could not be used. Instead, leaves originated from different landrace individuals which had been selected for uniformity on the basis of morphological criteria. Yet in the light of high genetic intra-landrace diversity, this may be a serious issue. Secondly, only one sample per landrace was analysed. Lee *et al.* (2000) stressed the need of both biological and technical replication due to the considerable variability of any microarray gene expression experiment and suggested the use of at least three replications. Nevertheless, contrasting the \log_2 ratios obtained through MPSS with those from microarrays indicated good consistency between both platforms for highly to medium expressed genes. \log_2 ratios were usually in a similar range, with only sporadic discrepancies. Due to measuring absolute transcript abundances, MPSS is considered to provide a greater dynamic range of gene expression than microarrays, in which, dependent on the sensitivity of scanning parameters, hybridisation signals are expected to level off at certain thresholds (Jongeneel *et al.*, 2005; Chen *et al.*, 2007). On the one hand, highly expressed genes may either result in saturation of the fluorescent signal, or, on the other hand, genes with low abundance may fall beyond the detection level. Therefore, absolute transcript abundances are weakly correlated between technologies, but logarithmised expression ratios allow a better comparison (Chen *et al.*, 2007; Liu *et al.*, 2007). However, one substantial difference was observed for MPSS cluster 110. MPSS indicated the presence of landrace-specific variation, whereas microarrays suggested consistent downregulation in all landraces. The latter may be regarded as a sign that

microarrays came closer to the biological truth in this case and 454 sequencing or subsequent clustering may have produced an artefact. In turn, this issue seemed less clear in the subsequent water-deficit stress experiment (discussed later).

With increasing depth of analysis, logarithmised gene expression ratios showed greater disagreement between both technologies, which has also been observed in other comparative studies (Liu *et al.*, 2007; Nygaard *et al.*, 2008). Again, microarrays tended to display higher consistency between samples than MPSS. Liu *et al.* (2007) stated that neither technology is able to reliably detect transcripts expressed at very low frequencies. However, commercially fabricated microarrays available for major crops or model species contain more than 10,000 genes and are used as standard for genome-wide gene expression profiling (Close *et al.*, 2004; Swanson-Wagner *et al.*, 2006; Gallardo *et al.*, 2007; Müller *et al.*, 2007), which underpins their significance in profound gene expression experiments. The dubiety of MPSS expression profiling started around cluster 200, i.e. the gene with the 200th most abundant copies in the full multiplexed cDNA library which corresponded to around 550 TPM. The least expressed gene chosen for microarray validation, cluster 1236, was counted 19-times (96 TPM) in the MPSS library and possessed good processability in the microarray analysis. In relation to the 45,010 tentative unique sequence tags discovered through MPSS, this gene still displayed reasonably high expression. Therefore, it would stand to reason that microarrays provide a more suitable tool to investigate differential gene expression below a certain abundance threshold due to the aforementioned technical limitations of the MPSS approach, namely a strong bias towards the most abundant transcripts and, in this particular case, the partitioning of one 454 sequencing run into eight sub-libraries.

Another concern that has often been raised is the higher frequency of sequencing errors in the 454 technology compared to traditional Sanger sequencing (Wicker *et al.*, 2006; Bräutigam *et al.*, 2008). The ‘non-redundant clustering’ procedure to figure out the sequences for oligonucleotide design indeed revealed occasional nucleotide inconsistencies among sequence tags from individual genotypes especially towards the end of the reads and in stretches of low complexity, e.g. homopolymers like polyA. However, as transcript counting involves a high degree of oversampling, it was not problematic to identify the correct consensus sequence. Consequently, all 132 oligonucleotides designed from the 454 data resulted in clear hybridisation signals in at least one treatment variant (stressed vs. non-stressed). Furthermore, microarray analysis confirmed that the lack of homologous transcript tags in one genotype (LunT) was attributed to sequencing a different position of the cDNA in 12 of 13 cases (Fig. 15). However, one such difference (MPSS cluster 124) remained to be true.

4.2.3 Water-deficit stress reaction in bambara groundnut

4.2.3.1 Physiological changes

While the work presented focussed on gene expression, several approaches were undertaken to link these data to physiological processes in bambara groundnut in response to water-deficit stress. The first visible sign of altered physiology was a change in leaf orientation, which started two days after initiating the water-deficit treatment in the second CE experiment. While untreated plants generally oriented their leaves in a horizontal position, drying soil partly led to leaflet angles steepening towards the light source. Such paraheliotropic movements may decrease light interception and thus lower leaf temperatures, which entails less transpiratory water loss and a reduced risk of damage through photoinhibition (Ludlow & Björkman, 1984; Yu & Berg, 1994; Pastenes *et al.*, 2004). The first CE experiment had shown that this trait is particularly pronounced in landrace AS-17. Simultaneously, the plant used for MPSS expression profiling constantly maintained comparably high expression levels of energy- or metabolism-related genes that were strongly repressed upon the stress treatment, e.g. MPSS clusters 8, 9, 11, 16, 19 and 26, which is reflected by the lowest repression factors of the four genotypes observed (Table 17). Hence, these findings suggest that paraheliotropism may play a pivotal role in dehydration avoidance. However, in order to minimise the influence of differing degrees of plant intrinsic stress, two landraces were chosen for the subsequent water-deficit stress experiment that do not (LunT) or only slightly (DipC) change their leaf orientation in response to water-deficit (Fig. 16), as previously observed by Collinson *et al.* (1999).

LunT exhibited another strategy to minimise water loss. Three days after initiating the stress treatment in the second CE experiment, the oldest leaves began to show necrotic lesions which continued until complete dieback while shoot tips continued growth (Fig. 17). In cowpea, this trait was described as one common mechanism of drought tolerance (Mai-Kodomi *et al.*, 1999). DipC leaves did not display any signs of damage. Other physiological mechanisms of drought accommodation in bambara groundnut may include decreased leaf initiation and expansion, which reduce the photosynthetic area and thus shorten water use (Collinson *et al.*, 1997; Collinson *et al.*, 1999). However, both treatments in the CE experiments were not carried out for a time sufficient to observe the latter phenomena. The same authors measured steadily decreasing osmotic potentials until close to the end of the life cycle (around 100 DAS) regardless of the watering regime. Leaf osmolality (inversely correlated to osmotic potential) was determined at the end of the stress treatment in the second CE experiment (69 DAS, nine days of treatment). Both treatments resulted in slight, statistically not significant increase in the concentration of osmotically active solutes for both landraces (Table 7). It was not possible to unequivocally ascertain the actual extent of osmotic adjustment as osmolality was not measured in fully hydrated tissues. It is expected that leaf relative water content decreases in response to limiting soil moisture, although in bambara

groundnut, relatively high levels are maintained (Collinson *et al.*, 1997). Thus, it seems likely that the little changes were caused by concentration rather than active accumulation of osmolytes. The role of osmoprotection in terms of yield stability is not fully understood due to the high number of factors influencing yield and the dynamic nature of osmotic adjustment (Blum, 1996; Serraj & Sinclair, 2002). Most evidence on the benefit of osmotic adjustment was shown in roots, where enhanced levels of osmolytes brought about greater soil moisture extraction (Morgan & Condon, 1986; Ludlow *et al.*, 1990). Furthermore, osmoprotective mechanisms seem to be mainly effective when severe dehydration occurs, where mere plant survival is of more importance than sustained growth and yield production (Chaves *et al.*, 2003).

Subsequent to the stress phase in the second CE experiment, leaves were detached and subjected to dehydration for one hour. This was done in order to investigate the effects of the water-deficit treatment on photosynthetic parameters. Chlorophyll fluorescence measurement was carried out to determine photosynthesis yield (Y) in light-adapted leaves, i.e. the ratio between variable fluorescence (F'_v), which is the difference between maximal fluorescence (F'_m) induced by a saturating light flash and the ground state (F'_0), and maximal fluorescence. This value provides an estimate of the maximum efficiency of photosystem II reaction centres regardless of the actual photosynthetic activity influenced by CO₂ availability through stomatal movements (Baker & Rosenqvist, 2004). All leaflets, irrespective of the landrace and the watering regime, showed initial Y values close to 0.8 (Fig. 18). This indicated the presence of healthy leaves (Öquist *et al.*, 1992) and demonstrated that the 2008 experimental conditions did not result in too severe stress that could damage the photosynthetic apparatus. With progressing dehydration, chlorophyll fluorescence steadily diminished, thus being a sign of increasing photoinhibition. After around 18 minutes, an effect of the pre-treatment became apparent. When plants had been grown under water-limiting conditions for eight or nine days, respectively, photoinhibition occurred less strongly than in untreated reference plants. Hence, mechanisms must have been activated in both landraces that allowed enhanced protection of reaction centres. A landrace-specific difference was only obvious in non-challenged leaves, where curves separated after about 40 minutes. DipC, originating from a drought-prone area, was less sensitive to damage than its counterpart LunT. As photoinhibition is thought to be caused by ROS produced in the absence of available water (Giardi *et al.*, 1996), two not mutually exclusive explanations come into account. The drought-adapted landrace may possess a more powerful constitutive antioxidant system, and leaf water may be available for a longer time so that ROS formation is delayed. Stomatal movements were reflected by simultaneously recording leaf temperatures (Fig. 18). Decreasing temperatures in both unstressed controls are to be interpreted as transpirational cooling due to open stomata and were visible for around five minutes. In the following approximately ten minutes, temperatures strongly increased which indicated stomata closure. Stressed DipC leaflets showed this effect in a milder form, which means that this landrace maintained a certain degree of gas exchange even under water-limiting conditions. Thus, carbon fixation and leaf

cooling may continue possibly at the expense of water loss. However, bambara groundnut has the potential to keep up high turgor pressure despite declining water availability (Collinson *et al.*, 1997). These authors used a landrace from Zimbabwe, which is probably comparable to DipC as climatic conditions are similar to Botswana and genetic diversity is generally low in Southern Africa (see above). In contrast, leaf temperatures in LunT indicated that the landrace adapted to a humid environment (high rainfall and coastal fog) responded to drought by complete stomata closure. Furthermore, sloppy leaves pointed towards declining turgor pressure.

Considered independently, it is difficult to estimate the benefits of these physiological traits with respect to water-deficit stress tolerance as plants under stress are always faced with the dilemma of production or protection. Therefore, yield serves as the most useful parameter to define tolerance (Passioura, 1996). Even though the CE experiments were not specifically designed to determine yield, the stress treatment for nine days during the stage of pod filling tended to have impacted yield parameters in the 2008 experiment (Table 7). At final harvest around three months after the stress treatment, DipC seeds were 9% lighter than those from corresponding reference plants. Seed number, however, was slightly higher (not significant), which resulted in unchanged total yield. In LunT, seed weight was not altered, but the number of seeds per plant was reduced by 40% from an already low base level. This difference was not statistically firm, most likely due to a considerable level of intra-landrace diversity. Disregarding two outliers with relatively high yield would offer statistical proof of substantial yield reduction in LunT upon a short phase of water-deficit. Thus, it can be concluded that, as initially hypothesised, the degree of drought tolerance in bambara groundnut landraces is reflected by adaptation to their areas of cultivation. Taking these findings into account, the maintenance of photosynthetically active leaf area due to effective stomatal regulation and turgor retention under water-limiting conditions may be regarded as key features of drought-tolerant bambara groundnut landraces. Root characteristics such as increased length, density and dry matter allocation, certainly equally important parameters (Begemann, 1988; Collinson *et al.*, 1996), were not part of this study, but were unlikely to have influenced results as rooting space was restricted in the presented pot experiments.

4.2.3.2 Differential gene expression under water-deficit stress

Depending on the depth of analysis, MPSS expression profiling revealed that more than half (based on the 2,425 strongest expressed genes) to two thirds (based on the 561 strongest expressed genes) of the genes were termed differentially expressed across four bambara groundnut genotypes in response to reducing the water dosage by 65% for eight days in the first CE experiment. The 561 largest transcript clusters were assigned to eleven functional categories (Fig. 11), spanning virtually the whole spectrum of protein functions.

Photosynthesis and respiration

The 'energy' functional category was the class with the greatest share of differentially expressed genes (96%). With the exception of two genes encoding for a protein of the ubiquinol-cytochrome c reductase complex, which is involved in the mitochondrial respiratory chain (Braun & Schmitz, 1995), and a mitochondrial F1-ATPase subunit, all genes were strongly downregulated in response to water-deficit stress. The vast majority encode for proteins of both photosystem I and II, which are post-translationally assembled by the chloroplast and inserted in thylakoid membranes (Jansson, 1994). Nine transcript clusters (MPSS clusters 8, 26, 32, 53, 55, 61, 75, 85 and 150; Table 17) showed homology to light-harvesting chlorophyll a/b-binding polypeptides which function as coordinators of antenna pigments that absorb light and transfer energy to the photosystem reaction centres (Pullerits & Sundström, 1996). Further genes were assigned to photosystem reaction centres, where charge separation of water takes place, or were less exactly characterised (e.g. MPSS clusters 11, 23, 35, 37, 38, 40 and 45).

Subsequent to the light reaction, this energy is used to fix atmospheric CO₂ through the Calvin cycle. Although functionally and in C₃ plants like bambara groundnut also spatially connected, carbon assimilation is assigned to the 'metabolism' functional category in the scheme used. Table 5 lists seven genes involved in the Calvin cycle as downregulated. Hierarchical clustering of developing gene expression from the second CE experiment displayed highly similar patterns suggesting coordinated expression of the photosynthesis genes investigated (Fig. 19, cluster B). Several studies implicated parallelised regulation of nuclear coded chlorophyll a/b-binding proteins and RuBisCO small subunit (MPSS cluster 9), the best-characterised Calvin cycle enzyme, with the involvement of plastid factors (Taylor, 1989; Rapp & Mullet, 1991). Even though most research has focussed on stress-inducible genes so far, it is recognised that photosynthesis-related genes are downregulated in response to limiting water supply, thereby possibly reducing photooxidative stress (Ingram & Bartels, 1996; Seki *et al.*, 2002; Kavar *et al.*, 2008; Degenkolbe *et al.*, 2009).

A wide quantitative difference was obvious between the two CE experiments. While the 2006 conditions often resulted in delogarithmised repression factors in the double-digit range (Table 17), none of the photosynthesis genes used in the microarray analysis of the 2008 experiment was downregulated below the significance threshold. In conclusion, the temperature difference by 2°C (30°C in 2006 vs. 28°C in 2008) produced milder stress in the second experiment. Furthermore, these repression factors seem to be in relationship with the annual precipitation in the respective areas of landrace origins. South Africa and Botswana are similarly drought-prone regions, and landraces showed the smallest average repression factors for energy-related genes (5.3 and 7.1 for AS-17 and DipC, respectively). Swazi Red is grown in place with more than twice the annual rainfall of Botswana and was characterised by higher repression factors (13.4). Strongest repression was observed in LunT (22.4), which is cultivated in a region with more than 3.000mm of annual precipitation. Therefore, the degree

of downregulating photosynthesis-related genes may not only serve as an indicator of the severity of water-deficit stress, but could also be used a selection criterion in terms of drought-tolerance, particularly in view of the importance of maintained photosynthesis under water-limiting conditions as mentioned above.

After photosynthetic carbon fixation, glucose can be subjected to glycolysis, an anaerobic process that generates energy in the form of ATP (Plaxton, 1996). Seven genes involved in this pathway were found to be transcriptionally downregulated as was the case for photosynthesis genes in the 2006 experiment. The four genes chosen for microarray analysis (MPSS clusters 19 – fructose biphosphate aldolase, 49 and 398 – glyceraldehyde 3-phosphate dehydrogenase subunits and 162 – phosphoglycerate kinase) also showed expression profiles similar to those of photosynthesis genes under milder stress conditions (Fig. 19, cluster B). In contrast, the literature mainly reports induction of glycolytic enzymes under drought (Espartero *et al.*, 1995; Rizhsky *et al.*, 2002; Chaves *et al.*, 2003). However, downregulation has also been described in potato under slowly progressing water-deficit (Schafleitner *et al.*, 2007). Pyruvate generated through glycolysis is further respired in the Krebs cycle, again producing energy and providing carbon skeletons for the biosynthesis of amino acids and several metabolites (Fornie *et al.*, 2004). The two genes discovered by MPSS expression profiling, clusters 320 – nucleoside diphosphate kinase and 349 – glyoxysomal malate dehydrogenase, were repressed after one week of severe water-deficit stress. However, regarding the time course of expression in 2008, the latter gene showed slight upregulation after one day and downregulation with proceeding stress. This suggests that the Krebs cycle is controlled by mechanisms different from those of photosynthesis and glycolysis. In summary, respiratory energy generation is reduced in bambara groundnut under water-limiting conditions at least in the medium term. This is further confirmed by the downregulation of two ATP synthase genes (clusters 229 and 373), two ferredoxins (44 and 223) and a Rieske iron-sulfur protein precursor (cluster 88), which play a role in the mitochondrial respiratory chain (Castresana & Saraste, 1995).

When the availability of CO₂ becomes limited due to closed stomata, RuBisCO may shift its properties towards an oxygenase and activate the photorespiratory pathway, leading to a net loss of carbon and ATP. An intermediate of this pathway is glycolate, which converted to glyoxylate by glycolate oxidase, releasing H₂O₂ (Asada & Takahashi, 1987). The gene encoding for this enzyme (MPSS cluster 69) was downregulated in the 2006 experiment, which possibly reduced the levels of toxic H₂O₂. A similar observation was made in drought-stressed pea leaves (Moran *et al.*, 1994). After conversion of glyoxylate to glycine, serine is produced through the glycine cleavage complex, represented by two genes in the MPSS dataset (clusters 187 and 242), and serine hydroxymethyltransferase (cluster 144) (Oliver, 1994). These genes were again substantially downregulated in 2006, while the latter showed no response to water-deficit in 2008. Thus, while photorespiration can also exhibit a stress-protective function through the dissipation of excess energy, which may lower the danger of photoinhibition (Wingler *et al.*, 2000), its components are transcriptionally repressed under

severe water-deficit stress in bambara groundnut, suggesting the avoidance of its negative effects as ROS production and energy depletion.

Carbohydrate metabolism

Seven differentially expressed genes involved in carbohydrate metabolism were identified, of which six were upregulated after seven days of relatively severe water-deficit stress (Table 5). Two β -amylase genes, encoding for chloroplast-targeted (MPSS cluster 114) and β -amylase 1 (cluster 224) degrade starch into maltose and have been described to be activated under drought in many plant species (Yang *et al.*, 2001; Seki *et al.*, 2002; Boominathan *et al.*, 2004). Maltose may function as a compatible solute to protect stromal proteins and the functionality of the thylakoid membrane for the photosynthetic electron transport chain (Kaplan & Guy, 2004). Maltose can then be further metabolised to glucose units and/or sucrose via cytosolic glucosyltransferases (Chia *et al.*, 2004; Kaplan *et al.*, 2006). MPSS cluster 563 may function in this way and was termed generally upregulated regarding the average expression changes in four bambara groundnut landraces. Both sucrose and monosaccharides may function as osmoprotectants (Keller & Ludlow, 1993; Pinheiro *et al.*, 2001). Sucrose is phloem-transported sugar in most higher plants, whereas it cannot be used for metabolic processes unless it is cleaved to hexoses by invertases (Kim *et al.*, 2000). The expression of a putative invertase inhibitor gene (cluster 166) was found to be induced upon water-deficit treatment. Thus, this enzyme may be involved in carbon mobilisation from source/storage tissues to sink organs like shoot tips and developing pods. Such imbalances on the whole-plant level may in turn affect the regulation of photosynthesis (Paul & Foyer, 2001). Another catabolic enzyme may be encoded by MPSS cluster 34. However, its annotation as an endo-1,4- β -mannanase is dubious as it has mainly been described to play a role in galactomannan mobilisation from germinating seeds (Spyropoulos & Reid, 1988) and, though well-studied in many leguminous species, only one BLAST hit was obtained.

myo-Inositol is synthesised from glucose 6-phosphate through *myo*-inositol 1-phosphate synthase (cluster 457) and is the precursor of pinitol (Streeter *et al.*, 2001; Gomes *et al.*, 2005), which is the one of the main solute accumulating in drought-stressed pigeonpea leaves (Kellwe & Ludlow, 1993). While upregulated in chickpea under water-deficit conditions (Boominathan *et al.*, 2004), the *myo*-inositol 1-phosphate synthase gene was downregulated at least in non-drought-adapted bambara groundnut. Clearly upregulated in all genotypes was the expression of the *myo*-inositol oxygenase 5 gene (cluster 152). This enzyme catalyses the oxidation of *myo*-inositol to glucuronate, which forms part of plant pectins and hemicelluloses (Loewus & Murthy, 2000) and may thus contribute to increased cell wall stability. However, Lorence *et al.* (2004) suggested *myo*-inositol oxygenase to play a role in the biosynthesis of ascorbate, the major antioxidant in plant cells.

Lignin metabolism

Three differentially expressed genes are presumed to function in the biosynthesis of lignin. Being another major component of plant cell walls, it may confer mechanical strength to conducting vessels to withstand the negative pressure generated by transpiration (Vincent *et al.*, 2005). Lignin is derived from polymerisation of three different monolignol units, and the last step in their formation is catalysed by cinnamyl alcohol dehydrogenase (CAD). MPSS cluster 102 showed high homology to the CPRD14 gene from cowpea, which is strongly induced by dehydration and high-salinity stress, but not by cold, heat or exogenous ABA (Iuchi *et al.*, 1996) and in turn is highly homologous to CAD genes from various species. Due to its specific role in the monolignol biosynthetic pathway, CAD has been considered to be an indicator of lignin biosynthesis (Walter *et al.*, 1988; Sibout *et al.*, 2005). MacKay *et al.* (1997) showed that a CAD-deficient maize mutant has decreased lignin content. Its strong transcript accumulation in the 2006 CE experiment, which represented severe water-deficit stress, and induction in the early stages of the milder stress experiment therefore point towards increased lignin biosynthesis in bambara groundnut.

MPSS cluster 112 followed the same expression patterns as CAD, as indicated by induction factors (Table 8) and hierarchical clustering (Fig. 19, cluster E). Annotated as a low temperature and salt responsive protein from potato, it also shows homology to a strawberry *o*-methyltransferase mRNA. While *o*-methyltransferases are involved in numerous biochemical pathways, this one may methylate caffeate or 5-hydroxyferulate to produce ferulate or sinapate, respectively (Whetten & Sederoff, 1995) and thus lead to the synthesis of coniferyl alcohol and/or sinapyl alcohol monolignol units instead of *p*-coumaryl alcohol, which is the preferred pathway in angiosperms (Boerjan *et al.*, 2003). However, the above-named enzyme properties are usually attributed more specifically to caffeic acid methyltransferase (COMT; Whetten & Sederoff, 1995). The corresponding gene (cluster 99) was strongly downregulated in bambara groundnut under water-deficit stress according to MPSS profiling and tended to follow a pattern of downregulation different from the one of photosynthesis genes (Fig. 19, cluster A). COMT has been associated to the *brown midrib3* mutation in maize (Vignols *et al.*, 1995) and its antisense suppression resulted in decreased lignin content and altered lignin composition (Piquemal *et al.*, 2002). In summary, lignin biosynthesis seems to change in bambara groundnut upon reduced irrigation, possibly resulting in *de novo* synthesis of an altered form of stress lignin.

Protein synthesis and modification

Compared to other functional categories, genes grouped under ‘protein synthesis’ were less affected by the water-deficit treatments. Only one third (16/48) of the genes were termed generally differentially expressed over four bambara groundnut genotypes in the first CE experiment. mRNAs similar to a SUI1 translation initiation factor (MPSS cluster 92) and a ribosomal protein (cluster 433) slightly accumulated in response to severe water-deficit.

Thirteen of 14 downregulated genes encode for ribosomal proteins. However, repression factors were usually below the range of photosynthesis genes, e.g. in clusters 33, 237, 304 and 348. None of the genes chosen for microarray analysis proved to show significantly altered expression in the 2008 CE experiment. Consequently, with the exception of ribosomal protein S27, which tended to cluster with early induced genes (Fig. 19, cluster E), ribosomal protein genes did not fall into specific expression groups. The literature is ambiguous: Several studies reported upregulation of ribosomal proteins under water-deficit or osmotic stress; downregulated genes were, however, not regarded (Boominathan *et al.*, 2004; Wang *et al.*, 2007; Reddy *et al.*, 2008). The same seems to apply for elongation factors and translation initiation factors. Therefore, protein synthesis-related genes may be relatively stably expressed as housekeeping genes and are little influenced by at least mild water-deficit stress.

The proportion of differentially expressed genes involved in the 'protein fate' functional category, i.e. the modification or degradation of proteins, was also comparatively low at 32%. However, regarding the averaged expression profiles from four genotypes (Fig. 11), most of the differentially expressed transcripts exhibited enhanced accumulation after one week under water-limiting conditions in the 2006 experiment. Five of these genes (among the 561 largest transcript clusters) were attributed to proteolysis. This process involves the removal of abnormal or misfolded proteins and recycling the resulting amino acids to the plant nitrogen cycle, which is essential for plant adaptation to environmental conditions (Vierstra, 1996). Increased activity and expression of proteolytic enzymes has been observed in various legumes under drought stress (Cruz de Carvalho *et al.*, 2001; Hieng *et al.*, 2004; Reddy *et al.*, 2008). Ubiquitination assumes a central position in the proteolytic pathway (Vierstra, 1996). MPSS clusters 300 and 570 fulfil this task and were upregulated in the 2006 experiment, and another gene (ubiquitin-conjugation enzyme, cluster 77) was significantly induced in the 2008 experiment. Proteins destined for breakdown are selectively tagged by ubiquitin and can then be degraded by proteases. Genes encoding for a cysteine protease (cluster 131), a protease precursor (cluster 472), a zinc-dependent protease (cluster 493; all in the 2006 experiment) and another cysteine protease (cluster 86; 2008 experiment) were found upregulated upon water-deficit in bambara groundnut. Expression kinetics was similar for all genes spotted on the microarray. Early induction was followed by approaching or even falling below the levels of unstressed plants with progressing adaptation to water-limiting conditions. This indicated the importance of diversified proteolytic regulation depending on different stages of water-deficit stress. Furthermore, one ubiquitin-like protein mRNA (cluster 550) tended to be repressed by severe stress and upregulation was also observed for a protease inhibitor gene (cluster 413) counteracting protein degradation. Considering the fact that protein modification may also affect transcription factors and cell receptors, it is obvious that it is part of a highly complex cascade of regulatory events controlling a plethora of metabolic processes (Vierstra, 1996).

Cell protection

Enhanced accumulation of five different LEA transcripts was revealed through MPSS expression profiling (Table 6). All genes were expressed at high levels after applying reduced irrigation for one week, which is indicated by low cluster numbers (2, 4, 36, 52 and 71). MPSS cluster 2 is highly similar to a drought-inducible gene from cowpea that also responds to salt stress and exogenous ABA, but not to cold and heat stress (Iuchi *et al.*, 1996). Further sequence alignment identified this gene encoding for a dehydrin, also known as group 2 LEA proteins (93% nucleotide identity). While expressed around the detection limit under well-watered conditions, transcript concentrations rose to between 16,833 TPM in AS-17 and 48,950 TPM in LunT. This clearly makes these mRNAs the most abundant in the stress dataset and supports the assumption that LEA proteins comprise the vast majority of stress-responsive proteins (Bartels & Sunkar, 2005). In contrast to the consistent expression profiles under severe water-deficit, the second CE experiment demonstrated different patterns of expression under milder conditions. MPSS clusters 52 and 71 accumulated mainly during the first two days after imposing reduced irrigation (Fig. 19, cluster E), whereas clusters 2 and 36 were upregulated throughout the eight-day stress period (Fig. 19, cluster D). Surprisingly, cluster 4 did not emerge to be induced at all but clustered together with photosynthesis genes (Fig. 19, cluster B). Detailed biochemical interpretation is difficult due to limited knowledge about the function of LEA proteins (Wise & Tunnacliffe, 2004). Dehydrins (MPSS clusters 2, 4 and 71) may be capable of inhibiting the coagulation of protein and membrane structures, thereby preserving structural integrity under water-limiting conditions (Close, 1997). Many studies reported a positive correlation between enhanced expression of LEA proteins/dehydrins and drought-tolerance, e.g. in sunflower (Cellier *et al.*, 1998), in winter wheat, where dehydrin accumulation is related to the maintenance of greater shoot dry matter production (Lopez *et al.*, 2003) or in barley, where QTLs for drought-tolerance are associated with dehydrin genes (Forster *et al.*, 2004). Furthermore, experiments using transgenic plants corroborate the beneficial role of LEA proteins in terms of performance under water-deficit stress conditions (Cheng *et al.*, 2002; Bahieldin *et al.*, 2005; Park *et al.*, 2005).

Similarly to LEA proteins, heat shock proteins (HSPs) have also been shown to accumulate in response to dehydration stress (Almoguera *et al.*, 1993; Seki *et al.*, 2002; Senthil-Kumar *et al.*, 2007) and to act as molecular chaperones ensuring correct folding and assembly of proteins (Wang *et al.*, 2004). Four genes with homology to HSP sequences, MPSS clusters 158, 209, 404 and 451, were found upregulated in the 2006 CE experiment (Table 6). For the three genes used for microarray analysis, enhanced expression at at least one time point was confirmed in the 2008 experiment (Table 10). However, MPSS expression profiling also identified two HSPs downregulated upon severe water-deficit stress (clusters 312 and 356). This may be attributed to the involvement of HSPs in a magnitude of plant developmental processes that are not necessarily related to drought (Vierling, 1991).

Three different drought stress-induced transcripts were identified as lipid transfer proteins (LTPs; Table 6). One of them, MPSS cluster 5, was characterised by 37-fold change as averaged over four genotypes, while two (clusters 24 and 250) accumulated less dramatically (Table 17). The former, LTPI, exhibited equally strong induction throughout the entire stress period in the moderate water-deficit experiment (Table 10). In contrast, the expression of LTPII was not significantly altered. Yet again, this suggests differential regulatory control of genes with putatively similar functions and also implies varied significance for the water-deficit reaction in bambara groundnut. In general, LTPs were proposed to be capable of transporting cutin monomers to sites of cutin synthesis in or on the outside of the cell wall (Sterk *et al.*, 1991). Thus, LTPs may be involved in the formation of epicuticular wax (Kader, 1997), which may lead to increasing cuticle thickness and has been described as one mechanism to protect plants from water loss under water-limiting conditions (Ramanjulu & Bartels, 2002). Romo *et al.* (2001) have shown that LTPs preferentially accumulate in young tissues and during early developmental stages in osmotically stressed chickpea, but that they are insensitive to ABA treatment.

Detoxification

Another three highly expressed genes, MPSS clusters 12, 13 and 21, encode for metallothioneins. These heavy metal-binding proteins are ubiquitous in both eukaryotes and prokaryotes (Palmiter, 1998) and have been shown to respond to a broad range of environmental stresses in plants (Hsieh *et al.*, 1995), including water-deficit (Seki *et al.*, 2002; Ozturk *et al.*, 2002; Boominatan *et al.*, 2004). Due to their metal-binding properties, it has been suggested that metallothioneins participate in maintaining the homeostasis of physiological metals like copper and zinc, but also in the detoxification of toxic metals such as cadmium and arsenic (Roosens *et al.*, 2004; Zimeri *et al.*, 2005). Consequently, this involvement in the micronutrient balance may have an impact on cofactor-requiring proteins. Recently, Yang *et al.* (2009) have demonstrated enhanced accumulation of several zinc finger transcription factors in transgenic rice overexpressing a metallothionein gene. Furthermore, metallothioneins are capable of scavenging ROS in drought-stressed plants (Akashi *et al.*, 2004; Xue *et al.*, 2009). It is yet unclear whether this antioxidant function directly relies on metallothioneins or on the activation of further antioxidant enzymes (Yang *et al.*, 2009). Nevertheless, there is growing evidence that metallothioneins confer increased tolerance under water-limiting conditions as reported in transgenic studies (Xue *et al.*, 2009; Yang *et al.*, 2009). The three metallothioneins in bambara groundnut behaved distinctly in response to severe water-deficit stress. A type 2 metallothionein (MPSS cluster 12) displayed a mean twofold induction, while type 1 metallothioneins (clusters 13 and 21) were repressed after one week. When applying milder stress in 2008, the former was not responsive, and MPSS cluster 21 tended to accumulate early upon treatment and to be repressed in later stages including rehydration (Table 10). Such diverging expression profiles were also observed in challenged

barley (Ozturk *et al.*, 2002) and may be an indicator for different physiological functions of various metallothionein types, which is supported by different organ-specificities (Zhou & Goldsbrough, 1994).

Ferritin (cluster 261) is another metal-chelating protein that may be involved in the defence reaction by sequestering iron ions which are required for the generation of hydroxyl radicals through the Fenton reaction (Deák *et al.*, 1999). The corresponding transcripts accumulated in bambara groundnut after severe water-deficit, whereas on the other hand, milder stress led to significant decreases in the late stress and recovery stages (Table 10). This may, beside their detoxifying properties, underpin the importance of metal-binding proteins in regulating cellular cofactor concentrations.

The expression of the “classical” detoxifying enzymes catalase (cluster 66), glutathione peroxidase (GPX, cluster 130), glutathione S-transferase (cluster 228) and ascorbate peroxidase (APX, cluster 283) did not substantially change in the 2006 CE experiment. Average upregulation of APX and glutathione S-transferase was mainly caused by single genotypes. The time course of catalase and GPX gene expression were monitored in the 2008 experiment. GPX exhibited clear induction up to two days after imposing reduced irrigation. Thereafter, no significant difference to the unstressed reference was observed, but a tendency towards downregulation after rehydration. Catalase showed a similar profile, however statistically insignificant during the stress period (Fig. 19, cluster C). Thus, the antioxidant enzymes mentioned above may be involved in the immediate response to limited water supply, but are likely to play a minor role in the medium to long term drought adaptation in bambara groundnut.

The case was different for the copper/zinc-superoxide dismutase gene (SOD; cluster 454). A number of studies reported the activation of SODs under water-deficit stress (Ingram & Bartels, 1996; Acar *et al.*, 2001; Borsani *et al.*, 2001). On the contrary, all bambara groundnut genotypes investigated displayed considerably reduced expression after one week of severe water-deficit, as revealed through MPSS and microarray analysis (Table 18). Likewise, the second CE experiment brought about significant gene repression (Table 10). Literature search did not yield information about the downregulation of SODs under drought, but global gene expression projects also do not report SOD induction (Seki *et al.*, 2002; Ozturk *et al.*, 2002; Boominathan *et al.*, 2004; Senthil-Kumar *et al.*, 2007; Reddy *et al.*, 2008). In the light of metal cofactor regulation and possibly reduced ROS pressure through downregulated photorespiration as described above, the repression of the Cu/Zn-SOD in bambara groundnut appears more plausible.

ROS also mediate lipid peroxidation in biological systems, which leads to the formation of toxic aldehydes (Bartels & Sunkar, 2005). Transcripts encoding for an aldo/keto reductase (cluster 254) accumulated after one week of severe water-deficit stress. This enzyme may reduce aldehydes to alcohols, as supported by transgenic plants (Oberschall *et al.*, 2000). The resulting alcohols may then be detoxified by alcohol dehydrogenase (cluster 239). Both CE experiments showed induced expression in response to water-deficit. The same is reported in

numerous publications (De Bruxelles, 1996; Seki *et al.*, 2002; Boominathan *et al.*, 2004; Senthil-Kumar *et al.*, 2007; Yue *et al.*, 2008).

Signalling and transcriptional control

While there exist ABA-independent pathways of gene regulation, ABA-mediated signalling plays a crucial role in many aspects of plant water-deficit responses (Shinozaki & Yamaguchi-Shinozaki, 1997; Bray, 2002). The best-characterised gene in ABA biosynthesis encodes for the key regulator 9-*cis*-epoxycarotenoid dioxygenase (NCED; Qin & Zeevaart, 1999; Iuchi *et al.*, 2000). However, neither this nor any other gene involved in ABA biosynthesis was identified in bambara groundnut. While the general assumption is that ABA is primarily synthesised in roots as the primary sensors of drought and then transported to the shoot (Christmann *et al.*, 2005), enhanced expression of NCED has been reported in leaves of several species (Qin & Zeevaart, 1999; Iuchi *et al.*, 2000; Thompson *et al.*, 2000). It is possible that the absence of NCED is due to the technical limitations of the MPSS approach (see 4.2.1). However, as NCED is expressed very early after dehydration (Iuchi *et al.*, 2000), sampling after one week of stress may have been too late and the differential gene expression profiles obtained may be the result of increased mRNA stability or ABA-independent gene regulation. Furthermore, it is conceivable that ABA inhibits its own biosynthesis in a feedback mechanism (Bartels & Sunkar, 2005).

Another signalling pathway is triggered by ethylene. It has been shown that ethylene interacts with ABA (Morgan & Drew, 1997; Sharp & LeNoble, 2002). This implicates its involvement in the water-deficit stress response in plants. 1-Aminocyclopropane-1-carboxylic acid oxidase (MPSS cluster 185) catalyses the final step in ethylene biosynthesis (John, 1997). Both CE experiments resulted in significant decrease of this enzyme (Table 8, Table 17). The literature is ambiguous about the regulation of ethylene biosynthesis under water-limiting conditions (Morgan & Drew, 1997). As 1-aminocyclopropane 1-carboxylic acid oxidase requires iron ions as cofactors (John, 1997), its activity may be influenced by ferritin, which may also explain coexpression of these genes (Fig. 19, cluster A).

Five out of eleven genes assigned to the 'cellular communication/signal transduction mechanism' category were termed generally differentially expressed in the 2006 experiment. Strong accumulation was observed for a nodule-enhanced protein phosphatase type 2C mRNA (cluster 269). Reversible phosphorylation of proteins is a common way of rapid post-transcriptional regulation of transcription factor activities in response to extracellular stimuli (Karin & Hunter, 1995). This specific gene has been suggested to be involved in initiating nitrogen fixation through plant-microbe interaction in *Lotus japonicus* roots (Kapranov *et al.*, 1999). As it is not known whether this influence is positive or negative, and gene expression was only investigated in dehydrating leaves, the impact of water-deficit on nodule formation remains speculative. A related gene found to be induced after severe water-deficit treatment encodes for a serine/threonine protein kinase (cluster 471). Studies on an ABA-insensitive

Arabidopsis mutant identified a missense mutation in this gene (Meyer *et al.*, 1994), and the resulting disrupted regulation of guard cell ion channels negatively affects stomata closure (Armstrong *et al.*, 1995). By contrast, a CDPK-related protein kinase gene (cluster 143) exhibited strong downregulation in 2006 (Table 17) and slight (insignificant) downregulation upon milder water-deficit. The involvement of CDPKs in Ca²⁺-mediated signalling in osmotic stress is well-documented (Shinozaki & Yamaguchi-Shinozaki, 1997; Sanders *et al.*, 1999; Ramanjulu & Bartels, 2002). Water-deficit stress experiments usually report increasing expressions (Seki *et al.*, 2002; Ozturk *et al.*, 2002). The different results in bambara groundnut suggest a much more complex network of signal transduction involving further protein kinases not discovered through MPSS gene expression profiling.

In addition, 14-3-3 proteins can interfere with signalling pathways by binding to CDPKs and MAPKs (Fanger *et al.*, 1998; Camoni *et al.*, 1998). One such gene (cluster 486) was repressed upon severe water-deficit stress (Table 17), while another gene (cluster 412) did not respond significantly to both treatments. Post-transcriptional control of gene expression may be modulated by RNA-binding proteins. The levels of two RNA-binding protein mRNAs (clusters 117 and 301) were decreased in the 2006 experiment. The former gene was chosen for microarray analysis and downregulation was confirmed in the second CE water-deficit experiment (Table 11). Another RNA-binding protein gene (cluster 377) was induced in 2006, but did not show significant expression difference upon milder stress. Lectin mRNA (cluster 325) accumulated in both experiments. Due to its carbohydrate-binding properties, it may be involved in sugar-mediated signalling (Sharon & Lis, 1989; Rolland *et al.*, 2002).

Among the 561 largest transcript clusters, six were homologous to transcription factors. On an average over the four genotypes, three were termed upregulated and two downregulated after one week of severe water-deficit. However, with the exception of an ethylene-responsive protein mRNA (cluster 313), which was not differentially expressed, all transcription factors identified through 454 sequencing and BLAST search, i.e. also those with transcript abundance lower than 40 tags in the dataset, were included in microarray analysis. Three transcription factors share homology with MYB transcription factors. MYB proteins are involved in an ABA-dependent pathway of gene expression under water-deficit stress (Shinozaki & Yamaguchi-Shinozaki, 1997), as evidenced by dehydration experiments (Urao *et al.*, 1993; Abe *et al.*, 1997). Stomatal opening in *Arabidopsis* has recently been described to be influenced by MYB transcription factors (Cominelli *et al.*, 2005; Liang *et al.*, 2005). In bambara groundnut, the MYB139 gene (cluster 227) was strongly induced in the 2006 experiment. Analysis of expression dynamic in the second CE experiment suggested MYB139 being co-expressed with three uncharacterised genes (MPSS clusters 82, 154 and 418) and several genes assigned to the 'cell rescue, defence and virulence' functional category, i.e. two HSPs, alcohol dehydrogenase and a Snakin-like cysteine-rich protein (Fig. 20, DipC cluster III and LunT cluster IIIb). Further genes, mainly involved in protein synthesis, also appeared to be co-expressed with MYB in DipC, but were grouped in a different cluster (IIIa) for LunT. MYB123 (cluster 555) displayed no clear common

expression signature in the MPSS approach, possibly due to the generally low expression level, but microarray validation of the 2006 experiment suggested downregulation for all genotypes (Table 18). k-Means clustering assigned its expression pattern to cluster Va for both landraces used in the 2008 experiment. Including clusters Vb, which are only marginally different from Va, it appears that many genes co-expressed with MYB123 may function in cellular communication processes such as binding to proteins or other factors (MPSS clusters 139 - acireductone dioxygenase/metal ion-binding protein, 325 - lectin, 412 - 14-3-3 protein and 551 - selenium-binding protein), involvement in transport routes (275 - ras-related GTP-binding protein, 307 - cytochrome-c oxidase and 357 - small GTP-binding protein), modification of proteins (70 - ubiquitin-like protein, 157 - isoprenylated protein and 617 - cysteine proteinase precursor) or ethylene signalling (496 - S-adenosylmethionine decarboxylase). In both landraces, these genes quickly responded to reduced irrigation by slight upregulation for the first two days and had unaltered or slightly decreased levels afterwards. Expression changes were usually below the statistical significance thresholds. Early induction may also explain why most of these genes were not found to be differentially expressed in the 2006 experiment, where sampling was carried out seven days after initiating the water-deficit treatment. In the literature, the MYB123 transcription factor has been described as a regulator of proanthocyanidin biosynthesis induced by various biotic and abiotic stresses in poplar (Mellway *et al.*, 2009). Another MYB transcription factor, MYB93 (MPSS cluster 811), was 4.4-fold upregulated on average over the four genotypes used in the 2006 experiment. In 2008, no significant expression difference to the unstressed reference was obvious at any time point. While the literature reports the expression of MYB93 being linked to cold and heavy metal stress (Ogundiwin *et al.*, 2008; Liu *et al.*, 2009), its relevance in water-deficit stress appears to depend on the degree of dehydration.

Homeodomain-leucine zipper (HD-ZIP) proteins represent another class of transcription factors responsive to dehydration (Söderman *et al.*, 1999; Dezar *et al.*, 2005). MPSS cluster 766 is highly homologous to a HD-ZIP transcription factor activated in virus-inoculated soybean leaves (Wang *et al.*, 2005). Its dynamic expression followed the pattern of MYB139. Thus, it would stand to reason that both transcription factors are regulated by a common signal.

Two genes encode for basic region leucine zipper (bZIP) transcription factors. Similarly to MYB transcription factors, these were also shown to function in the regulation of ABA-inducible genes (Shinozaki & Yamaguchi-Shinozaki, 1997). *Cis*-acting factors involved in the transcriptional regulation by ABA (ABREs) are bound by bZIP transcription factors (Bray, 2002). In bambara groundnut, bZIP33 (MPSS cluster 393) was found to be upregulated after one week of severe water-deficit. Yet no significant change was recorded when milder water-deficit stress was applied. Thus, the course of gene expression does not allow clear assignment to other genes. The bZIP6 gene (cluster 823) was induced upon water-deficit in both experiments (Table 9, Table 18). Beside five uncharacterised genes, two LEA protein

genes (clusters 52 and 71) and two genes of the carbohydrate metabolism (clusters 114 – β -amylase and 166 – invertase inhibitor) were coexpressed with the bZIP6 transcription factor. The expression of one member of a fourth class of transcription factors, a zinc-finger B-box protein mRNA (cluster 438), was also induced upon severe water-deficit. Similarly to MYB93 and bZIP33, the treatment in 2008 did not result in a stress intense enough to see significant differential expression. Nevertheless, a number of studies have pointed out the involvement of zinc-finger proteins in coordinating the complex molecular response to water-deficit (Sugano *et al.*, 2003; Mukhopadhyay *et al.*, 2004; Sakamoto *et al.*, 2004).

4.2.3.3 Differential gene expression between landraces

Single genes

As described above, the response mechanisms to limiting irrigation are highly similar in contrasting bambara groundnut landraces and, except from few aspects, have been described in numerous water-deficit challenged plant species. Nevertheless, physiological observations exemplified by comparing a landrace adapted to a drought-prone region, DipC, and one from an environment characterised by high moisture levels, LunT, suggested differences in terms of stomatal movements and yield formation (see 4.2.3.1). In order to identify disparities at the transcriptional level, induction or repression factors, respectively, obtained through MPSS gene expression profiling were compared. With respect to differences among the four genotypes used, a gene was termed differentially expressed when at least one landrace individual displayed a more than twofold change and at least another one had an inverse expression factor at the same time. This resulted in one third of the 561 largest transcript clusters carrying a potential landrace-specific difference. As shown in Fig. 12a, the number of genes assigned to protein synthesis was disproportionately high. Regarding only differences between the landraces used for the second CE experiment, the percentage was reduced to 8.6. Out of these 48 genes, 18 were chosen for microarray analysis. When contrasting \log_2 transformed expression ratios obtained through MPSS and microarray analyses (Table 18), a landrace difference between DipC and LunT was only confirmed for one gene (MPSS cluster 189). As mentioned above, direct comparison of both methods is difficult due to missing replication and non-uniform genotypes. Furthermore, the frequency of expression differences detected through MPSS increased with diminishing numbers of tags per cluster, when reliability of MPSS expression profiling became questionable. Thirdly, the greatest fraction of annotated genes was assigned to protein synthesis regarding differential expression profiles within all four genotypes as well as between DipC and LunT (Fig. 12a, b). These genes were found to show a relatively weak response to water-deficit stress. Therefore, it is more likely that variation around the induction/repression threshold occurred. Consequently, genes with a strongly developed reaction to severe water-deficit stress, i.e. photosynthesis and stress- and defence-related genes, were clearly underrepresented. By combining each two genotypes from

dry (DipC and AS-17) and from humid areas (LunT and Swazi Red), it was attempted to substantiate the MPSS expression profiles with respect to climatic adaptation. Four genes out of the nine clusters presented in Fig. 13 were printed on the microarray. Again, only MPSS cluster 189 encoding for a valine transfer-RNA proved to be differentially expressed between genotype groups after microarray validation of the 2006 experiment. However, biological relevance in the stress-adaptation of plants remains obscure. When milder water-deficit was applied in 2008, statistically significant differences were detected neither between treatments nor between genotypes.

Due to the problems associated with MPSS expression profiling for transcripts of lower abundance and relatively late sampling in 2006, the 2008 experiment may have yielded more meaningful results. Nine genes were significantly higher expressed in stressed LunT (Table 13) at at least one sampling date. It is conspicuous that all differences but one with insignificant higher expression occurred in the latest stress stage investigated eight days after reducing irrigation. Two of these genes encode for LEA proteins (MPSS clusters 2 and 52) and three are presumably involved in lignin metabolism (clusters 99, 102 and 112). As expression was higher in the putatively drought-sensitive landrace, it appears that these genes reflect plant-intrinsic stress rather than that they explain the better drought-tolerance of DipC. Instead, DipC is likely to feature some drought avoidance mechanism which, after an initial adaptive response, makes the prolonged expression of lignin and some LEA genes unnecessary. The same pattern was observed for an early light-induced protein, which is thought to function in the assembly of the photosynthetic apparatus (Adamska & Klopstech, 1991) and a polyubiquitin 2 gene, which may be involved in the removal of damaged proteins or protein-mediated signalling.

Therefore, structural genes with higher expression in the tolerant landrace are of more value in terms of protection against adverse conditions. Three genes met this criterion at single time points in the treated variant (Table 13). MPSS cluster 20 is similar to a soybean stored cotyledon mRNA, which shares high homology with a defensin gene from *Tephrosia villosa* ($4e^{-71}$, 86% nucleotide identity). Defensins are antimicrobial proteins usually involved in the defence reaction against pathogens (Hanks *et al.*, 2005). However, their induction pathways may share some elements of the water-deficit response, e.g. ethylene signalling (Penninckx *et al.*, 1998). Being repressed at least in the drought-adapted genotypes after one week of severe water-deficit, the 2008 CE experiment showed induction of the defensin gene in the first two days after reducing irrigation (Table 10). One LEA protein gene (cluster 36) surprisingly was expressed 4.5-fold more in DipC than in LunT at the first stress recovery stage. During the stress period, no statistically significant difference was observed, although at stage 4RI, expression was 6-fold higher. At the same stage, the *myo*-inositol oxygenase 5 gene (cluster 152) was expressed twice as strong as in LunT.

MPSS cluster 124, or at least the region sequenced, does not share homology with a published sequence. It was not detected in LunT, neither through MPSS nor microarray technology. As this gene was upregulated after one week of severe water-deficit stress as well as in the first

two days after applying milder stress in DipC (Table 12), it may serve as a useful, potentially novel candidate gene for drought tolerance in bambara groundnut. Cluster 110, another uncharacterised gene, was also induced upon the 2008 treatment in DipC, whereas it was more or less stably expressed in LunT (Table 12). Hence, significant landrace differences were recorded until the fourth day after initiating the stress treatment in 2008 (Table 13). However, microarray analysis of the 2006 experiment suggested strong downregulation of this gene in all four genotypes, which is in contrast to MPSS results. This special case deserves further attention, but does not render this gene less interesting.

Apart from landrace-specific differences under stress, seven genes were found to be constitutively (i.e. in control plants) higher expressed in DipC (Table 13). While for six of these, no clear biological function could be determined, cluster 1, the most abundant mRNA in the full MPSS dataset, encoding for a proline-rich protein precursor, was 2.6-fold higher expressed than in LunT at one time point. A threefold difference between unstressed DipC and LunT plants was also observed in the MPSS profiling experiment. Proline-rich proteins are structural cell wall proteins conferring mechanical strength (Keller, 1993). Therefore, the drought-tolerant bambara groundnut genotype may be able to form more rigid cell walls in the absence of water-deficit, which protect cells from turgor loss when water becomes scarce. However, all seven genes had a statistically significant higher constitutive expression at only one time point in DipC, which means that general evidence is not given.

Expression kinetics

In addition to the analysis of single time points, gene expression curves over four time points of limiting irrigation and two time points after re-applying full irrigation were compared between differently drought-adapted DipC and LunT using k-means clustering (Fig. 20). Genes of clusters I, including two LEA genes, two genes functioning in carbohydrate metabolism and one LTP gene, were found upregulated already one day after initiation of the water-deficit treatment in both landraces. However, the stress reaction was more pronounced in the drought-sensitive landrace, where the average log₂ ratio was 4 while in DipC, this value was 3. Thus it appears that these genes do not actively prevent bambara groundnut from dehydration, but that they are expressed as a response of water-deficit stress, which is perceived more strongly by LunT. In DipC, gene expression steadily increased until four days after reducing irrigation. In contrast, LunT showed maximum expression at stage 2RI, followed by a slump at stage 4RI and a second phase of transcript accumulation eight days after imposing stress. This may imply that the drought-tolerant genotype undergoes a prolonged adaptive phase, possibly involving enrichment of LEA proteins and intensified cuticle development at the expense of stored carbohydrates, after which metabolism is again directed towards regular growth. LunT did show such a fading upregulation at the end of the stress period for the genes of clusters I. Extrapolating this phenomenon, the drought-sensitive landrace was characterised by a higher investment in certain defence-related metabolic

pathways with increasing duration of water-deficit conditions. Though not resolved through k-means clustering, the stress response of three genes (MPSS clusters 36 – LEA5 protein, 182 – cytidine or deoxycytidylate deaminase and 195 – no homology) may have been delayed in DipC, but statistical analysis did not support this assumption.

Similarly, the genes in clusters II were expressed in a different order of magnitude between DipC and LunT during the first two stress stages and tended to show prolonged upregulation in DipC at stage 4RI. However, in contrast to the genes of cluster I, a second phase of upregulation was not observed in LunT. In terms of membership, only MPSS cluster 110 is a considerable outlier as already mentioned above.

The genes grouped under DipC k-means cluster III tended to respond to water-deficit more slowly. Statistically significant differences to the corresponding unstressed reference were usually detected at four days after imposing stress (MPSS clusters 158 and 451 encoding for HSPs and 154 and 418 without homology to a published gene). In LunT, these genes split into two clusters. One (IIIa) was barely responsive to water-deficit, while the other one (IIIb) displayed expression kinetics different from DipC. The first two stress stages were characterised by distinctive induction, whereas at stage 4RI, no difference to the control was obvious. This and corresponding induction factors obtained from the 2006 experiment (Table 14) suggest that these genes may be of particular importance under jeopardising stress, which was probably circumvented by DipC when milder water-deficit was applied in 2008.

Another difference between landraces was obvious when expression curves of downregulated, non-photosynthesis genes, i.e. those of hierarchical cluster A (Fig. 19), were compared. In DipC, these formed a separate k-means cluster IVb, with the strongest water-deficit response again becoming visible at stage 4RI. On the other hand, these genes were grouped together with photosynthesis genes for LunT, which were barely affected by the 2008 conditions. Thus, not only the induction, but also downregulation of genes may be an important factor in distinguishing bambara groundnut landraces with different degrees of water-deficit tolerance. Three of these genes may be involved in signalling pathways (MPSS clusters 117 – RNA-binding protein, 167 – UDP-glucuronosyltransferase and 185 – 1-aminocyclopropane 1-carboxylic acid oxidase), and another two genes possibly exhibit similar function through regulating the levels of ROS (clusters 261 – ferritin and 454 – Cu/Zn-SOD). With respect to photosynthesis genes, no difference between landraces was noticed under mild stress. However, sustained expression appears to be of importance under severe water-deficit which was shown by comparing expression levels after the 2006 treatment (Fig. 12c).

Table 14. Induction/repression factors for six genes and four genotypes obtained through MPSS expression profiling in 2006 for which upregulation was marginal for DipC, but considerable for LunT in the 2008 experiment. Under severe water-deficit stress, induction was more pronounced in genotypes adapted to humid environments.

Gene homology	Arid environment		Humid environment	
	DipC	AS-17	Swazi Red	LunT
82: No homology	30.9	14.0	66.2	83.8
158: Heat shock protein 22	1.1	3.4	15.3	14.4
227: MYB139	15.2	4.2	13.8	34.3
239: Alcohol dehydrogenase 1	1.4	1.6	4.4	1.8
243: Snakin-like cysteine rich protein	2.3	2.2	4.3	16.6
451: Heat shock protein 81-1	-1.3	0.0	-1.3	7.2

k-Means clusters V did not reveal differences between landraces during the stress period. However, genes differed in their expression when full irrigation was resumed. At stage 1REC, downregulation was significant for eight genes of diverse functions (MPSS clusters 21, 66, 93, 212, 262, 323 and 349) in LunT, while DipC did not display significant differences. Upregulation during the first two stress stages implies a role in the adaptive reaction. Hence, a certain degree of constitutive gene expression appears necessary, and LunT seems to neglect this when shifting from stress metabolism back to regular growth under non-limiting water supply. Many genes of clusters V may be involved in signalling processes and are likely to be transcriptionally controlled by MYB123 (see 4.2.3.3, signalling and transcriptional control).

4.3 Conclusion and outlook

4.3.1 DArT markers and genetic diversity

The first part of this study reports the discovery of solid-state, hybridisation-based genetic markers in a microarray format (DArT) specific for bambara groundnut. After resolving a problem so far unobserved with this technology, a full-size genotyping array comprising 7,680 clones was assembled from three different fragment libraries. Hybridisation with targets prepared using two individual enzyme combinations resulted in a total of 658 clones polymorphic between 87 genotypes, of which 296 showed unique segregation signatures. These values were essentially confirmed in a further experiment using an extended screening population. Due to the high locus specificity of DArT markers, these can be easily arranged into genetic linkage maps (Akbari *et al.*, 2006; Wenzl *et al.*, 2006) and will complement the existing initial map based on a subintraspecific cross (wild x cultivated subspecies) and currently containing 81 AFLP and two SSR markers (Basu *et al.*, 2007c; Mayes *et al.*, 2009). Such a medium-density genetic map would then provide the framework for identifying QTLs dissecting quantitative traits, like stable yield under limited water supply, into their single genetic components (Tuberosa & Salvi, 2006). Given that close linkage between markers and QTLs is established, which is relatively plausible as DArT markers preferentially map to gene-rich telomeric regions (Akbari *et al.*, 2006), these diagnostic markers can be utilised for marker-assisted selection (MAS) in breeding programmes. Although in practice, the general impact of MAS in improving drought tolerance has been low so far (Reynolds & Tuberosa, 2008), promising approaches were reported for rice (Steele *et al.*, 2006), maize (Ribaut & Ragot, 2007) and pearl millet (Serraj *et al.*, 2005). MAS can further be used to combine favourable alleles with other, usually more simply inherited traits influencing agronomic performance or end-user acceptance, such as internode length, days to emergence and seed coat colour in bambara groundnut (Basu *et al.*, 2007a). DArT allows for investigation of thousands of loci in parallel and is therefore particularly suited for gene/QTL pyramiding.

Another issue of marker-based whole-genome profiling is the creation of population genetic data. Through fingerprinting of accessions in a germplasm collection, it is possible to reveal genetic relationships among materials and to identify duplications. In order to save resources, a core collection with a reduced number of samples but essentially conserved genetic diversity can be extracted from such a dataset and form the basis of a breeding programme. Schenkel *et al.* (2002) had applied the concept of a “hierarchical nested core collection”, containing between eight and 102 accessions at five cluster levels, to half of the IITA bambara groundnut germplasm collection. However, its implementation was restricted by the exclusive use of morphological descriptors, which are influenced by environmental factors. This problem can be circumvented by drawing the core collection upon molecular marker data. Eventually, by

having a reliable estimate of the genetic diversity within a species, the number of test crosses in a breeding programme can be reduced through efficient selection of suitable parental lines. However, compared with other crops to which DArT has been applied, genetic diversity in cultivated bambara groundnut appears to be generally low and especially in Southern Africa, where landraces are adapted to low water availability. This may pose a serious constraint to any genetic improvement programme. Thus, it seems evident to increasingly make use of germplasm from the centre of origin and/or wild relatives. Such a strategy has been successfully applied to species like wheat (Huang *et al.*, 2003), barley (Von Korff *et al.*, 2006), tomato (Gur & Zamir, 2004) and common bean (Blair *et al.*, 2006) by means of advanced backcross QTL analysis.

DArT analysis furthermore revealed comparatively high genetic variation within landraces. From an academic point of view, it is clear that advanced ‘omics’ technologies do not make much sense without genetically defined material being available. For an inbreeding species like bambara groundnut, it should be relatively easy to develop pure lines from landraces, which are believed to consist of a series of inbred lines as confirmed using codominant molecular markers (Basu *et al.*, 2007b; Mayes *et al.*, 2009). Seeds from individual plants used for DArT marker discovery were harvested and may form the basis for future investigation. Nevertheless, it has been proposed that intra-landrace variability is one factor for adaptiveness of a population to a broad range of diverse environmental stresses (Zeven, 1998). Therefore, instead of homogeneous cultivars, bambara groundnut may be developed as a mixture of improved genotypes to meet the demands of low-input farming systems (Massawe *et al.*, 2005).

4.3.2 Gene expression profiling

High-throughput sequencing of pooled cDNA populations and subsequent counting of transcript tags created expression profiles for thousands of genes under reduced irrigation. Functional annotation of genes showed that virtually every metabolic aspect is affected by water-deficit stress and that there are manifold interactions between them. Hence, it is not surprising that a simple answer to a trait as complex as drought tolerance cannot be given. Both investigation of severe stress for four bambara groundnut genotypes adapted to environments differing in moisture availability as well as a time course experiment representing milder soil dehydration indicated a highly similar response between drought-tolerant and sensitive materials. Most differences were observed on a quantitative level, with sensitive genotypes/landraces usually exhibiting a more pronounced stress reaction, i.e. stronger induction of stress- and defence-related genes and stronger repression of photosynthesis-related genes. Furthermore, the adaptive phase appeared to be temporally confined for the more drought-tolerant DipC line, whereas the sensitive landrace LunT constantly invested resources into protective mechanisms, at least throughout the period

investigated. This suggests that control mechanisms underlying these responses are to be searched upstream of the genes presented in this work. With respect to transcription factors, only a small fraction was captured considering that in *Arabidopsis*, 190 members of the MYB family (Riechmann *et al.*, 2000) and 75 bZIP proteins (Jakoby *et al.*, 2002) have been described. Other transcription factors playing a role in plant water-deficit response like MYC, DREB and WRKY proteins (Chen *et al.*, 2002; Bartels & Sunkar, 2005) could not be identified in the bambara groundnut dataset. Likewise, the genes of the signal transduction cascade leading to the activation of transcription involving abscisic acid and MAPKs were probably expressed at levels beyond the MPSS detection limit.

When gene expression profiles were compared between MPSS and microarrays, it was also noticed that the tag counting-based approach delivered results with high resolution for strongly expressed genes, but quickly lost in power the fewer tags were present per cluster. On the contrary, microarrays still provided robust results for less strongly expressed genes. Thus, even with the ongoing further development of high-throughput sequencing technologies, it is unlikely that MPSS will catch up with microarrays in terms of transcription profiling in the near future. However, as shown in this work, MPSS coupled with 454 sequencing provides an invaluable tool for creating oligonucleotide microarrays. Setting aside (reliable) expression profiling but focussing on mere transcript discovery, possibly by means of normalised cDNA populations, fast development of a (nearly) whole-transcriptomic microarray through the information gained in one single sequencing run is not unrealistic.

Despite these technical limitations, at least two promising candidate genes for drought tolerance in bambara groundnut were identified. Further validation in various genomic backgrounds as well as multi-locational field testing are necessary to confirm their contribution to this complex trait. In parallel, as both transcript tags did not show homology to published sequences, it should be attempted to obtain full-length cDNAs by means of RACE (rapid amplification of cDNA ends)-PCR-based approaches (Frohman *et al.*, 1988). In case the transcripts prove to be novel, a number of *in silico* approaches are available to predict protein function (Engelhardt *et al.*, 2005; Friedberg, 2006). In addition, genes should also be investigated *in vivo*. This can be either through transformation into model plants like *Arabidopsis* or tobacco or gene silencing by means of RNA interference (Small, 2007) or virus-induced gene silencing (Holzberg *et al.*, 2002).

The complexity of drought tolerance, however, requires a more holistic view of biological processes. Firstly, while leaves are directly affected by water-deficit through their photosynthetic action, roots as the primary sensors and transmitters of water scarcity may not be disregarded. Secondly, gene expression allows for inference, but does not necessarily reflect the actual metabolism as, for example, post-transcriptional modification or substrate availability/affinity may influence the impact of expressed genes or translated proteins, respectively. Therefore, transcriptomics should be complemented by further disciplines that focus on investigation of the proteome and metabolome in order to integrate transcriptomic

data into more consistent picture of the physiology of bambara groundnut under drought conditions.

4.3.3 Subsumption of the work and future direction

In summary, two recently developed high-throughput technologies, MPSS coupled with 454 sequencing and DArT, plus a small oligonucleotide microarray, were successfully applied to bambara groundnut. Compared to decades of research and the enormous financial contribution dedicated to major crops, the presented approaches proved to be most cost- and time-efficient and may be regarded as a useful example for other under-utilised crops. Especially regarding MPSS, further applications are conceivable. Within a dataset as obtained in this study, it is possible to mine for expressed microsatellite (EST-SSR) motifs (Cheung *et al.*, 2006). In our case, read lengths were too short to capture flanking regions suitable for primer design. Secondly, genic SNP markers can be derived from the transcript tags. Due to the higher polymorphism rate, sequences from the 3'-UTR, as presented here, are particularly suitable.

One elegant way combining genetic markers and gene expression profiling is expression QTL (eQTL) analysis (Jansen & Nap, 2001). Considering expression levels as quantitative traits, they can be mapped in a segregating population. As a result, it will be indicated what portion of the variation in gene expression is attributed to the gene itself (*cis*-acting factors) and to what extent other genomic locations (*trans*-acting factors) influence gene expression. Thus, it is possible to detect regulatory candidate genes which may have been missed by expression profiling because of low expression levels or influential expression before sampling. This will provide additional insight into the regulatory network of complex traits such as drought tolerance as has been reported for rice (Hazen *et al.*, 2005) and poplar (Street *et al.*, 2006).

While the work presented focussed on the *de novo* generation of information specific for bambara groundnut, it is also possible to borrow resources developed for other species, e.g. major crops or model species such as *Arabidopsis thaliana* or, probably more suitable for bambara groundnut, *Medicago truncatula*. For example, Das *et al.* (2008) reported the successful application of single feature polymorphisms (SFPs), i.e. microarray markers based on expressed sequences, to cowpea using a readily available soybean microarray. Likewise, microarray gene expression studies can be conducted across species (Hammond *et al.*, 2006), or markers can be developed based on conserved gene synteny between plant genomes (Fulton *et al.*, 2002; Choi *et al.*, 2004). The latter two approaches are currently under investigation in bambara groundnut and may soon complement the data presented in this thesis (Mayes *et al.*, 2009).

Consistent improvement, accompanied by significant cost reduction, of biomolecular technologies allows for a deeper understanding of plant stress resistance mechanisms and, moreover, makes these accessible to a broader research community. However, practical

application requires a more holistic approach integrating ‘omics’ datasets derived from the cellular or organ level into experimental results on the whole-plant and canopy level (Wollenweber *et al.*, 2005; Witcombe *et al.*, 2008). Extensive and robust phenotyping is the key condition for such a ‘systems biology’ approach, but, due to high cost and labour-intensity, probably the most limiting factor for the progress of genomic studies on drought tolerance (Cattivelli *et al.*, 2008). The BAMLINK programme was designed to cover a broad range of agronomic aspects of bambara groundnut including field evaluation, assessment of genotype-by-environment interaction of key traits for yield stability under limiting water supply, development of cross-links to other abiotic stresses and functional/nutritional evaluation. In close cooperation with potential end-users, the near future will tell about the impact of such a concerted, integrative methodology on the contribution of an under-utilised crop to food security in semi-arid regions of the world.

5. Summary

Bambara groundnut [*Vigna subterranea* (L.) Verdc.] is an indigenous African legume which bears the potential to produce protein-rich seeds for human consumption even in resource-poor, drought-prone environments. However, it has fallen into disuse for a number of cultural and economic reasons and been replaced by major crops like maize or peanut. Consequently, no supra-regional markets exist at present and modern breeding research has ignored bambara groundnut during the last decades. In this context, the EU-funded BAMLINK project was launched in 2006 in order to promote the use of bambara groundnut for food security in (semi-)arid regions of the world. The work presented in this thesis was conducted with the overall aim to create a molecular genetic basis specific for bambara groundnut using state-of-the-art high-throughput technologies. In particular, technical objectives included a) the development of a Diversity Arrays Technology (DArT) molecular marker array for whole-genome profiling and b) its implementation for genotyping a representative fraction of germplasm held at genebanks and landrace material used by project partners. Secondly, it was intended to analyse gene expression patterns in leaves under water-deficit stress by c) expression profiling using the massively parallel signature sequencing (MPSS) approach coupled with a ‘next-generation sequencing’ technology (454 pyrosequencing), d) validation of these data using a custom-made cDNA microarray developed from a subset of the sequences generated through MPSS and e) investigating the time-dependent behaviour of these genes in a second water-deficit experiment. Thereby, it was aimed to integrate the data into a regulatory network and to identify candidate genes potentially explaining different degrees of drought tolerance in contrasting bambara groundnut landraces.

For genetic diversity studies, an initial DArT marker discovery array was developed from 38 diverse bambara groundnut DNA samples using the restriction enzyme combination *PstI/AluI* and analysed with 32 of these genotypes. Results indicated 5.5% polymorphic markers with a high proportion of redundancy. By increasing the genetic basis to 94 samples, adding an additional restriction digest to eliminate redundant markers and including fragments prepared from a second enzyme combination (*PstI/TaqI*), a full-size DArT array containing 7,680 clones was created. Of these, 658 (8.6%) were polymorphic in total and 296 (3.9%) showed unique segregation signatures. Compared to other crops, this indicates a low level of overall genetic diversity in bambara groundnut, which, however, is in the range of other legumes with a similar breeding system. These findings were confirmed by fingerprinting another 342 genotypes.

Using non-redundant markers to cluster accessions revealed a clear structure in bambara groundnut germplasm correlating well with the geographic origins of materials. Three relatively diverse clusters emerged from Western and Central African genotypes, including a cluster with probably not fully domesticated materials. In contrast, genotypes from Eastern and Southern Africa formed a single cluster which is characterised by a fairly narrow genetic

basis. Thus, it was concluded that a genetic bottleneck has occurred in the process of moving bambara groundnut out of its area of domestication. This became even more evident by investigating materials from Indonesia, where genetic diversity is almost absent. In addition, six landraces were analysed for intra-landrace diversity. Results indicated a broad margin from nearly pure lines to highly heterogeneous mixtures of genotypes.

To study the effects of water-deficit on gene expression, a controlled-environment experiment was conducted using four differently drought-adapted bambara groundnut landrace individuals. After the onset of flowering, reference leaf samples were harvested. Afterwards, relatively severe stress was applied by reducing irrigation to one third of control conditions for one week and the same plants were sampled again. 3'-untranslated regions from eight cDNA populations were sequenced in parallel using the 454 sequencing technology. After sequence cleaning, this produced a total of 197,400 sequence tags with an average read length of 87 nucleotides. Clustering these yielded 10,583 transcript groups represented by two to 3,026 sequence tags, and 34,427 singletons. Clusters were broken down into genotype and treatment-specific libraries and the number of tags was counted as a measure of gene expression. Results showed that two thirds of the most highly expressed genes were affected by water-deficit, with few genotypic differences. Downregulation was mainly observed for genes functioning in energy generating processes and protein synthesis, while stress and defence-related genes and genes for the modification and degradation of proteins usually exhibited induced expression. These results were validated by means of an oligonucleotide microarray designed from a subset of 132 MPSS-derived sequences. Comparison of both technologies indicated good correlation for highly expressed genes, but also decreasing accuracy of MPSS when transcript abundance became low. Microarrays were also employed to analyse a second water-deficit experiment, representing milder stress and using the two most contrasting landraces and six sampling dates. Most gene expression changes were observed in the first two days after initiating the treatment. Expression patterns were clustered and partitioned well into groups of similar biological functions, suggesting the presence of common regulatory mechanisms. Landrace-specific differences were usually of quantitative nature with a generally stronger response in the non-drought-adapted landrace, while the drought-tolerant landrace tended to react in a more concerted, temporary way. Nevertheless, when the data from both experiments were combined, two promising, potentially novel candidate genes, being induced upon water-deficit stress and displaying higher expression in the drought-adapted landrace, were identified.

In conclusion, two novel high-throughput technologies, DArT and MPSS, were successfully applied to bambara groundnut for the first time. Compared to the investment in major crops, significant data justifying further fundamental research on drought-tolerance, but also for direct breeding applications were created in a short time and a cost-efficient way. Thus, the approaches presented could serve as an example for molecular genetic research in other underutilised crops in order to increase food security in underprivileged regions of the world.

6. Zusammenfassung

Die Bambara-Erdnuss [*Vigna subterranea* (L.) Verdc.] ist eine in Afrika heimische Hülsenfrucht, die das Potential trägt, auch in ressourcenarmen, trockengefährdeten Gebieten proteinreiche Samen für die menschliche Ernährung hervorzubringen. Allerdings verlor sie aus einigen kulturellen wie wirtschaftlichen Gründen an Bedeutung und wurde durch Hauptkulturarten wie Mais oder die südamerikanische Erdnuss ersetzt. Folglich existieren derzeit keine überregionalen Märkte, und die moderne Züchtungsforschung ließ die Bambara-Erdnuss lange Zeit außer Acht. Vor diesem Hintergrund wurde 2006 das von der EU geförderte Projekt BAMLINK ins Leben gerufen, um die Bambara-Erdnuss als Beitrag zur Ernährungssicherheit in (semi-)ariden Regionen der Erde voranzutreiben. Grundlegendes Ziel der vorgestellten Arbeit war es, unter Einsatz aktueller Hochdurchsatz-Techniken eine für die Bambara-Erdnuss spezifische molekulargenetische Grundlage zu schaffen. Insbesondere umfasste die Studie a) die Entwicklung eines ‚Diversity Arrays Technology‘ (DArT) molekularen Markerarrays zur Genom-Profilierung und b) dessen Anwendung zur Genotypisierung eines repräsentativen Anteils von Genbank-Akzessionen und von Projektpartnern verwendeter Landsorten. Zweitens wurde angestrebt, Genexpressionsmuster in Blättern unter Wassermangel-Stress zu analysieren. Dies geschah durch c) Expressionsprofilierung mittels ‚massively parallel signature sequencing‘ (MPSS), gekoppelt mit einer ‚next-generation‘ Sequenzieretechnologie (454-Pyrosequenzierung), d) Validierung dieser Daten anhand eines individuell angefertigten cDNA-Mikroarrays und e) Untersuchung des zeitabhängigen Verhaltens dieser Gene in einem weiteren Wassermangel-Experiment. Es wurde beabsichtigt, die Daten in ein regulatorisches Netzwerk zu integrieren und Kandidatengene zu identifizieren, die möglicherweise verschiedene Grade an Trockenheitstoleranz in kontrastierenden Bambara-Erdnuss-Landsorten erklären.

Zur Erarbeitung von DArT-Markern für genetische Diversitätsstudien wurde zuerst ein Test-Array aus 38 unterschiedlichen DNA-Proben mit der Restriktionsenzymkombination *PstI/AluI* erstellt, der mit 32 dieser Genotypen analysiert wurde. Daraus ergaben sich 5,5 % polymorphe Marker, die jedoch einen hohen Anteil an Redundanz aufwiesen. Durch Ausweitung der genetischen Basis auf 94 Genotypen, Anwendung eines zusätzlichen Restriktionsverdau zur Entfernung redundanter Marker und Hinzunahme mit einer zweiten Enzymkombination (*PstI/TaqI*) generierter Fragmente wurde ein Genotypisierungsarray mit 7 680 Klonen zusammengestellt. Davon erwiesen sich insgesamt 658 (8,6 %) als polymorph, und 296 (3,9 %) zeigten unikale Spaltungsmuster. Im Vergleich zu anderen Kulturarten bedeutet dies einen geringen Grad an genetischer Diversität in der Bambara-Erdnuss, der sich jedoch im Bereich anderer Hülsenfrüchte mit ähnlicher Befruchtungsbiologie und Züchtungsgeschichte befindet. Diese Erkenntnisse wurden durch die Genotypisierung weiterer 342 Individuen bestätigt.

Die Verwendung nicht-redundanter Marker zur Clusteranalyse ließ eine klare Struktur im Bambara-Erdnuss-Genpool erkennen, die stark mit der Herkunft der Akzessionen korrelierte. Aus west- und zentralafrikanischen Genotypen bildeten sich drei vergleichsweise diverse Cluster heraus, wovon eines vermutlich nicht vollständig domestiziertes Material beinhaltet. Im Gegensatz dazu ergaben Genotypen aus Ost- und Südafrika ein einziges Cluster, das durch eine schmale genetische Basis gekennzeichnet ist. Daraus wurde gefolgert, dass die Verbrei-

tung der Bambara-Erdnuss außerhalb ihres Domestikationsgebietes zu einem genetischen Engpass führte. Dies wurde noch deutlicher durch die Analyse indonesischen Materials, das kaum mehr genetische Diversität in sich trägt. Zusätzlich wurden sechs Landsorten auf inhärente Diversität hin untersucht. Es zeigte sich eine weite Spanne von nahezu reinen Linien bis hin zu hochgradig heterogenen Saatgut-Mischungen.

Zur Erforschung der Auswirkungen von Wassermangel auf die Genexpression wurde ein Experiment unter kontrollierten Bedingungen mit vier unterschiedlich an Trockenheit angepassten Landsorten-Individuen durchgeführt. Nach Beginn der Blüte erfolgte die Entnahme von Referenz-Blattproben. Daran schloss sich eine einwöchige Phase relativ starken Stresses durch Reduktion der Bewässerung auf ein Drittel der Kontrollbedingungen und die abermalige Beprobung derselben Pflanzen. Mit der 454-Sequenziertechnologie wurden 3'-untranslatierte Bereiche aus acht cDNA-Populationen parallel ansequenziert. Dies ergab nach Bereinigung der Sequenzen 197 400 Tags mit einer durchschnittlichen Länge von 87 Nukleotiden. Diese clusterten in 10 583 Transkriptgruppen aus zwei bis 3 026 Tags und 34 427 Einzelsequenzen. Die Cluster wurden nach Genotyp und Behandlung unterteilt und die Anzahl der Sequenzen als Maß der Genexpression verwendet. Die Ergebnisse zeigten, dass zwei Drittel der am stärksten exprimierten Gene durch Wassermangel beeinflusst waren und wenige genotypische Unterschiede bestanden. Herunterregulierung war hauptsächlich bei Genen für energiegewinnende Prozesse und Proteinsynthese zu beobachten, während Gene der Stressabwehr und für den Um- oder Abbau von Proteinen in der Regel durch erhöhte Expression gekennzeichnet waren. Die Validierung dieser Ergebnisse erfolgte anhand eines Oligonukleotid-Mikroarrays, der aus 132 MPSS-Sequenzen erstellt wurde. Ein Vergleich der beiden Techniken ergab eine gute Korrelation für stark exprimierte Gene, MPSS zeigte jedoch nachlassende Präzision bei geringerer Transkriptmenge. Die Mikroarrays dienten auch zur Analyse eines weiteren Wassermangel-Experiments, welches milderen Stress darstellte und die beiden unterschiedlichsten Landsorten und sechs Beprobungszeitpunkte umfasste. Die meisten Genexpressionsänderungen waren in den ersten beiden Tagen nach Behandlungsbeginn zu beobachten. Clusteranalysen der Expressionsmuster erbrachten Gruppen mit ähnlicher biologischer Funktion, was auf das Vorhandensein gemeinsamer regulatorischer Mechanismen hindeutete. Landsorten-spezifische Unterschiede waren meist quantitativer Natur, wobei die nicht an Trockenheit adaptierte Landsorte für gewöhnlich eine stärkere Reaktion zeigte, während die trockenheitstolerante Landsorte zu einer konzertierteren, zeitlich begrenzten Stressantwort neigte. Dennoch wurden, wie die Kombination der Daten beider Experimente zeigte, zwei aussichtsreiche, möglicherweise noch nicht charakterisierte Kandidatengene identifiziert, die unter Wassermangel-Stress induziert und in der trockenadaptierten Landsorte stärker exprimiert waren.

Schlussfolgernd lässt sich der Einsatz zweier neuartiger Hochdurchsatz-Techniken, DArT und MPSS, bei der Bambara-Erdnuss als erfolgreich betrachten. Aussagekräftige Daten, die weiterführende Grundlagenforschung zur Trockenheitstoleranz rechtfertigen, aber auch für die direkte züchterische Anwendung, wurden, verglichen mit den Investitionen in Hauptkulturarten, auf zeit- und kostensparende Weise erzeugt. Somit können die vorgestellten Ansätze als Beispiel für molekulargenetische Forschung an anderen wenig genutzten Kulturen dienen, um die Ernährungssicherheit in den unterprivilegierten Regionen der Welt zu erhöhen.

7. Literature cited

- Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., Shinozaki, K. (1997): Role of Arabidopsis MYC and MYB Homologs in Drought- and Abscisic Acid-Regulated Gene Expression. *The Plant Cell* 9, 1859-1868.
- Acar, O., Türkan, I., Özdemir, F. (2001): Superoxide dismutase and peroxidase activities in drought sensitive and resistant barley (*Hordeum vulgare* L.) varieties. *Acta Physiologiae Plantarum* 23, 351-356.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., Venter, J.C. (1991): Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science* 252, 1651-1656.
- Adamska, I., Kloppstech, K. (1991): Evidence for an association of the early light-inducible protein (ELIP) of pea with photosystem II. *Plant Molecular Biology* 16, 209-223.
- Akashi, K., Nishimura, N., Ishida, Y., Yokota, A. (2004): Potent hydroxyl radical-scavenging activity of drought-induced type-2 metallothionein in wild watermelon. *Biochemical and Biophysical Research Communications* 323, 72-78.
- Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M.J., Howes, N., Sharp, P., Vaughan, P., Rathmell, B., Huttner, E., Kilian, A. (2006): Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and Applied Genetics* 113, 1409-1420.
- Alamerew, S., Chebotar, S., Huang, X., Röder, M., Börner, A. (2004): Genetic diversity in Ethiopian hexaploid and tetraploid wheat germplasm assessed by microsatellite markers. *Genetic Resources and Crop Evolution* 51, 559-567.
- Almoguera, C., Coca, M.A., Jordano, J. (1993): Tissue-specific expression of sunflower heat shock proteins in response to water stress. *The Plant Journal* 4, 232-239.
- Alobo, A.P. (1999): Production and organoleptic assessment of akara from bambara groundnut (*Voandzeia subterranea* (L.) Thouars). *Plant Foods for Human Nutrition* 53, 313-320.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Miller, W., Lipman, D.J. (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389-3402.
- Alwine, J.C., Kemp, D.J., Stark, G.R. (1977): Method for detection of specific RNAs on agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5350-5354.
- Amadi, E.N., Uneze, R., Barimalaa, I.S., Achinewhu, S.C. (1999): Studies on the production of bambara groundnut (*Vigna subterranea*) tempe. *Plant Foods and Human Nutrition* 53, 199-208.
- Amadou, H.I., Bebeli, P.J., Kaltsikes, P.J. (2001): Genetic diversity in Bambara groundnut (*Vigna subterranea* L.) germplasm revealed by RAPD markers. *Genome* 44, 995-999.
- Amarteifio, J.O., Moholo, D. (1998): The Chemical Composition of Four Legumes Consumed in Botswana. *Journal of Food Composition and Analysis* 11, 329-332.
- Anderson, J.A., Churchill, G.A., Autrique, J.E., Tanksley, S.D., Sorrells, M.E. (1993): Optimizing parental selection for genetic linkage maps. *Genome* 36, 181-186.
- Apel, K., Hirt, H. (2004): Reactive oxygen species: metabolism, oxidative stress and signal transduction. *Annual Review of Plant Biology* 55, 373-399.

- Armstrong, F., Leung, J., Grabov, A., Brearley, J., Giraudat, J., Blatt, M.R. (1995): Sensitivity to abscisic acid of guard cell K^+ channels is suppressed by *abi-1*, a mutant *Arabidopsis* gene encoding a putative protein phosphatase. Proceedings of the National Academy of Sciences of the United States of America 92, 9520-9524.
- Asada, K. (1999): The water-water cycle in chloroplasts: Scavenging of active oxygens and dissipation of excess photons. Annual Review of Plant Physiology and Plant Molecular Biology 50, 601-639.
- Asada, K., Takahashi, M. (1987): Production and scavenging of active oxygen in photosynthesis. In: Kyle, D.J., Osmond, C.B., Arntzen, C.J. (eds.): Photoinhibition. Elsevier, Amsterdam, pp. 227-287.
- Bahieldin, A., Mahfouz, H.T., Eissa, H.F., Saleh, O.M., Ramadan, A.M., Ahmed, I.A., Dyer, W.E., El-Itriby, H.A., Madkour, M.A. (2005): Field evaluation of transgenic wheat plants stably expressing the *HVA1* gene for drought tolerance. Physiologia Plantarum 123, 421-427.
- Baker, N.R., Rosenqvist, E. (2004): Applications of chlorophyll fluorescence can improve crop production strategies: an examination of future possibilities. Journal of Experimental Botany 55, 1607-1621.
- Bartels, D., Sunkar, R. (2005): Drought and Salt Tolerance in Plants. Critical Reviews in Plant Sciences 24, 23-58.
- Basu, S., Mayes, S., Davey, M., Roberts, J.A., Azam-Ali, S.N., Mithen, R., Pasquet, R.S. (2007a): Inheritance of 'domestication' traits in bambara groundnut (*Vigna subterranea* (L.) Verdc.). Euphytica 157, 59-68.
- Basu, S., Roberts, J.A., Azam-Ali, S.N., Mayes, S. (2007b): Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.) – an underutilized African legume crop species. Molecular Ecology Notes 7, 1326-1328.
- Basu, S., Roberts, J.A., Azam-Ali, S.N., Mayes, S. (2007c): Bambara Groundnut. In: Kole, C. (ed.): Genome Mapping and Molecular Breeding in Plants. Volume 3. Pulses, Sugar and Tuber Crops. Springer, Berlin, Heidelberg, pp 159-173.
- Begemann, F. (1988): Ecogeographic differentiation of bambara groundnut (*Vigna subterranea*) in the collection of the International Institute of Tropical Agriculture (I.I.T.A.). PhD thesis, Technische Universität München.
- Berchie, J.N., Azam-Ali, S.N., Collinson, S. (2002): Dry matter production and partitioning of bambara groundnut landraces in relation to soil moisture. In: Begemann, F., Mukema, I., Obel-Lawson, E. (eds.): Promotion of bambara groundnut (*Vigna subterranea*): Latest Developments of Bambara Groundnut Research. Proceedings of the 2nd International Workshop of the International Bambara Groundnut Network (BAMNET), 23-25 September 1998, CSIR, Accra, Ghana. International Plant Genetic Resources Institute (IPGRI), Nairobi, pp 84-91.
- Blair, M.W., Iriarte, G., Beebe, S. (2006): QTL analysis of yield traits in an advanced backcross population from cultivated Andean x wild common bean (*Phaseolus vulgaris* L.) cross. Theoretical and Applied Genetics 112, 1149-1163.
- Blum, A. (1996): Crop responses to drought and the interpretation of adaptation. Plant Growth Regulation 20, 135-148.
- Boerjan, W., Ralph, J., Baucher, M. (2003): Lignin Biosynthesis. Annual Review of Plant Biology 54, 519-546.
- Boguski, M.S., Lowe, T.M.J., Tolstoshev, C.M. (1993): dbEST – database for “expressed sequence tags”. Nature Genetics 4, 332-333.
- Bohnert, H.J., Sheveleva, E. (1998): Plant stress adaptations – making metabolism move. Current Opinion in Plant Biology 1, 267-274.

- Boominathan, P., Shukla, R., Kumar, A., Manna, D., Negi, D., Verma, P.K., Chattopadhyay, D. (2004): Long Term Transcript Accumulation during the Development of Dehydration Adaptation in *Cicer arietinum*. *Plant Physiology* 135, 1-13.
- Borsani, O., Diaz, P., Agius, M.F., Valpuesta, V., Monza, J. (2001): Water stress generates an oxidative stress through the induction of a Cu/Zn superoxide dismutase in *Lotus corniculatus* leaves. *Plant Science* 161, 757-763.
- Bowler, C., Van Montagu, M., Inzé, D. (1992): Superoxide dismutase and stress tolerance. *Annual Review of Plant Physiology and Plant Molecular Biology* 43, 83-116.
- Braun, H.-P., Schmitz, U.K. (1995): Molecular Features and Mitochondrial Import Pathway of the 14-Kilodalton Subunit of Cytochrome *c* Reductase from Potato. *Plant Physiology* 107, 1217-1223.
- Bräutigam, A., Shrestha, R.P., Whitten, D., Wilkerson, C.G., Carr, K.M., Froehlich, J.E., Weber, A.P.M. (2008): Low-coverage massively parallel pyrosequencing of cDNAs enables proteomics in non-model species: Comparison of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. *Journal of Biotechnology* 136, 44-53.
- Bray, E.A. (1993): Molecular responses to water deficit. *Plant Physiology* 103, 1035-1040.
- Bray, E.A. (1997): Plant responses to water deficit. *Trends in Plant Science* 2, 48-54.
- Bray, E.A., Bailey-Serres, J., Weretilnyk, E. (2000): Responses to abiotic stresses. In: Buchanan, B., Gruissem, W., Jones, R. (eds.): *Biochemistry and molecular biology of plants*. American Society of Plant Physiologists, Rockville, pp 1158-1249.
- Bray, E.A. (2002): Abscisic acid regulation of gene expression during water-deficit stress in the era of the Arabidopsis genome. *Plant, Cell and Environment* 25, 153-161.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridg, R.B., Kirchner, J., Fearon, K., Mao, J., Corcoran, K. (2000): Gene expression analysis by massively parallel signature sequencing (MPSS) on microarray beads. *Nature Biotechnology* 18, 630-634.
- Brink, M. (1999): Development, growth and dry matter partitioning in bambara groundnut (*Vigna subterranea*) as influenced by photoperiod and shading. *Journal of Agricultural Science* 133, 159-166.
- Brough, S.H., Azam-Ali, S.N. (1992): The Effect of Soil Moisture Content on the Proximate Composition of Bambara Groundnut (*Vigna subterranea* (L) Verdc). *Journal of the Science of Food and Agriculture* 60, 197-203.
- Brough, S.H., Azam-Ali, S.N., Taylor, A.J. (1993): The potential of bambara groundnut (*Vigna subterranea*) in vegetable milk production and basic protein functionality systems. *Food Chemistry* 47, 277-283.
- Camoni, L., Harper, J.F., Palmgren, M.G. (1998): 14-3-3 proteins activate a plant calcium-dependent protein kinase (CDPK). *FEBS Letters* 430, 381-384.
- Castresana, J., Saraste, M. (1995): Evolution of energetic metabolism: the respiration-early hypothesis. *Trends in Biochemical Sciences* 20, 443-448.
- Cattivelli, L., Rizza, F., Badeck, F.-W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., Marè, C., Tondelli, A., Stanca, A.M. (2008): Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Research* 105, 1-14.
- Ceccarelli, S., Grando, S. (1996): Drought as a challenge for the plant breeder. *Plant Growth Regulation* 20, 149-155.
- Cellier, F., Conéjéro, G., Breitler, J.-C., Casse, F. (1998): Molecular and Physiological Responses to Water Deficit in Drought-Tolerant and Drought-Sensitive Lines of Sunflower. *Plant Physiology* 116, 319-328.

- Chalker-Scott, L. (1999): Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology* 70, 1-9.
- Chang, S., Puryear, J., Cairney, J. (1993): A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11, 113-116.
- Charlesworth, D. (2003): Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 358, 1051-1070.
- Chaves, M.M. (1991): Effects of water deficits on carbon assimilation. *Journal of Experimental Botany* 42, 1-16.
- Chaves, M.M., Maroco, J.P., Pereira, J.S. (2003): Understanding plant responses to drought – from genes to the whole plant. *Functional Plant Biology* 30, 239-264.
- Chen, J., Agrawal, V., Rattray, M., West, M.A.L., St Clair, D.A., Michelmore, R.W., Coughlan, S.J., Meyers, B.C. (2007): A comparison of microarray and MPSS technology platforms for expression analysis of Arabidopsis. *BMC Genomics* 8, 414.
- Chen, W., Provar, N.J., Glazebrook, J., Katagiri, F., Chang, H.-S., Euglem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., Budworth, P.R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J.A., Harper, J.F., Si-Ammour, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J.L., Wang, X., Zhu, T. (2002): Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses. *The Plant Cell* 14, 559-574.
- Cheng, Z., Targolli, J., Huang, X., Wu, R. (2002): Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (*Oryza sativa* L.). *Molecular Breeding* 10, 71-82.
- Cheung, F., Haas, B.J., Goldberg, S.M.D., May, G.D., Xiao, Y., Town, C.D. (2006): Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics* 7, 272.
- Chia, T., Thorneycroft, D., Chapple, A., Messerli, G., Chen, J., Zeeman, S.C., Smith, S.M., Smith, A.M. (2004): A cytosolic glucosyltransferase is required for conversion of starch to sucrose in Arabidopsis leaves at night. *Plant Journal* 37, 853-863.
- Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M., Rafalski, A.J. (2002): SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3, 19.
- Choat, B., Ball, M., Lully, J., Holtem, J. (2003): Pit membrane porosity and water stress-induced cavitation in four co-existing dry rainforest tree species. *Plant physiology* 131, 41-48.
- Choi, H.-K., Mun, J.-H., Kim, D.-J., Zhu, H., Baek, J.-M., Mudge, J., Roe, B., Ellis, N., Doyle, J., Kiss, G.B., Young, N.D., Cook, D.R. (2004): Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America* 101, 15289-15294.
- Christmann, A., Hoffmann, T., Teplova, I., Grill, E., Müller, A. (2005): Generation of Active Pools of Abscisic Acid Revealed by In Vivo Imaging of Water-Stressed Arabidopsis. *Plant Physiology* 137, 209-219.
- Close, D.C., McArthur, C. (2002): Rethinking the role of many plant phenolics – protection from photodamage not herbivores? *Oikos* 99, 166-172.
- Close, T.J. (1997): Dehydrins: A commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* 100, 291-296.
- Close, T.J., Wanamaker, S.I., Caldo, R.A., Turner, S.M., Ashlock, D.A., Dickerson, J.A., Wing, R.A., Muehlbauer, G.J., Kleinhofs, A., Wise, R.P. (2004): A New Resource for Cereal Genomics: 22K Barley GeneChip Comes of Age. *Plant Physiology* 134, 960-968.

- Collinson, S.T., Azam-Ali, S.N., Chavula, K.M., Hodson, D. (1996): Growth, development and yield of bambara groundnut (*Vigna subterranea*) in response to soil moisture. *Journal of Agricultural Science* 126, 307-318.
- Collinson, S.T., Clawson, E.J., Azam-Ali, S.N., Black, C.R. (1997): Effects of soil moisture deficits on the water relations of bambara groundnut (*Vigna subterranea* L. Verdc.). *Journal of Experimental Botany* 48, 877-884.
- Collinson, S.T., Berchie, J., Azam-Ali, S.N. (1999): The effect of soil moisture on light interception and the conversion coefficient for three landraces of bambara groundnut (*Vigna subterranea*). *Journal of Agricultural Science* 133, 151-157.
- Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., Leonhardt, N., Dellaporta, S.L., Tonelli, C. (2005): A Guard-Cell-Specific MYB Transcription Factor Regulates Stomatal Movements and Plant Drought Tolerance. *Current Biology* 15, 1196-1200.
- Coudert, J. (1984): Market openings in West Africa for cowpeas and bambara groundnuts. *International Trade Forum* 20, 14-15, 28-29.
- Crowe, J.H., Hoekstra, F.A., Crowe, L.M. (1992): Anhydrobiosis. *Annual Review of Physiology* 54, 579-599.
- Cruz de Carvalho, M.H., d'Arcy-Lameta, A., Roy-Macauley, H., Gareil, M., El Maarouf, H., Pham-Thi, A.-T., Zuily-Fodil, Y. (2001): Aspartic protease in leaves of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* L. Walp): enzymatic activity, gene expression and relation to drought susceptibility. *FEBS Letters* 492, 242-246.
- Das, S., Bhat, P.R., Sudhakar, C., Ehlers, J.D., Wanamaker, S., Roberts, P.A., Cui, X., Close, T.J. (2008): Detection and validation of single feature polymorphisms in cowpea (*Vigna unguiculata* L. Walp) using a soybean genome array. *BMC Genomics* 9, 107.
- De Bruxelles, G.L., Peacock, W.J., Dennis, E.S., Dolferus, R. (1996): Abscisic Acid Induces the Alcohol Dehydrogenase Gene in *Arabidopsis*. *Plant Physiology* 111, 381-391.
- Deák, M., Horváth, G.V., Davletova, S., Török, K., Sass, L., Vass, I., Barna, B., Király, Z., Dudits, D. (1999): Plants ectopically expressing the iron-binding protein, ferritin, are tolerant to oxidative damage and pathogens. *Nature Biotechnology* 17, 192-196.
- Degenkolbe, T., Do, P.T., Zuther, E., Repsilber, D., Walther, D., Hinch, D.K., Köhl, K.I. (2009): Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. *Plant Molecular Biology* 69, 133-153.
- Demmig-Adams, B., Adams, W.W. (2002): Antioxidants in photosynthesis and human nutrition. *Science* 298, 2149-2153.
- Dezar, C.A., Gago, G.M., González, D.H., Chan, R.L. (2005): *Hahb-4*, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants. *Transgenic Research* 14, 429-440.
- Doku, E.V. (1968): Flowering, pollination and pod formation in bambara groundnut (*Voandzeia subterranea*) in Ghana. *Experimental Agriculture* 4, 41-48.
- Doku, E.V. (1969): Growth habit and pod production in bambara groundnut (*Voandzeia subterranea*). *Ghana Journal of Agricultural Science* 2, 91-95.
- Doku, E.V., Karikari, S.K. (1971): Bambara groundnut. *Economic Botany* 25, 225-262.
- Duke, J.A. (1981): *Voandzeia subterranea* (L.) Thouars. In: *Handbook of legumes of world economic importance*. Plenum Press, New York, London, pp 307-310.
- Dure, L. III, Greenway, S.C., Galau, C.A. (1981): Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* 20, 4162-4168.
- Dure, L. III, Crouch, M., Harada, J., Ho, T.-H. D., Quatrano, R., Mundy, J., Thomas, T., Sung, Z.R. (1989): Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Molecular Biology* 12, 475-486.

- Dure, L. III (1993): A repeating 11-mer amino acid motif and plant desiccation. *The Plant Journal* 3, 363-369.
- Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998): Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95, 14863-14868.
- Emrich, S.J., Barbazuk, W.B., Li, L., Schnable, P.S. (2007): Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research* 17, 69-73.
- Engelhardt, B.E., Jordan, M.I., Muratore, K.E., Brenner, S.E. (2005): Protein Molecular Function Prediction by Bayesian Phylogenomics. *PLoS Computational Biology* 1, e45.
- Enwere, N.J., Hung, Y.-C. (1996): Some chemical and physical properties of bambara groundnut (*Voandzeia subterranea Thouars*) seed and products. *International Journal of Food Sciences and Nutrition* 47, 469-475.
- Espartero, J., Sánchez-Aguayo, I., Pardo, J.M. (1995): Molecular characterization of glyoxalase-I from a higher plant; upregulation by stress. *Plant Molecular Biology* 29, 1223-1233.
- Eveland, A.L., McCarthy, D.R., Koch, K.E. (2008): Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiology* 146, 32-44.
- Evenson, R.E., Gollin, D. (2003): Assessing the impact of the green revolution, 1960 to 2000. *Science* 300, 758-762.
- Fanger, G.R., Widmann, C., Porter, A.C., Sather, S., Johnson, G.L., Vaillancourt, R.R. (1998): 14-3-3 Proteins Interact with Specific MEK Kinases. *The Journal of Biological Chemistry* 273, 3476-3483.
- FAO (1997): *The State of the World's Plant Genetic Resources for Food and Agriculture*. Food and Agriculture Organization of the United Nations, Rome.
- FAO (2009): FAOSTAT. Available online: <http://faostat.fao.org/site/567/default.aspx#ancor>. Accessed 13 Jan 2009.
- FAO/WHO (1973): *Energy and Protein Requirements*. Report of a joint FAO/WHO *Ad Hoc* Committee. World Health Organization, Technical Report Series no. 522, Geneva.
- Fernie, A.R., Carrari, F., Sweetlove, L.J. (2004): Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology* 7, 254-261.
- Fleißner, K.W.E. (2006): *The improvement of bambara groundnut production in Northern Namibia by means of breeding strategies and agronomic investigations*. PhD thesis, Technische Universität München.
- Forster, B.P., Ellis, R.P., Moir, J., Talamè, V., Sanguineti, M.C., Tuberosa, R., This, D., Teulat-Merah, B., Ahmed, I., Mariy, S.A.E.E., Bahri, H., El Ouahabi, M., Zoumarou-Wallis, N., El-Fellah, M., Ben Salem, M. (2004): Genotype and phenotype associations with drought tolerance in barley tested in North Africa. *Annals of Applied Biology* 144, 157-168.
- Frahm-Leliveld, J.A. (1953): Some chromosome numbers in tropical leguminous plants. *Euphytica* 2, 46-48.
- Francia, E., Tacconi, G., Crossati, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., Valè, G. (2005): Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture* 82, 317-342.
- Friedberg, I. (2006): Automated protein function prediction – the genomic challenge. *Briefings in Bioinformatics* 7, 225-242.
- Frohman, M.A., Dush, M.K., Martin, G.R. (1988): Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America* 85, 8998-9002.

- Fulton, T.M., Van der Hoeven, R., Eannetta, N.T., Tanksley, S.D. (2002): Identification, Analysis, and Utilization of Conserved Ortholog Set Markers for Comparative Genomics in Higher Plants. *The Plant Cell* 14, 1457-1467.
- Gal, T.Z., Glazer, I., Koltai, H. (2004): An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. *FEBS Letters* 577, 21-26.
- Gallardo, K., Firnhaber, C., Zuber, H., Héricher, D., Belghazi, M., Henry, C., Küster, H., Thompson, R. (2007): A Combined Proteome and Transcriptome Analysis of Developing *Medicago truncatula* Seeds. *Molecular and Cellular Proteomics* 6, 2165-2179.
- Giardi, M.T., Cona, A., Geiken, B., Kučera, T., Masojídek, J., Mattoo, A.K. (1996): Long-term drought stress induces structural and functional reorganization of photosystem II. *Planta* 199, 118-125.
- Glew, R.H., VanderJagt, D.J., Lockett, C., Grivetti, L.E., Smith, G. C., Pastuszyn, A., Millson, M. (1997): Amino Acid, Fatty Acid, and Mineral Composition of 24 Indigenous Plants of Burkina Faso. *Journal of Food Composition and Analysis* 10, 205-217.
- Gomes, C.I., Obendorf, R.L., Horbowicz, M. (2005): *myo*-Inositol, *D-chiro*-Inositol, and *D*-Pinitol Synthesis, Transport, and Galactoside Formation in Soybean Explants. *Crop Science* 45, 1312-1319.
- Gruenbaum, Y., Naveh-Many, T., Cedar, H., Razin, A. (1981): Sequence specificity of methylation in higher plant DNA. *Nature* 292, 860-862.
- Gur, A., Zamir, D. (2004): Unused Natural Variation Can Lift Yield Barriers in Plant Breeding. *PLoS Biology* 2, e245.
- Hamdy, A., Ragab, R., Scarascia-Mugnozza, E. (2003): Coping with water scarcity: water saving and increasing water productivity. *Irrigation and Drainage* 52, 3-20.
- Hammer, K., Heller, J., Engels, J. (2001): Monographs on underutilized and neglected crops. *Genetic Resources and Crops Evolution* 48, 3-5.
- Hammond, J.P., Bowen, H.C., White, P.J., Mills, V., Pyke, K.A., Baker, A.J.M., Whiting, S.N., May, S.T., Broadley, M.R. (2006): A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist* 170, 239-260.
- Hanks, J.N., Snyder, A.K., Graham, M.A., Shah, R.K., Blaylock, L.A., Harrison, M.J., Shah, D.M. (2005): Defensin gene family in *Medicago truncatula*: structure, expression and induction by signal molecules. *Plant Molecular Biology* 58, 385-399.
- Harris, D., Azam-Ali, S.N. (1993): Implications of daylength sensitivity in bambara groundnut (*Vigna subterranea*) for production in Botswana. *Journal of Agricultural Science* 120, 75-78.
- Havaux, M. (1998): Carotenoids as membrane stabilizers in chloroplasts. *Trends in Plant Science* 3, 174-151.
- Hazen, S.P., Pathan, M.S., Sanchez, A., Baxter, I., Dunn, M., Estes, B., Chang, H.-S., Zhu, T., Kreps, J.A., Nguyen, H.T. (2005): Expression profiling of rice segregating for drought tolerance QTLs using a rice genome array. *Functional & Integrative Genomics* 5, 104-116.
- Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M. (1996): Real Time Quantitative PCR. *Genome Research* 6, 986-994.
- Hepper, F.N. (1963): The bambara groundnut (*Voandzeia subterranea*) and Kersting's groundnut (*Kerstingiella geocarpa*) wild in West Africa. *Kew Bulletin* 16, 395-407.
- Hieng, B., Ugrinović, K., Šuštar-Vozlič, J., Kidrič, M. (2004): Different classes of proteases are involved in the response to drought of *Phaseolus vulgaris* L. cultivars differing in sensitivity. *Journal of Plant Physiology* 161, 519-530.
- Holzberg, S., Brosio, P., Gross, C., Pogue, G.P. (2002): *Barley stripe mosaic virus*-induced gene silencing in a monocot plant. *The Plant Journal* 30, 315-327.

- Howell, J.A., Eshbaugh, W.H., Guttman, S., Rabakonandrianina, E. (1994): Common names given to bambara groundnut (*Vigna subterranea*; Fabaceae) in Central Madagascar. *Economic Botany* 48, 217-221.
- Hsieh, H.-M., Liu, W.-K., Huang, P.C. (1995): A novel stress-inducible metallothionein-like gene from rice. *Plant Molecular Biology* 28, 381-389.
- Huang, X., Madan, A. (1999): CAP3: a DNA sequence assembly program. *Genome Research* 9, 868-877.
- Huang, X. Q., Cöster, H., Ganai, M-W., Röder, M.S. (2003): Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 106, 1379-1389.
- Hundertmark, M., Hinch, D.K. (2008): LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9, 118.
- Huttner, E., Wenzl, P., Akbari, M., Caig, V., Carling, J., Cayla, C., Evers, M., Jaccoud, D., Peng, K., Patarapuwadol, S., Uszynski, G., Xia, L., Yang, S., Kilian, A. (2005): Diversity Arrays Technology: A Novel Tool for Harnessing the Genetic Potential of Orphan Crops. In: Serageldin, I., Persely, G.J. (eds.): *Discovery to Delivery: BioVision Alexandria 2004*. Bibliotheca Alexandrina, El-Shatby-Alexandria, pp. 145-155.
- Hyten, D.L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., Cregan, P.B. (2006): Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 103, 16666-16671.
- Ingram, J., Bartels, D. (1996): The molecular basis of dehydration tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 47, 377-403.
- IPGRI, IITA, BAMNET (2000): Descriptors for bambara groundnut (*Vigna subterranea*). International Plant Genetic Resources Institute, Rome, Italy; International Institute of Tropical Agriculture, Ibadan, Nigeria; The International Bambara Groundnut Network, Germany.
- Iqbal, M.J., Reddy, O.U.K., El-Zik, K.M., Pepper, A.E. (2001): A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics* 103, 547-554.
- Iuchi, S., Kobayashi, M., Yamaguchi-Shinozaki, K., Shinozaki, K. (2000): A Stress-Inducible Gene for 9-cis-Epoxycarotenoid Dioxygenase Involved in Abscisic Acid Biosynthesis under Water Stress in Drought-Tolerant Cowpea. *Plant Physiology* 123, 553-562.
- Iuchi, S., Yamaguchi-Shinozaki, K., Urao, T., Terao, T., Shinozaki, K. (1996): Novel Drought-Inducible Genes in the Highly Drought-Tolerant Cowpea: Cloning of cDNAs and Analysis of the Expression of the Corresponding Genes. *Plant and Cell Physiology* 37, 1073-1082.
- Jaccard, P. (1908): Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles* 44, 223-270.
- Jaccoud, D., Peng, K., Feinstein, D., Kilian, A. (2001): Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29, e25.
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F. (2002): bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* 7, 106-111.
- Jansen, R.C., Nap, J.-P. (2001): Genetical genomics: the added value from segregation. *Trends in Genetics* 17, 388-391.
- Jansson, S. (1994): The light-harvesting chlorophyll *a/b*-binding proteins. *Biochimica et Biophysica Acta* 1184, 1-19.

- John, P. (1997): Ethylene biosynthesis: The role of 1-aminocyclopropane-1-carboxylate (ACC) oxidase, and its possible evolutionary origin. *Physiologia Plantarum* 100, 583-592.
- Johns, M.A., Skroch, P.W., Nienhuis, J., Hinrichsen, P., Bascur, G., Muñoz-Schick (1997): Gene Pool Classification of Common Bean Landraces from Chile Based on RAPD and Morphological Data. *Crop Science* 37, 605-613.
- Jonak, C., Ligerink, W., Hirt, H. (1999): MAP kinases in plant signal transduction. *Cellular and Molecular Life Sciences* 55, 204-213.
- Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J.G., Vasicek, T.J. (2007): An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Research* 15, 1007-1014.
- Kadam, S.S., Reddy, N.R., Patil, G.D. (1989): Other legumes. In: Salunkhe, D.K., Kadam, S.S. (eds.): *CRC handbook of world food legumes: nutritional chemistry, processing technology, and utilization*. Volume III. CRC Press, Boca Raton, pp 67-88.
- Kader, J.-C. (1997): Lipid-transfer proteins: a puzzling family of plant proteins. *Trends in Plant Science* 2, 66-70.
- Kaplan, F., Guy, C.L. (2004): β -Amylase Induction and the Protective Role of Maltose during Temperature Shock. *Plant Physiology* 135, 1674-1684.
- Kaplan, F., Sung, D.Y., Guy, C.L. (2006): Roles of β -amylase and starch breakdown during temperatures stress. *Physiologia Plantarum* 126, 120-128.
- Kapranov, P., Jensen, T.J., Poulsen, C., de Bruijn, F., Szczyglowski, K. (1999): A protein phosphatase 2C gene, *LjNPP2C1*, from *Lotus japonicus* induced during root nodule development. *Proceedings of the National Academy of Sciences of the United States of America* 96, 1738-1743.
- Karin, M., Hunter, T. (1995): Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Current Biology* 5, 747-757.
- Kavar, T., Maras, M., Kidrič, M., Šuštar-Vozlič, J., Meglič, V. (2008): Identification of genes involved in the response of leaves of *Phaseolus vulgaris* to drought stress. *Molecular Breeding* 21, 159-172.
- Keller, B. (1993): Structural Cell Wall Proteins. *Plant Physiology* 101, 1127-1130.
- Keller, F., Ludlow, M.M. (1993): Carbohydrate metabolism in drought-stressed leaves of pigeonpea (*Cajanus cajan*). *Journal of Experimental Botany* 44, 1351-1359.
- Kilian, A., Huttner, E., Wenzl, P., Jaccoud, D., Carling, J., Caig, V., Evers, M., Heller-Uszynska, K., Uszynski, G., Cayla, C., Patarapuwadol, S., Xia, L., Yang, S., Thomson, B. (2005): The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In Tuberosa, R., Phillips, R.L., Gale, M. (eds.): *Proceedings of the International Congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution"*, May 27-31 2003, Bologna, Italy. Avenue media, Bologna, pp 443-461.
- Kim, J.-Y., Mahé, A., Brangeon, J., Prioul, J.-L. (2000): A Maize Vacuolar Invertase, *IVR2*, Is Induced by Water Stress. *Organ/Tissue Specificity and Diurnal Modulation of Expression*. *Plant Physiology* 124, 71-84.
- Knight, H., Knight, M.R. (2001): Abiotic stress signalling pathways: specificity and cross-talk. *Trends in Plant Science* 6, 262-267.
- Kovtun, Y., Chiu, W.-L., Tena, G., Sheen, J. (2000): Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proceedings of the National Academy of Sciences of the United States of America* 97, 2940-2945.
- Koster, K.L. (1991): Glass formation and desiccation tolerance in seeds. *Plant Physiology* 96, 302-304.

- Lawrence, J.B., Singer, R.H. (1985): Quantitative analysis of *in situ* hybridization methods for the detection of actin gene expression. *Nucleic Acids Research* 13, 1777-1799.
- Lavin, M., Herendeen, P.S., Wojciechowski, M.F. (2005): Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary. *Systematic Biology* 54, 575-594.
- Lee, M.-L.T., Kuo, F.C., Whitmore, G.A., Sklar, J. (2000): Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of the National Academy of Sciences of the United States of America* 91, 9834-9839.
- Levitt, J. (1980): Responses of plants to environmental stresses. Volume II. Water, radiation, salt, and other stresses. Academic Press, New York.
- Liang, Y.-K., Dubos, C., Dodd, I.C., Holroyd, G.H., Hetherington, A.M., Campbell, M.M. (2005): *AtMYB61*, an R2R3-MYB Transcription Factor Controlling Stomatal Aperture in *Arabidopsis thaliana*. *Current Biology* 15, 1201-1206.
- Linnemann, A.R. (1991): Preliminary observations on photoperiod regulation of phenological development in bambara groundnut (*Vigna subterranea*). *Field Crops Research* 26, 295-304.
- Linnemann, A.R., Azam-Ali, S. (1993): Bambara groundnut (*Vigna subterranea*). In: Williams, J.T. (ed.): Pulses and vegetables. Chapman & Hall, London, pp 13-58.
- Linnemann, A.R., Westphal, E., Wessel, M. (1995): Photoperiod regulation of development and growth in bambara groundnut (*Vigna subterranea*). *Field Crops Research* 40, 39-47.
- Liu, F., Jenssen, T.-K., Trimarchi, J., Punzo, C., Cepko, C.L., Ohno-Machado, L., Hovig, E., Kuo, W.P. (2007): Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates. *BMC Genomics* 8, 153.
- Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., Shinozaki, K. (1998): Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *The Plant Cell* 10, 1391-1406.
- Liu, T., Liu, S., Guan, H., Ma, L., Chen, Z., Gu, H., Qu, L.-J. (2009): Transcriptional profiling of *Arabidopsis* seedlings in response to heavy metal lead (Pb). *Environmental and Experimental Botany*, *in press*.
- Loewus, F.A., Murthy P.P.N. (2000): *myo*-Inositol metabolism in plants. *Plant Science* 150, 1-19.
- Lopez, C.G., Banowitz, G.M., Peterson, C.J., Kronstad, W.E. (2003): Dehydrin Expression and Drought Tolerance in Seven Wheat Cultivars. *Crop Science* 43, 577-582.
- Lorence, A., Chevone, B.I., Mendes, P., Nessler, C.L. (2004): *myo*-Inositol Oxygenase Offers a Possible Entry Point into Plant Ascorbate Biosynthesis. *Plant Physiology* 134, 1200-1205.
- Ludlow, M.M., Björkman, O. (1984): Paraheliotropic leaf movement in Siratro as a protective mechanism against drought-induced damage to primary photosynthetic reactions: damage by excessive light and heat. *Planta* 161, 505-518.
- Ludlow, M.M., Santamaria, J.M., Fukai, S. (1990): Contributions of Osmotic Adjustment to Grain Yield in Sorghum bicolor (L.) Moench under Water-limited Conditions. II. Water stress after Anthesis. *Australian Journal of Agricultural Research* 41, 67-78.
- Mace, E.S., Xia, L., Jordan, D.R., Halloran, K., Parh, D.K., Huttner, E., Wenzl, P., Kilian, A. (2008): DArT markers: diversity analyses and mapping in *Sorghum bicolor*. *BMC Genomics* 9, 26.

- MacKay, J.J., O'Malley, D.M., Presnell, T., Booker, F.L., Campbell, M.M., Whetten, R.W., Sederoff, R.R. (1997): Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* 94, 8255-8260.
- Mai-Kodomi, Y., Singh, B.B., Myers Jr., O., Yopp, J.H., Gibson, P.J., Terao, T. (1999): Two mechanisms of drought tolerance in cowpea. *Indian Journal of Genetics* 59, 309-316.
- Mardis, E.R. (2008): The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24, 133-141.
- Maréchal, R. (1982): Arguments for a global conception of the genus *Vigna*. *Taxon* 31, 280-283.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Hegelsen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M. (2005): Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- Maroco, J., Rodrigues, M.L., Lopes, C., Chaves, M.M. (2002): Limitations to leaf photosynthesis in field-grown grapevine under drought – metabolic and modelling approaches. *Functional Plant Biology* 29, 451-459.
- Massawe, F.J., Dickinson, M., Roberts, J.A., Azam-Ali, S.N. (2002): Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces revealed by AFLP markers. *Genome* 45, 1175-1180.
- Massawe, F.J., Roberts, J.A., Azam-Ali, S.N., Davey, M.R. (2003): Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) Verdc) landraces assessed by Random Amplified Polymorphic DNA (RAPD) markers. *Genetic Resources and Crop Evolution* 50, 737-741.
- Massawe, F.J., Mwale, S.S., Azam-Ali, S.N., Roberts, J.A. (2005): Breeding in bambara groundnut (*Vigna subterranea* (L.) Verdc.): strategic considerations. *African Journal of Biotechnology* 4, 463-471.
- Mayes, S., Basu, S., Murchie, E., Roberts, J.A., Azam-Ali, S.N., Stadler, F., Mohler, V., Wenzel, G., Massawe, F., Kilian, A., Bonin, A., Beena, A., Sheshshayee, M.S. (2009): BAMLINK – a Cross Disciplinary Programme to Enhance the Role of Bambara Groundnut (*Vigna subterranea* L. Verdc.) for Food Security in Africa and India. *Acta Horticulturae* 806, 137-150.
- McKersie, B.D., Leshem, Y.Y. (1994): Water and drought stress. In: McKersie, B.D., Leshem, Y.Y. (eds.): *Stress and stress coping in cultivated plants*. Kluwer Academic Publishers, Dordrecht, Boston, London, pp 148-180.
- Meijer, H.J.G., Munnik, T. (2003): Phospholipid-based signalling in plants. *Annual Review of Plant Biology* 54, 265-306.
- Mellway, R.D., Tran, L.T., Prouse, M.B., Campbell, M.M., Constabel, C.P. (2009): The Wound-, Pathogen-, and Ultraviolet B-Responsive *MYB134* Gene Encodes an R2R3 MYB Transcription Factor That Regulates Proanthocyanidin Synthesis in Poplar. *Plant Physiology* 150, 924-941.
- Meyer, K., Leube, M.P., Grill, E. (1994): A Protein Phosphatase 2C Involved in ABA Signal Transduction in *Arabidopsis thaliana*. *Science* 264, 1452-1455.

- Mooney, B.P., Thelen, J.J. (2004): High-throughput mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry* 65, 1733-1744.
- Moran, J.F., Becana, M., Iturbe-Ormaetxe, I., Frechilla, S., Klucas, R.V., Aparicio-Tejo, P. (1994): Drought induces oxidative stress in pea plants. *Planta* 194, 346-352.
- Morgan, J.M., Condon, A.G. (1986): Water Use, Grain Yield, and Osmoregulation in Wheat. *Australian Journal of Plant Physiology* 13, 523-532.
- Morgan, P.W., Drew, M.C. (1997): Ethylene and plant responses to stress. *Physiologia Plantarum* 100, 620-630.
- Mukhopadhyay, A., Vij, S., Tyagi, A.K. (2004): Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. *Proceedings of the National Academy of Sciences of the United States of America* 101, 6309-6314.
- Müller, R., Morant, M., Jarmer, H., Nilsson, L., Nielsen, T.H. (2007): Genome-Wide Analysis of the Arabidopsis Leaf Transcriptome Reveals Interaction of Phosphate and Sugar Metabolism. *Plant Physiology* 143, 156-171.
- Munné-Bosch, S., Alegre, L. (2003): Drought-induced changes in the redox state of α -tocopherol, ascorbate, and the diterpene carnosic acid in chloroplasts of Labiatae species differing in carnosic acid contents. *Plant Physiology* 131, 1816-1825.
- Murdock, G.P. (1959): *Africa: Its Peoples and Their Culture History*. McGraw-Hill, New York.
- Mwale, S.S., Azam-Ali, S.N., Massawe, F.J., Roberts, J.A. (2003): Effect of soil moisture on growth and development of bambara groundnut (*Vigna subterranea* (L.) Verdc.). In: *Proceedings of the International Bambara Groundnut Symposium*, Botswana College of Agriculture, Botswana, 8-12 August 2003, pp 65-84.
- Mwale, S.S., Azam-Ali, S.N., Massawe, F.J. (2007): Growth and development of bambara groundnut (*Vigna subterranea*) in response to soil moisture. 1. Dry matter and yield. *European Journal of Agronomy* 26, 345-353.
- Nanjo, T., Kobayashi, M., Yoshiba, Y., Sanada, Y., Wada, K., Tsukaya, H., Kakubari, Y., Yamaguchi-Shinozaki, K., Shinozaki, K. (1999): Biological functions of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. *The Plant Journal* 18, 185-193.
- Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2003): Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. *The Plant Journal* 34, 137-148.
- National Research Council (2006): *Lost crops of Africa. Volume II: Vegetables*. The National Academies Press, Washington, D.C., pp 52-73.
- Nei, M. (1973): Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70, 3321-3323.
- Nelson, O.E. (1969): Genetic Modification Of Protein Quality In Plants. *Advances in Agronomy* 21, 171-194.
- Noctor, G., Foyer, C.H. (1998): Ascorbate and glutathione: keeping active oxygen under control. *Annual Review of Plant Physiology and Plant Molecular Biology* 49, 249-279.
- Noctor, G., Veljovic-Jovanovic, S., Driscoll, S., Novitskaya, L., Foyer, C.H. (2002): Drought and oxidative load in the leaves of C3 plants: a predominant role of photorespiration? *Annals of Botany* 89, 841-850.

- Ntundu, W.H., Bach, I.C., Christiansen, J.L., Andersen, S.B. (2004): Analysis of genetic diversity in bambara groundnut [*Vigna subterranea* (L.) Verdc] landraces using amplified fragment length polymorphism (AFLP) markers. *African Journal of Biotechnology* 3, 220-225.
- Ntundu, W.H., Shillah, S.A., Marandu, W.Y.F., Christiansen, J.L. (2006): Morphological diversity of bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces in Tanzania. *Genetic Resources and Crop Evolution* 53, 367-378.
- Nwokolo, E. (1987): A Nutritional Assessment of African Yam Bean *Sphenostylis stenocarpa* (Hochst ex A. Rich) Harms. and Bambara Groundnut *Voandzeia subterranea* L. *Journal of the Science of Food and Agriculture* 41, 123-129.
- Nygaard, V., Liu, F., Holden, M., Kuo, W.P., Trimarchi, J., Ohno-Machado, L., Cepko, C.L., Frigessi, A., Glad, I.K., van de Wiel, M.A., Hovig, E., Lyng, H. (2008): Validation of oligoarrays for quantitative exploration of the transcriptome. *BMC Genomics* 9, 258.
- Oberschall, A., Deák, M., Török, K., Sass, L., Vass, I., Kovács, I., Fehér, A., Dudits, D., Horváth, G.V. (2000): A novel aldose/aldehyde reductase protects transgenic plants against lipid peroxidation under chemical and drought stress. *The Plant Journal* 24, 437-446.
- Obizoba, I.C., Egbuna, H.I. (1992): Effect of germination and fermentation on the nutritional quality of bambara nut (*Voandzeia subterranea* L. Thouars) and its product (milk). *Plant Foods for Human Nutrition* 42, 13-23.
- Odeigah, P.G.C., Osanyinpeju, A.O. (1998): Evaluating the genetic biodiversity of Bambara groundnut accessions from Nigeria using SDS-polyacrylamide gel electrophoresis. *Genetic Resources and Crop Evolution* 45, 451-458.
- Ogundiwin, E.A., Martí, C., Forment, J., Pons, C., Granell, A., Gradziel, T.M., Peace, C.P., Crisosto, C.H. (2008): Development of ChillPeach genomic tools and identification of cold-responsive genes in peach fruit. *Plant Molecular Biology* 68, 379-397.
- Oliver, D.J. (1994): The glycine decarboxylase complex from plant mitochondria. *Annual Review of Plant Physiology and Plant Molecular Biology* 45, 323-337.
- Onimawo, I.A., Momoh, A.H., Usman, A. (1998): Proximate composition and functional properties of four cultivars of bambara groundnut (*Voandzeia subterranea*). *Plant Foods for Human Nutrition* 53, 153-158.
- Öquist, G., Chow, W.S., Anderson, J.M. (1992): Photoinhibition of photosynthesis represents a mechanism for the long-term regulation of photosystem II. *Planta* 186, 450-460.
- Orozco-Castillo, C., Chalmers, K.J., Waugh, R., Powell, W. (1994): Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics* 87, 934-940.
- Ozturk, Z.N., Talamé, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R., Bohnert, H.J. (2002): Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Molecular Biology* 48, 551-573.
- Palmiter, R.D. (1998): The elusive function of metallothioneins. *Proceedings of the National Academy of Sciences of the United States of America* 95, 8428-8430.
- Parella, L., Gepts, P. (1992): Genetic relationships within *Vigna unguiculata* (L.) Walp. based on isozyme analyses. *Genetic Resources and Crop Evolution* 39, 71-88.
- Park, B.-J., Liu, Z., Kanno, A., Kameya, T. (2005): Genetic improvement of Chinese cabbage for salt and drought tolerance by constitutive expression of a *B. napus* LEA gene. *Plant Science* 169, 553-558.
- Pasquet, R.S., Schwedes, S., Gepts, P. (1999): Isozyme diversity in bambara groundnut. *Crop Science* 39, 1228-1236.
- Passioura, J.B. (1996): Drought and drought tolerance. *Plant Growth Regulation* 20, 79-83.

- Pastenes, C., Porter, V., Baginsky, C., Horton, P., González, J. (2004): Paraheliotropism can protect water-stressed bean (*Phaseolus vulgaris* L.) plants against photoinhibition. *Journal of Plant Physiology* 161, 1315-1323.
- Paul, M.J., Foyer, C.H. (2001): Sink regulation of photosynthesis. *Journal of Experimental Botany* 52, 1383-1400.
- Penninckx, I.A.M.A., Thomma, B.P.H.J., Buchala, A., Métraux, J.-P., Broekaert, W.F. (1998): Concomitant Activation of Jasmonate and Ethylene Response Pathways Is Required for Induction of a Plant Defensin Gene in Arabidopsis. *The Plant Cell* 10, 2103-2113.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J. (2003): TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651-652.
- Pinheiro, C., Chaves, M.M., Ricardo, C.P. (2001): Alterations in carbon and nitrogen metabolism induced by water-deficit in the stem and leaves of *Lupinus albus* L. *Journal of Experimental Botany* 52, 1063-1070.
- Piquemal, J., Chamayou, S., Nadaud, I., Beckert, M., Barrière, Y., Mila, I., Lapierre, C., Rigau, J., Puigdomenech, P., Jauneau, A., Digonnet, C., Boudet, A.-M., Goffner, D., Pichon, M. (2002): Down-Regulation of Caffeic Acid *O*-Methyltransferase in Maize Revisited Using a Transgenic Approach. *Plant Physiology* 130, 1675-1685.
- Plaxton, W.C. (1996): The organization and regulation of plant glycolysis. *Annual Review of Plant Physiology and Plant Molecular Biology* 47, 185-214.
- Pullerits, T., Sundström, V. (1996): Photosynthetic Light-Harvesting Pigment-Protein Complexes: Toward Understanding How and Why. *Accounts of Chemical Research* 29, 381-389.
- Qin, X., Zeevaart, J.A.D. (1999): The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stressed bean. *Proceedings of the National Academy of Sciences of the United States of America* 96, 15354-15361.
- Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., McCombie, W.R., Martienssen, R.A. (1999): Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genetics* 23, 305-308.
- Ramanjulu, S., Bartels, D. (2002): Drought- and desiccation-induced modulation of gene expression in plants. *Plant, Cell and Environment* 25, 141-151.
- Rapp, J.C., Mullet, J.E. (1991): Chloroplast transcription is required to express the nuclear genes *rbcS* and *cab*. Plastid DNA copy number is regulated independently. *Plant Molecular Biology* 17, 813-823.
- Reddy, P.C.O., Sairanganayakulu, G., Thippeswamy, M., Reddy, P.S., Reddy, M.K., Sudhakar, C. (2008): Identification of stress-induced genes from the drought tolerant semi-arid legume crop horsegram (*Macrotyloma uniflorum* (Lam.) Verdc.) through analysis of subtracted expressed sequence tags. *Plant Science* 175, 372-384.
- Reinartz, J., Bruyns, E., Lin, J.-Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M., Woychik, R. (2002): Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics and Proteomics* 1, 95-104.
- Reynolds, M., Skovmand, B., Trethowan, R., Pfeiffer, W. (1999): Evaluating a Conceptual Model for Drought Tolerance. In: Ribaut, J.-M., Poland, D. (eds.): *Molecular Approaches for the Genetic Improvement of Cereals for Stable Production in Water-Limited Environments. A Strategic Planning Workshop held at CIMMYT, El Batán, Mexico, 21-25 June 1999*. CIMMYT, Mexico D.F..

- Reynolds, M., Tuberosa, R. (2008): Translational research impacting on crop productivity in drought-prone environments. *Current Opinion in Plant Biology* 11, 171-179.
- Ribaut, J.-M., Ragot, M. (2007): Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany* 58, 351-360.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G.-L. (2000): *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science* 290, 2105-2110.
- Rizhsky, L., Liang, H., Mittler, R. (2002): The Combined Effect of Drought Stress and Heat Shock on Gene Expression in Tobacco. *Plant Physiology* 130, 1143-1151.
- Roberts, J.K., DeSimone, N.A., Lingle, W.L., Dure, L. III (1993): Cellular concentrations and uniformity of cell-type accumulation of two lea proteins in cotton embryos. *The Plant Cell* 5, 769-780.
- Roberts, R.J., Vincze, T., Posfai, J., Macelis, D. (2003): REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Research* 31, 418-420.
- Rohlf, F.J. (2006): NTSYSp: Numerical Taxonomy System, ver. 2.20. Exeter Publishing, Ltd., Setauket, NY.
- Rolland, F., Moore, B., Sheen, J. (2002): Sugar sensing and signalling in plants. *The Plant Cell* 14, S185-S205.
- Romo, S., Labrador, E., Dopico, B. (2001): Water stress-regulated gene expression in *Cicer arietinum* seedlings and plants. *Plant Physiology and Biochemistry* 39, 1017-1026.
- Roosens, N.H., Bernard, C., Leplae, R., Verbruggen, N. (2004): Evidence for copper homeostasis function of metallothionein (MT3) in the hyperaccumulator *Thlaspi caerulescens*. *FEBS Letters* 577, 9-16.
- Rothberg, J.M., Leamon, J.H. (2008): The development and impact of 454 sequencing. *Nature Biotechnology* 26, 1117-1124.
- Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Güldener, U., Mannhaupt, G., Münsterkötter, M., Mewes, H.W. (2004): The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Research* 32, 5539-5545.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J. (2003): TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374-378.
- Saghai-Marouf, M.A., Soliman, K.M., Jorgensen, R.A., Allard, R.W. (1984): Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 81, 8014-8018.
- Sakamoto, A., Murata, N. (2002): The role of glycine betaine in the protection of plants from stress: clues from transgenic plants. *Plant, Cell and Environment* 25, 163-171.
- Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2004): *Arabidopsis* Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. *Plant Physiology* 136, 2734-2746.
- Sanders, D., Brownlee, C., Harper, J.F. (1999): Communicating with calcium. *The Plant Cell* 11, 691-706.
- Savenije, H.H.G. (2000): Water scarcity indicators; the deception of the numbers. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* 25, 199-204.

- Schafleitner, R., Rosales, R.O.G., Gaudin, A., Aliaga, C.A.A., Martinez, G.N., Marca, L.R.T., Bolivar, L.A., Delgado, F.M., Simon, R., Bonierbale, M. (2007): Capturing candidate drought tolerance traits in two native Andean potato clones by transcription profiling of field grown plants under water stress. *Plant Physiology and Biochemistry* 45, 673-690.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995): Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* 270, 467-470.
- Schenkel, W., Sticksel, E., Wenzel, G. (2002): Development of a bambara groundnut core collection from IITA germplasm based on characterisation and evaluation data. In: Sesay, A., Edje, O.T., Cornelissen, R. (eds.): *Increasing the Productivity of Bambara Groundnut (*Vigna subterranea* L.) for Sustainable Food Production in Semi-Arid Africa*. Proceedings of a Mid-Project Workshop Held at the University of Swaziland, Kwaluseni Campus, Swaziland, 28-30 August 2001, pp 115-122.
- Schoper, J.B., Lambert, R.J., Vasilas, B.L., Westgate, M.E. (1987): Plant factors controlling seed set in maize. *Plant Physiology* 83, 121-125.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., Shinozaki, K. (2002): Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal* 31, 279-292.
- Senthil-Kumar, M., Govind, G., Kang, L., Mysore, K.S., Udayakumar, M. (2007): Functional characterization of *Nicotiana benthamiana* homologs of peanut water deficit-induced genes by virus-induced gene silencing. *Planta* 225, 523-539.
- Serraj, R., Hash, C.T., Rizvi, S.M.H., Sharma, A., Yadav, R.S., Bidinger, F.R. (2005): Recent Advances in Marker-Assisted Selection for Drought Tolerance in Pearl Millet. *Plant Production Science* 8, 334-337.
- Serraj, R., Sinclair, T.R. (2002): Osmolyte accumulation: can it really help increase crop yield under drought conditions? *Plant, Cell and Environment* 25, 333-341.
- Sharon, N., Lis, H. (1989): Lectins as cell recognition molecules. *Science* 246, 227-234.
- Sharp, R.E., LeNoble, M.E. (2002): ABA, ethylene and the control of shoot and root growth under water stress. *Journal of Experimental Botany* 53, 33-37.
- Shinozaki, K., Yamaguchi-Shinozaki, K. (1997): Gene expression and signal transduction in water-stress response. *Plant Physiology* 115, 327-334.
- Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapiere, C., Jouanin, L., Séguin, A. (2005): *CINNAMYL ALCOHOL DEHYDROGENASE-C* and *-D* Are the Primary Genes Involved in Lignin Biosynthesis in the Floral Stem of *Arabidopsis*. *The Plant Cell* 17, 2059-2076.
- Singr n, C., Schenkel, W. (2003): Fingerprinting of bambara groundnut Germplasm with molecular markers. In: *Proceedings of the International Bambara Groundnut Symposium*, Botswana College of Agriculture, Botswana, 8-12 August 2003, pp 161-173.
- Small, I. (2007): RNAi for revealing and engineering plant gene functions. *Current Opinion in Biotechnology* 18, 148-153.
- Smartt, J. (1990): *Grain legumes. Evolution and genetic resources*. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney.
- Smirnoff, N. (1993): The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytologist* 125, 27-58.
- Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., Efstratiadis, A. (1994): Construction and characterization of a normalized cDNA library. *Proceedings of the National Academy of Sciences of the United States of America* 91, 9228-9232.

- Söderman, E., Hjällström, Fahleson, J., Engström, P. (1999): The HD-Zip gene *ATHB6* in *Arabidopsis* is expressed in developing leaves, roots and carpels and up-regulated by water-deficit conditions. *Plant Molecular Biology* 40, 1073-1083.
- Sokal, R.R., Michener, C.D. (1958): A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38, 1409-1438.
- Solomon, S., Qin, D., Manning, M., Alley, R.B., Berntsen, T., Bindoff, N.L., Chen, Z., Chidthaisong, A., Gregory, J.M., Hegerl, G.C., Heimann, M., Hewitson, B., Hoskins, B.J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Stocker, F., Whetton, P., Wood, R.A., Wratt, D. (2007): Technical summary. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (eds.): *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, New York.
- Soukas, A., Cohen, P., Socci, N.D., Friedman, J.M. (2000): Leptin-specific patterns of gene expression in white adipose tissue. *Genes & Development* 14, 963-980.
- Spyropoulos, C.G., Reid, J.S.G. (1988): Water stress and galactomannan breakdown in germinated fenugreek seeds. Stress affects the production and the activities in vivo of galactomannan-hydrolysing enzymes. *Planta* 174, 473-478.
- Stacy, R.A.P., Aalen, R.B. (1998): Identification of sequence homology between the internal hydrophilic repeated motifs of group 1 late-embryogenesis-abundant proteins in plants and hydrophilic repeats of the general stress protein GsiB of *Bacillus subtilis*. *Planta* 206, 476-478.
- Steele, K.A., Price, A.H., Shashidhar, H.E., Witcombe, J.R. (2006): Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. *Theoretical and Applied Genetics* 112, 208-221.
- Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., De Vries, S.C. (1991): Cell-Specific Expression of the Carrot EP2 Lipid Transfer Protein Gene. *The Plant Cell* 3, 907-921.
- Street, N.R., Skogström, O., Sjödin, A., Tucker, J., Rodríguez-Acosta, M., Nilsson, P., Jansson, S., Taylor, G. (2006): The genetics and genomics of the drought response in *Populus*. *The Plant Journal* 48, 321-341.
- Streeter, J.G., Lohnes, D.G., Fioritto, R.J. (2001): Patterns of pinitol accumulation in soybean plants and relationships to drought tolerance. *Plant, Cell and Environment* 24, 429-438.
- Sugano, S., Kaminaka, H., Rybka, Z., Catala, R., Salinas, J., Matsui, K., Ohme-Takagi, M., Takatsuji, H. (2003): Stress-responsive zinc-finger gene ZPT2-3 plays a role in drought tolerance in petunia. *The Plant Journal* 36, 830-841.
- Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., Schnable, P.S. (2006): All possible modes of gene action are observed in a global comparison of gene expression in a maize F₁ hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America* 103, 6805-6810.
- Syvänen, A.-C. (2005): Toward genome-wide SNP genotyping. *Nature Genetics* 37, S5-S10.
- Tattini, M., Galardi, C., Pinelli, P., Massai, R., Remorini, D., Agati, G. (2004): Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytologist* 163, 547-561.
- Taylor, W.C. (1989): Regulatory interactions between nuclear and plastid genomes. *Annual Review of Plant Physiology and Plant Molecular Biology* 40, 211-233.

- Thompson, A.J., Jackson, A.C., Parker, R.A., Morpeth, D.R., Burbidge, A., Taylor, I.B. (2000): Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-*cis*-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid. *Plant Molecular Biology* 42, 833-845.
- Tinker, N.A., Kilian, A., Wight, C.P., Heller-Uszynska, K., Wenzl, P., Rines, H.W., Bjørnstad, Å, Howarth, C.J., Jannink, J.-L., Anderson, J.M., Rosnagel, B.G., Stuthman, D.D., Sorrells, M.E., Jackson, E.W., Tuvešson, S., Kolb, F.L., Olsson, O., Federizzi, L.C., Carson, M.L., Ohm, H.W., Molnar, S.J., Scoles, G.J., Eckstein, P.E., Bonman, J.M., Ceplitis, A., Langdan, T. (2009): New DArT markers for oat provide enhanced map coverage and global germplasm characterization. *BMC Genomics* 10, 39.
- Torres, T.T., Metta, M., Ottenwälder, B., Schlötterer (2008): Gene expression profiling by massively parallel sequencing. *Genome Research* 18, 172-177.
- Tuberosa, R., Salvi, S. (2006): Genomics-based approaches to improve drought tolerance of crops. *Trends in Plant Science* 11, 405-412.
- Tunnacliffe, A., Wise, M.J. (2007): The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94, 791-812.
- Turner, L.B. (1993): The effect of water stress on floral characters, pollination and seed set in white clover (*Trifolium repens* L.). *Journal of Experimental Botany* 264, 1155-1160.
- Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T., Shinozaki, K. (1999): A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. *Plant Cell* 11, 1743-1754.
- Urao, T., Yamaguchi-Shinozaki, K., Urao, S., Shinozaki, K. (1993): An Arabidopsis *myb* Homolog Is Induced by Dehydration Stress and Its Gene Product Binds to the Conserved MYB Recognition Sequence. *The Plant Cell* 5, 1529-1539.
- Van der Maesen, L.J.G., Somaatmadja, S. (1989): Plant resources of South-East Asia. No. 1. Pulses. Pudoc, Wageningen, p 75.
- Varshney, R.K., Close, T.J., Singh, N.K., Hoisington, D.A., Cook, D.R. (2009): Orphan crops enter the genomics era! *Current Opinion in Plant Biology* 12, 202-210.
- Vierling, E. (1991): The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 42, 579-620.
- Vierstra, R.D. (1996): Proteolysis in plants: mechanisms and functions. *Plant Molecular Biology* 32, 275-302.
- Vignols, F., Rigau, J., Torres, M.A., Capellades, M., Puigdomènech, P. (1995): The *brown midrib3* (*bm3*) Mutation in Maize Occurs in the Gene Encoding Caffeic Acid *O*-Methyltransferase. *The Plant Cell* 7, 407-416.
- Vincent, D., Lapierre, C., Pollet, B., Cornic, G., Negroni, L., Zivy, M. (2005): Water Deficits Affect Caffeate *O*-Methyltransferase, Lignification, and Related Enzymes in Maize Leaves. A Proteomic Investigation. *Plant Physiology* 137, 949-960.
- Von Korff, M., Wang, H., Léon, J., Pillen, K. (2006): AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). *Theoretical and Applied Genetics* 112, 1221-1231.
- Walter, M.H., Grima-Pettenati, J., Grand, G., Boudet, A.M., Lamb, C.J. (1988): Cinnamyl-alcohol dehydrogenase, a molecular marker specific for lignin synthesis: cDNA cloning and mRNA induction by fungal elicitor. *Proceedings of the National Academy of Sciences of the United States of America* 85, 4446-5550.
- Wang, H., Zhang, H., Gao, F., Li, J., Li, Z. (2007): Comparison of gene expression between upland and lowland rice cultivars under water stress using cDNA microarray. *Theoretical and Applied Genetics* 115, 1109-1126.

- Wang, W., Vinocur, B., Shoseyov, O., Altman, A. (2004): Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science* 9, 244-252.
- Wang, Y.-J., Li, Y.-D., Luo, G.-Z., Tian, A.-G., Wang, H.-W., Zhang, J.-S., Chen, S.-Y. (2005): Cloning and characterization of an HDZip I gene *GmHZ1* from soybean. *Planta* 221, 831-843.
- Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C., Ohlrogge, J.B. (2007): Sampling the Arabidopsis Transcriptome with Massively Parallel Pyrosequencing. *Plant Physiology* 144, 32-42.
- Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., Kilian, A. (2004): Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences of the United States of America* 101, 9915-9920.
- Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E., Hearnden, P., Maier, C., Xia, L., Caig, V., Ovesná, J., Cakir, M., Poulsen, D., Wang, J., Raman, R., Smith, K.P., Muehlbauer, G.J., Chalmers, K.J., Kleinhofs, A., Huttner, E., Kilian, A. (2006): A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and phenotypic traits. *BMC Genomics* 7, 206.
- Whetten, R., Sederoff, R. (1995): Lignin Biosynthesis. *The Plant Cell* 7, 1001-1013.
- Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B., Stein, N. (2006): 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 7, 275.
- Willekens, H., Chamnongpol, S., Davey, M., Schraudner, M., Langebartels, C., Van Montagu, M., Inzé, D., Van Camp, W. (1997): Catalase is a sink for H₂O₂ and is indispensable for stress defence in C₃ plants. *The EMBO Journal* 16, 4806-4816.
- Wilson, R.F., Burke, J.J., Quisenberry, J.E. (1987): Plant morphological and biochemical responses to field water deficits. *Plant Physiology* 84, 251-254.
- Wingler, A., Lea, P.J., Quick, W.P., Leegood, R.C. (2000): Photorespiration: metabolic pathways and their role in stress protection. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 355, 1517-1529.
- Wise, M.J., Tunnacliffe, A. (2004): POPP the question: what *do* LEA proteins do? *Trends in Plant Science* 9, 13-17.
- Witcombe, J.R., Hollington, P.A., Howarth, C.J., Reader, S., Steele, K.A. (2008): Breeding for abiotic stresses for sustainable agriculture. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 363, 703-716.
- Wittenberg, A.H.J., van der Lee, T., Cayla, C., Kilian, A., Visser, R.G.F., Schouten, H.J. (2005): Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Molecular Genetics and Genomics* 274: 30-39.
- Wollenweber, B., Porter, J.R., Lübberstedt, T. (2005): Need for multidisciplinary research towards a second green revolution. *Current Opinion in Plant Biology* 8, 337-341.
- Xia, L., Peng, K., Yang, S., Wenzl, P., de Vicente, M.C., Fregene, M., Kilian, A. (2005): DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theoretical and Applied Genetics* 110, 1092-1098.
- Xue, T., Li, X., Zhu, W., Wu, C., Yang, G., Zheng, C. (2009): Cotton metallothionein GhMT3a, a reactive oxygen species scavenger, increased tolerance against abiotic stress in transgenic tobacco and yeast. *Journal of Experimental Botany* 60, 339-349.
- Yang, J., Zhang, J., Wang, Z., Zhu, Q. (2001): Activities of starch hydrolytic enzymes and sucrose-phosphate synthase in the stems of rice subjected to water stress during grain filling. *Journal of Experimental Botany* 52, 2169-2179.

- Yang, S., Pang, W., Ash, G., Harper, J., Carling, J., Wenzl, P., Huttner, E., Zong, X., Kilian, A. (2006): Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theoretical and Applied Genetics* 113, 585-595.
- Yang, Z., Wu, Y., Li, Y., Ling, H.-Q., Chu, C. (2009): OsMT1a, a type 1 metallothionein, plays the pivotal role in zinc homeostasis and drought tolerance in rice. *Plant Molecular Biology* 70, 219-229.
- Yokota, A., Takahara, K., Akashi, K. (2006): Water stress. In: Madhava Rao, M.V., Raghavendra, A.S., Janardhan Reddy, K. (eds.): *Physiology and molecular biology of stress tolerance in plants*. Springer, Dordrecht, pp 15-39.
- Yoshida, Y., Kiyosue, T., Nakashima, K., Yamaguchi-Shinozaki, K., Shinozaki, K. (1997): Regulations of levels of proline as an osmolyte in plants under water stress. *Plant Cell Physiology* 38, 1095-1102.
- Yu, F., Berg, V.S. (1994): Control of Paraheliotropism in Two *Phaseolus* Species. *Plant Physiology* 106, 1567-1573.
- Yue, G., Zhuang, Y., Li, Z., Sun, L., Zhang, J. (2008): Differential gene expression analysis of maize leaf at heading stage in response to water-deficit stress. *Bioscience Reports* 28, 125-134.
- Zeven, A.C. (1998): Landraces: A review of definitions and classifications. *Euphytica* 104, 127-139.
- Zhang, Z., Schwartz, S., Wagner, L., Miller, W. (2000): A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 7, 203-214.
- Zhou, J., Goldsbrough, P.B. (1994): Functional Homologs of Fungal Metallothionein Genes from Arabidopsis. *The Plant Cell* 6, 875-884.
- Zhu, J.-K. (2001): Plant salt tolerance. *Trends in Plant Science* 6, 66-71.
- Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassel, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D., Cregan, P.B. (2003): Single-Nucleotide Polymorphisms in Soybean. *Genetics* 163, 1123-1134.
- Zimeri, A.M., Dhankher, O.P., McCaig, B., Meagher, R.B. (2005): The plant MT1 metallothioneins are stabilized by binding cadmium and are required for cadmium tolerance and accumulation. *Plant Molecular Biology* 58, 839-855.

8. Appendix

Table 15. Bambara groundnut accessions and landraces selected for a) the DArT diversity panel and b) DArT array expansion and genotyping.

a) DArT diversity panel			
Accession/landrace	Country of origin	Seed Source	Cluster no. according to Singrün & Schenkel (2003)
1691/2	Namibia	own stocks	9
1761/3	Namibia	own stocks	8
426/3	Namibia	own stocks	9
AHM753	Namibia	own stocks	11
AHM968	Namibia	own stocks	11
AS-17	South Africa	own stocks	11
Cibadak	Indonesia	own stocks	14
DipC	Botswana	own stocks	12
Dodoma Cream	Tanzania	own stocks	2
Dodoma Red	Tanzania	own stocks	1
GabC	Botswana	own stocks	12
Nyakeni C1	Swaziland	own stocks	13
Nyakeni C2	Swaziland	own stocks	13
Ramayana	Indonesia	own stocks	1
S19-3	Namibia	own stocks	8
Swazi Red	Swaziland	own stocks	14
Tiga Nicuru	Mali	Univ. Nottingham	-
TVsu144	Ghana	IITA	15
TVsu184	Tanzania	IITA	16
TVsu207	Benin	IITA	4
TVsu551	Cameroon	IITA	3
TVsu569	Cameroon	IITA	4
TVsu590	Nigeria	IITA	7
TVsu599	Nigeria	IITA	2
TVsu610	Nigeria	IITA	6
TVsu620	Nigeria	IITA	1
TVsu719	Zambia	IITA	16
TVsu740	Zambia	IITA	6
TVsu747	Zambia	IITA	7
TVsu796?	Madagascar	IITA	16
TVsu841	Nigeria	IITA	3
TVsu866	Zambia	IITA	10
TVsu927	Zambia	IITA	10
TVsu941	Zambia	IITA	17
TVsu999	Zimbabwe	IITA	17
TVsu1011	Zimbabwe	IITA	5
TVsu1033	Zimbabwe	IITA	5
VSSP11 (<i>V.s. var. spontanea</i>)	Cameroon	Univ. Nottingham	-

Table 15 (continued).

b) DArT array expansion and genotyping		
Accession/landrace	Country of origin	Seed Source
426/3	Namibia	own stocks
1761/3	Namibia	own stocks
1691/2	Namibia	own stocks
AHM753	Namibia	own stocks
AHM968	Namibia	own stocks
Ankpa 4	Nigeria	University of Nottingham
AS-17	South Africa	own stocks
Botswana 1	Botswana	University of Nottingham
Botswana 2	Botswana	University of Nottingham
Botswana 3	Botswana	University of Nottingham
Botswana 4	Botswana	University of Nottingham
Botswana 5	Botswana	University of Nottingham
Cibadak	Indonesia	own stocks
DipC	Botswana	own stocks
Dodoma Cream	Tanzania	own stocks
Dodoma Red	Tanzania	own stocks
GabC	Botswana	own stocks
GH CRI-3	Ghana	CRI Kumasi
GH CRI-6	Ghana	CRI Kumasi
GH CRI-8	Ghana	CRI Kumasi
KFBN9501	Namibia	own stocks
LunT	Sierra Leone	own stocks
Mahenene black	Namibia	University of Namibia
Malawi2	Malawi	own stocks
Malawi3	Malawi	own stocks
Manzini star	Swaziland	own stocks
NAV-4	Ghana	CRI Kumasi
NAV-red	Ghana	CRI Kumasi
Nyakeni C1	Swaziland	own stocks
Nyakeni C2	Swaziland	own stocks
OM1	Botswana	own stocks
Parung	Indonesia	own stocks
Ram R93	Botswana	own stocks
Ramayana	Indonesia	own stocks
S19-3	Namibia	own stocks
SB16-5A	Namibia	own stocks
SB2-1	Namibia	own stocks
SB4-2	Namibia	own stocks
Swazi black	Swaziland	own stocks
Swazi butterfly	Swaziland	own stocks
Swazi Red	Swaziland	own stocks
Swazi shade star	Swaziland	own stocks
Tiga Nicuru	Mali	University of Nottingham
TVsu1	Nigeria	IITA
TVsu6	Nigeria	IITA
TVsu9	Nigeria	IITA
TVsu10	Nigeria	IITA
TVsu23	Nigeria	IITA

Table 15 (continued).

Accession/landrace	Country of origin	Seed Source
TVsu85	Burkina Faso	IITA
TVsu88	Mali	IITA
TVsu89	Mali	IITA
TVsu118	Cote d'Ivoire	IITA
TVsu134	Ghana	IITA
TVsu138	Ghana	IITA
TVsu144	Ghana	IITA
TVsu155	Ghana	IITA
TVsu184	Tanzania	IITA
TVsu186	Benin	IITA
TVsu207	Benin	IITA
TVsu243	Gambia	IITA
TVsu246	Gambia	IITA
TVsu292	Burkina Faso	IITA
TVsu369	Tanzania	IITA
TVsu371	Tanzania	IITA
TVsu390	Sudan	IITA
TVsu391	Sudan	IITA
TVsu395	Cameroon	IITA
TVsu400	Cameroon	IITA
TVsu551	Cameroon	IITA
TVsu569	Cameroon	IITA
TVsu590	Nigeria	IITA
TVsu599	Nigeria	IITA
TVsu610	Nigeria	IITA
TVsu620	Nigeria	IITA
TVsu682	Zambia	IITA
TVsu719	Zambia	IITA
TVsu740	Zambia	IITA
TVsu747	Zambia	IITA
TVsu793	Kenya	IITA
TVsu796?	Madagascar	IITA
TVsu841	Nigeria	IITA
TVsu866	Zambia	IITA
TVsu927	Zambia	IITA
TVsu941	Zambia	IITA
TVsu999	Zimbabwe	IITA
TVsu1011	Zimbabwe	IITA
TVsu1033	Zimbabwe	IITA
TVsu1161	Burkina Faso	IITA
TVsu1164	Burkina Faso	IITA
TVsu1191	Burkina Faso	IITA
TVsu1205	Burkina Faso	IITA
TVsu1276	Central African Republic	IITA
TVsu1284	Central African Republic	IITA
VSSP6 (<i>V.s.</i> var. <i>spontanea</i>)	Cameroon	University of Nottingham

Table 16. Oligonucleotide sequences for microarray analysis.

MPSS cluster no.	Annotation	Sequence (5'-3')
1	Proline-rich protein precursor	CAT GCT TGT ACC CTG AAT TCA TAA ATA AAA AGT CTT TTA TGG TTT CAA AA
2	CPRD22 protein	CAT GGA TAT CTA AAA CCT ACC CTG TTG TAG AGC AGG GTG CTT CAG TTG TG
3	Unknown	TAT ATC TTC CAC TTT ATT TTG GGA GTG AAT GCT TTG TGT GGA CAA TAT TT
4	CPRD86 protein	GTG AGA GAA AAT GGT GGT GTG GTT GGT GAT GGA AGT GCT ATG AAT CTG TG
5	Lipid transfer protein I	GAG TTA ATG TTG TAA TGT TAT TTC CCC ATA TAT CCA CTT TAA ATT ATG TT
8	Type I chlorophyll a/b binding protein	CAT GTT CTC TGT GTT CGG GTT CTT CGT CCA GGC CAT TGT CAC CGG AAA GG
9	Ribulose-1,5-bisphosphate carboxylase small subunit	ATG TAT GGG TGC ACT GAT TCT TCT CAG GTG TTG AAG GAA CTT CAA GAG GC
10	RD22-like protein	ACT ATG TAG TTC TCT TCG TAT GCA TAA TAT TTT CTC TTG TTT GTT TAG AT
11	Photosystem II 10 kDa protein	TCA ACT ACT TTG TGG ATA TGT AAT GTT CCT TCC TTC TAT AAA TGC CTG GA
12	Type 2 metallothionein	TCT GTG TAA AGT TGG CTA ATC AGA ATT ATC AGT CTC TGC TTA TAA TAT TA
14	Translationally controlled tumor-like protein	TGA TGA TGG TTG CTT GGT CTT AGC CTA CTA CAA GGA CGG TGC CAC TGA TC
15	Unknown	GGT ACA AGG CTT GGT CTG TAG AAA CTA ATG GTA ATA AAT TTG TGC AGT GT
18	Seed albumin	CTT AAT AAT AAA TAA AAG TGG GTG GTT TAG TTT AAT ACA ATG ATA TTG TT
19	Fructose-bisphosphate aldolase	AAC CTT TTG CTA TGA CGA GTT GTG TAC TTT AAT ATG TCC CCA TAT ATA AC
20	Stored cotyledon mRNA	ATA CAA TAA GCC TTC TAC TGC ATC TAT ATA TGT ACC ATA CTG TTA CTT TC
21	Type 1 metallothionein	GTT CAA AAG TTT TAG AAT TTG TTG TTG GTG TTT TAA GCT ATG GTT ACT TT
24	Lipid transfer protein II	TGT AAT AAT GAT ATA TAA TAG TAT TGG AAT TCC TTC CTA CTT TTT CCG TT
25	No homology	TGG TGT ATG TAA CAA GGA TGT AAT ATG AAT ATA ATT GAC TAT TAC TTG AC
28	Oxygen-evolving enhancer protein 3 precursor	TTC TCA GAC ACT TTG TTG ACT GTT ATT TCA AGA AAA TCA AAT GAT GAG GG
29	ADP-ribosylation factor	CAA GTC TGG CGG TTC TTG GAG AAA GAT GCT TAT GTT TTA TAT AGT TAA GG
33	Ribosomal protein L41	GAG GAA GCG CCG AAA GAT GAG ACA GAG ATC CAA GTA GTG CGT TAA TCC TT
34	Endo-1,4-beta-mannanase	TAT ATA TAT GGG TCC AGT GAT ATA TCA GGG ACA GTT TGT AAG CAT ATA TT
36	LEA5 protein	AAC TAC TTT TAA GTT AAA CAA TGA AGA CTA GGA TAT GGT TCT TAT GAT GT
39	Auxin-repressed protein	GCC TTG AAC TTG GCA TCT GGT TAA TCT TTT TAC CTC TTC TGT TAA GTA TT
41	SUI1 homolog	GAT TAG AGT CTC AAA TAT CCA GAC TTA CCT GCC AAA CTA TAT GCT GGG CA
43	No homology	CAC TTG TAA TAT AAA TTT TCA TCA AAT GAC CCA AAA ATA AAA ACA AAA GA

Table 16 (continued).

MPSS cluster no.	Annotation	Sequence (5'-3')
44	Ferredoxin I	GTT GTG TCT GGT CAA GTC GAC CAA TCA GAC GGT AGC TTC CTT GAT GAC GA
46	No homology	TTG AGC GGT TGA AAT CGT GGA ATG ATG GAG TGG TAG CTT TGG GTG GTT TA
49	Glyceraldehyde-3-phosphate dehydrogenase A subunit	GCC ACT ACC ACA CTT GTT GCT TCA TTC CTT CTA AAC ACC TAC TGA GAA AC
50	Aquaporin-1	GTG TGA TGT ACG GCT GTG ATT GTG TAT GAT TAT GCA AAA TCA AAA ATA AC
52	LEA4 protein	TTT ACT GGT TGC TGT AAA AGT GTT TTC ACT GTA GTG TAT CGT TCC GTT GC
61	Type II chlorophyll a/b binding protein	TAA TAT GGC AAC TTC TGC ATA TAT GCA CAG AAT GAT AAT CCA ACT TGC CG
63	No homology	ATT GGA GCT CTT ACT TAA ATA CAT TCA AGC CTT GTG AAT AAC TCC ATC CA
64	Cinnamyl alcohol dehydrogenase-like protein	CAT GTG GGT GTT GTT GGT CTT GGT GGC CTT GGT CAT ATG GCT GTT AAG TT
65	Initiation factor 5A(2)	ATG GTG GAA AGC TGA AGT TGT AAC TGA GAT GAT ATA GTA TTT ATT TCT CT
66	Catalase	GTT TCC CTG CTA TTG TTG TGG ACT ATG TGT TCC TCG TTA TTC TAT TAA GA
68	No homology	GTT AAG TCC TTC CGA TCA TTG TAA GTA AAA TAA TTA CGA GGT ATT AAG AG
70	Ubiquitin-like protein	TCG AAC TCT ACT ACA ATT AAA CGT GCG TTT TAT ACA CAG TAT ATT ATG TG
71	Dehydrin	GTT TGA AGT GTA ATG AAA AGA AAA TGT ATT GAA ACA AAT TTA GTT AAA TA
74	No homology	GTT TGA TGT GGA TGT TAT GTT ACA TAA GGT GTG AAG TTT GGT GTG TGA AG
76	Unknown	TCA CTT GTA TAT CTA TTT CTG AAC TTT CAT CGT TTT CCT CTT TCC TTC GT
77	Ubiquitin-conjugation enzyme	TGT AAC TTT CTC TAT CGA GTA ATG TAG AAA TAG TGA TAT GTC GCA CAA GT
80	No homology	AAC TAT ATA TAA GAA ATA TGT ATT AAA GAA GAG TAT GGT TTT GTA CGT TT
82	No homology	TTA CGT AGC AGA GTT AGT TCT TGC TCT TGT AAC CTT TTA CCT GTT ACG TG
84	Cyclophilin I	TGG TCA TAG AGT GAT TGG AGT GTG TTA TGA GAA ACA TAA AAA TGA TGA GC
86	Cysteine proteinase	TAG AGC ATC TGA TGC CTG TAT TAA TGA GTA AAG ATA CGT GCA TTG TAT AA
89	DnaJ-like protein	ATG ATG ATG ATA CTT TTG ATG TTC CGT AAC TGT AGC GTG GAC TAT GGA CT
93	Pathogenesis-related protein 4.2	TGA TCC AAC CCA GTT ATC AAT ATC AGT GAT CAG TAA CAA GCT TTA TGT GA
97	No homology	ATG AGG AGT ACT CTC TCA CTG TTA TAT TAT AAT AAT AAG TTT ATT ATA TT
98	Aluminium-induced protein	GGT CTG GTA GCA GTA GTT TGT TCA TTG GAT AGG ATT TAC TAT GTT TCT CG
99	Caffeic acid methyltransferase	CTT CTG GCT CAC AAC GCT GGT GGG AAA GAG AGG ACC GAA GAG AAT TGG AA
101	No homology	TGC CCT GGT ATC GTT TGC ATT AGA GAT TGT ATA CTC GTT ATA TTA TTA AA

Table 16 (continued).

MPSS cluster no.	Annotation	Sequence (5'-3')
102	CPRD14 protein	TGT CTG TTT AAT GCA TTA TAT TTG TAT CAT TTG TAG CTT TAA TGG AAA AT
107	Unknown	ACT TCT TGG AAA ATA GGT GGT ATG TCT AAA TTA TAT GAG AAT AGC ACA AT
110	No homology	CAT GAA AAC TCG TGC TTG ATG CCT CTA CCA TAT GCT TTG GAT GTA ATA AA
111	No homology	AGG TTC ACA TCT GTT TGA AAT ACC TAT AAT GAA TCT TTC GAT ACA GAA CA
112	<i>o</i> -Methyltransferase	TGG GCG TTT CTG TCA TTC AAC TCA GAT ATT AGT GTT TCT TTA CTG TAT CA
114	β -Amylase	TTC TTG GAA CAT CCT TCT TCT GTT GTT TAT GAT AGG GAT TAG GAA ACT GT
115	Unknown	AAA AGG GTC ATT ATT TAA GCT TTA AAT AAA TAC AAA TGT ATT ATT GCA TA
116	No homology	TCA GAT GAT GGA ATC CTT GTA AAG CAC GGA TGT GTG TTT GCT TGT TCT AT
117	RNA-binding protein	TGT ATA TTT AGA AAA TCT GTT CGG ATT TGC TAC AAT ATG AAG TTG TGA AA
124	No homology	TCT GTG TTA TCA ACT TCA TTT AAA TAC ACA TCT TCG TCA TAC ATA CCT GC
126	Unknown	TAC TTC ATC AGA GGA AAT CAC TTC CTT ACA GCT ACA CAA CAA TGG CTA TA
129	Ribosomal protein S28	GTG AGA GAA GGA GAC ATT CTC ACC CTA CTC GAA TCT GAG AGG GAA GCA AG
130	Glutathione peroxidase	ATG TAC TGA AAC AGT TGC TCT TGT ACC TGA TAT TAT TAT GTT CAT AAC TA
133	Hypothetical protein	AAG CAG GTT CTG TAA AGA TAA ACA ATG GTG AAA CTC TGA CTT TAG AGT CT
139	Acireductone dioxygenase	CAT GCT GTT AAT GCT GCT GCG TAA AGA TCT GGT CTG CTT GAT CTA TGC AA
143	CDPK-related protein kinase	GCA TCT CAA CCT TTT CGC ATC TTT TCT TTT CTC ACT CTC TTC CAT TAT AT
144	Serine hydroxymethyltransferase	AAA GCC TTT GTA TGT ATG AAT CTA GAA AAT CTT GCT CAT TAT CTT CTT GA
152	<i>myo</i> -inositol oxygenase 5	TTT ATG ATG ATA GCC AAA CAG TTT CTT ATG TAG TTG AGG AAA TAA TAA GC
154	No homology	GGT TGA TGT AAT GGT TAT TCA GTA CTT GTG TGT AAA TAA GTG TTG TTT TC
155	Photosystem I psaH protein	GTT CAA CCA GCT GCC ATC AAT GGC CTT GCC GGA AGC TCC CTC ACC GGA AC
157	Isoprenylated protein	CTT CAA GGA GGA GAT GAA ACG AAA GCA TCA ACA CCA TTA CCG CGA GGC AT
158	Heat shock protein 22	GGA GGA CGA TGT GCT TGT CAT AAA TGG TGA TCA CAA GAG TGA ACA AGA AC
159	Catalytic subunit of protein phosphatase 1	CTA TTG GTT GCA AGG GAA AGT CAA GTT CCT TTC GTC TAT AAT ATT TTG GA
161	Ribosomal protein S27	GGA TGT TAA GTG TCA GGG TTG CTT TAA CAT AAC GAC TGT GTT TAG CCA CT
162	Phosphoglycerate kinase	ATT GAC TTC ACT CAT AAA TTG CGT GTA ACT GTA AAC ACT GGC TGA GAG CC
166	Invertase inhibitor	TAC CTG TTA TGT ATA AAT GCT GGT ATC TAT TTT AAA GCT CTT ATC ATA AG
167	UDP-glucuronosyltransferase	GTG CTT TCT CAT CCA TCT GTT GGT GTT TTC TAA CCC ATT GTG GTT GGA AC

Table 16 (continued).

MPSS cluster no.	Annotation	Sequence (5'-3')
168	Unknown	TGG AGC CAT AGT TGC TTG GCT AGT TTG CTT TGA GCT GTT GTA GAG TAA TT
169	No homology	AAA ATA ATT GAA GTT TCT GTC TTC CGC GCA GAT TAC TGC ATT TAT CAC GC
174	Unknown	AAT AAC GTC CTT GTT TTC TGT TCA ATT TGT ATA TTG TGT TCT GTC TTG TT
179	Elongation factor EF-1a	TCA AGA ACG TTG AGA AGA AGG ATC CCA CTG GAG CCA AGG TCA CCA AGG CT
182	Cytidine or deoxycytidylate deaminase	TGA ACC TTG CCC AAT GTG CTT CGG AGC AAT TCA CCT TTC ACG AGT TAA GA
185	1-Aminocyclopropane-1-carboxylic acid oxidase	TAC TAA TGT AAT ACA GTG TTC CTT TAT TAT TGT GTA TTG TAA TTC GTA CT
188	Glutamine synthetase	ATA TGA AAA TGC ACA TCA AGT TTC GTT GGT ACT ATT TGC TTC AGG ACA AA
189	Val-tRNA	TGT TGG GTT TTG AAA ACA AAT TAA ATC ATT TAG ATA ATA ATT AGT TTG AT
194	Histone H1D	GCG AAT TTC AAG AAG ATT CTA GGT CTG CAA TTG AAG AAT CAA GCA GCG AG
195	No homology	TTT TCT GGG AGA AAT TGT TCG AAA GAA TCA TCA CTG GTT TTC TTA AGA GA
209	Heat shock protein associated protein	CAA TGT AAT ATT TAT TTT CCT GAG TGA ATT TAA TTG TTC AAA GTT GAT AT
212	DNA-directed RNA polymerases I, II, and III 7kDa subunit	GTG CCG CGA GTG CGG TTA CCG TAT CCT TTA CAA GAA GCG CAC TCG TCG AA
217	Cyclin-dependent kinases regulatory subunit	AAC ACT TCC TTT GCG AAT CAG AAA GTG GGA ATA CAA TGT TGT TAT GTA GG
226	Early light-induced protein	GAA TTC CTG CTA TGT ACT TGA GAA GGA ATG TTA GCC TGA GAG TTA GGT CC
227	Transcription factor MYB139	CAT AAC AGT ATC TTT GAG CTT TCG GCA CGG TCC TCT GTA GTG TGA CAC AG
230	Photosystem II core complex protein psbY	AGC CGC TGC TGA AGT CGC GGC GAT CGC CGA AGC CGC GAG CGA CAA CAG GG
239	Alcohol dehydrogenase 1	TGG AAG AGT GAA TGA GTT ATG TTC TCT GGT GTG AGA ACA GTG TTT GTC TG
243	Snakin-like cysteine rich protein	GTC CTC GTC TCT GCC AAA GAG CTT GTG GAA CTT GTT GCA GAC GCT GCA AC
261	Ferritin	CAG TCT CTG TTT ACA TTC GCT CTT CCA TTA CAT TCT GGG CTG TTT TCA AA
262	Unknown	CAC TAT CAC GCA CAA GTG GTG TCA TCT TAA TTG TAT GTG CCA CCA TCG TA
270	Elongation factor EF-2	CAT GCA GCC AGC GAT TGC AGA CTT TGT TAT TAT AAT GGT TTG ATT TTG GA
275	Ras-related GTP-binding protein	TAT CCT ATT CCG ATA CTG TGA AAG AAA ACT GCG TAA GAT CTT TTT GAA GT
300	Polyubiquitin 2	TCT TAT GTC TAA TGT CAA TTG AAA TCG TGT GTT ACA GCT CTT GGT TGT GT
307	Cytochrome c oxidase subunit	GCT ACC TTG AAA GGT CCA AGT GTG GTA AAG GAG ATA TTA ATT GGA ATA AC
323	Ribulose-5-phosphate-3-epimerase	CTG GCT TGC TGT CAC CAC TGC CAA TTG CTT TTC TCT GGA ATT AAC TAT AA

Table 16 (continued).

MPSS cluster no.	Annotation	Sequence (5'-3')
325	Lectin	AGC GAA GAA ATA TTC GGT TTC ATT TGT GGT CAA CGT TAA AGA AAA CGG TT
336	Unknown	TTT GGC AGT GTT GGT GAT AAT GGT GAT AGT GAC AAA GAG AGT TTA GTC TC
349	Glyoxysomal malate dehydrogenase	TTT TGT TGA TTC TCA GGT TAC GGA ACT TCC CTT CTT TGC AAC CAA GGT AC
355	snRNP-related protein	CCA GTT TAT GAA TTT AGT TGT TGA CAA CAC CGT GGA AGT TAA TGG CAA TG
357	Small GTP-binding protein	CTA GTC AGC CCC TTC CAG ACG ACG ACG ATG ATG CAT TTG AGT AGG AAT TG
374	Ribosomal protein S17	GTT CCC GTA CAG GCA CCG CTA GCC TTC GGT CGC GGC GCC GCC GGA AGG AG
377	RNA binding protein 47	AAT TCA AGG GGC TTC CTG ACG GGG TTA GCT GAA TTG TGT GCT GAA GTA GG
393	Transcription factor bZIP33	AGT CTT ATG TGA TGT GTT ATG TGT TGT TTT GCA TCA AAA TCG TGT GGT CC
398	Glyceraldehyde-3-dehydrogenase	GAG TAG TCT GAA TAA ATC GGT TTC CGG AAC CAT TGT TGT TAC CCT GCT AG
412	14-3-3 protein	CTT CTC CGT GAC AAT CTG ACT TTG TGG ACA TCG GAC ATC ACG GAC GAT GG
418	No homology	TAT GAT ACG TAT TTA GGT TAA AAA TGG TAT GTG TGT ATG TAT ATA TGT AT
438	Salt-tolerance protein	TAT CTG TGT AGA TTG ATT ACT TCT CCT AGG ACA GTG AGT AAT GAA GTA TG
442	Unknown	CCA ACC CAA AAT TGG ATT TTG TTC ATT GTA TTT TAT AAA TGG TGT GAT AA
451	Heat-shock protein 81-1	AGG AGA GCA AGA TGG AAG AAG TAG ACT AAA TGC CAA CGA TCG TTT GTT TT
452	Progesterone 5-beta-reductase gene	CAT AAG TTG GAT AGA TAA GAG TAA GGG TTA TAA GAT TGT GCC TTG AAT AA
454	Copper/zinc-superoxide dismutase	TGT CCA TCT AAA GAC AAG TGC GCA AAT GAA TCC TGA AAA CAT ATC AGT CC
458	No homology	TCC GTT TTC CTT ATT GGA AGG ATA AAT CAG TAT CAC TTT TGA TTA GAA CT
472	Protease precursor	GGA GTC TAA GGG AAC TGC TTC TTG AAC AAC TGC ATT TTC TTT ATA TCA CA
479	CONSTANS-like 2 protein	TCG CAG TTG CAC TAA TCG GAT GCT TTT ATT ATT TTG TTT ACT GTA TTT TG
494	No homology	TGC TAA TAA AAT CTG AAT GTG AAT ACA AAT TAC TTT ACT GTA AAA TTA TT
496	S-adenosylmethionine decarboxylase	TTT TGT GGA TGT GAA AGG ATA CTA TCG TGA GGA GTG GAG CCA CGA AGG GC
551	Selenium binding protein	GGT GAC TGT ACT TCA GAT ATA TGG GTT TAA CCA CAC ATC TCA TTA TTT TC
555	Transcription factor MYB123	CCC CTC AAC GGT GTA CCT ATA TTG TTT ATT TCT AAT TCT GCA GCT CAC GA
617	Cysteine proteinase precursor	ACA GTC GAA CAT TGG CTG TGT ACA TAG GGC CAT ATT ATT TAT GCT CTT GT
766	HDZip I protein	TTT TAT CCT ATG TTT GTT TGT GTG TAG TGT TGT TTA TCT AGT ATT TGC AC
811	Transcription factor MYB93	TAA GGG TGT TCC TTG CCC GGC TCA TCA ATG TGG TCT TGT GTT GTT CAT CA
823	Transcription factor bZIP 6	TGT CAC CAT TGG TTA GAC AAT CTT GTG AAG TAG CAC TTT TCC TGG ATT TG
1236	Suppressor of CONSTANS1	ATA ATT TAA TGA TTA GGG TTT AGT TAA CCC CTG TCT TCC TCA TTT CTG AC

Table 17. Induction (positive) or repression (negative) factors for four bambara groundnut genotypes and BLAST sequence homologies for the 561 largest clusters (transcript abundance ≥ 40) in the 2006 MPSS expression profiling experiment. 0 indicates no change after the water-deficit stress treatment, while ± 1 stands for induction/repression by the factor 2 (n.d. = not detected). 454 sequencing-derived transcript tags were first searched for homologies to annotated mRNAs/genes in the ‘non redundant’ BLAST nucleotide databases. When no significant matches were obtained, the search was extended to unannotated non-human and non mouse EST (‘est_others’) databases and hits were used as queries in the ‘non-redundant’ databases again.

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Homology to sequences in ‘non-redundant’ (NR) and ‘est_others’ (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT	Mean				
1	49	-2.1	-0.7	-1.6	-0.1	-1.1	EST: PV_GEa0012C_G10.b1, <i>Phaseolus vulgaris</i> NR: proline-rich protein precursor mRNA, <i>Phaseolus vulgaris</i>	CV530090.1 U38229.1	9e-6 0	40/44 467/467
2	98	113.8	131.2	807.9	1008.1	515.2	NR: CPRD22 protein mRNA, <i>Vigna unguiculata</i>	D83972.1	8e-18	71/77
3	107	5.4	3.7	17.1	12.4	9.7	EST: CG18, <i>Phaseolus vulgaris</i> NR: no homology	EC997018.1 -	2e-12 -	70/80 -
4	99	10.5	2.3	11.5	14.2	9.6	EST: BE-1533, <i>Phaseolus vulgaris</i> NR: CPRD86 protein mRNA, <i>Vigna unguiculata</i>	EX304281.1 AB030294.1	e-13 3e-163	60/65 455/546
5	85	5.1	2.3	45.8	90.7	36.0	NR: lipid transfer protein I mRNA, <i>Vigna radiata</i>	AY300806.1	e-9	44/47
6	81	-14.3	-20.8	-59.5	-34.8	-32.4	EST: RF#1-T3_B10_078, <i>Vigna unguiculata</i> NR: no homology	ES884117.1 -	e-6 -	42/46 -
7	55	-0.5	0.3	-0.3	0.5	0.0	EST: 6-55, <i>Vigna unguiculata</i> NR: Drm3 mRNA, <i>Pisum sativum</i>	CB968062.1 AF515795.1	5e-5 5e-84	48/53 300/379
8	109	-7.0	-5.0	-32.8	-63.5	-27.1	NR: LHCII type I chlorophyll a/b binding protein mRNA, <i>Vigna radiata</i>	AF139467.2	e-41	95/97
9	92	-15.0	-9.7	-30.8	-406.1	-115.4	NR: ribulose 1,5-bisphosphate carboxylase small subunit mRNA, <i>Glycine max</i>	AF303941.1	7e-24	80/87
10	101	2.6	1.3	9.3	4.7	4.5	EST: NJ0553, <i>Phaseolus vulgaris</i> NR: RD22-like protein mRNA, <i>Vitis vinifera</i>	EC911284.1 AY634282.1	e-16 8e-128	65/70 495/652
11	127	-3.2	-2.8	-3.6	-5.5	-3.8	EST: BE-2176, <i>Phaseolus vulgaris</i> NR: chloroplast photosystem II 10 kDa protein mRNA, <i>Arachis hypogaea</i>	EX304046.1 DQ296038.1	3e-33 2e-132	106/115 337/384
12	85	0.4	0.9	1.5	1.8	1.1	NR: type 2 metallothionein mRNA, <i>Lablab purpureus</i>	AB176567.1	2e-27	71/73
13	96	-4.0	-2.5	-3.9	-1.4	-2.9	EST: PVEPSE3015H07, <i>Phaseolus vulgaris</i> NR: type 1 metallothionein mRNA, <i>Vigna angularis</i>	CV541493.1 AB176560.1	7e-18 5e-73	74/82 315/416
14	98	0.3	0.8	0.2	1.0	0.6	NR: translationally controlled tumor-like protein mRNA, <i>Glycine max</i>	AF421558.1	5e-25	79/85
15	104	0.3	-0.2	1.2	0.7	0.5	No homology	-	-	-
16	87	-8.2	-5.8	-43.0	-77.4	-33.6	NR: rubisco activase, <i>Vigna radiata</i>	AF126870.2	8e-8	59/68

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
17	84	-0.7	-0.6	-0.9	-0.8	-0.7	EST: Gm-c1073-4967, <i>Glycine max</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	BQ628869.1 EF147789.1	2e-11 4e-40	71/82 217/293	
18	79	236.9	166.1	193.6	44.9	160.4	EST: TC-T7-ST3_E06_020, <i>Vigna unguiculata</i> NR: seed albumin mRNA, <i>Vigna radiata</i>	CK151447.1 X70671.1	8e-5 e-163	33/35 493/614	
19	115	-6.6	-3.1	-20.2	-185.5	-53.8	EST: Gm-c1018-1459, <i>Glycine max</i> NR: putative fructose-bisphosphate aldolase mRNA, <i>Trifolium pratense</i>	BI941659.1 AB236776.1	e-10 3e-78	66/75 282/353	
20	90	-4.5	-21.1	2.3	0.5	-5.7	NR: stored cotyledon mRNA, <i>Glycine max</i>	X16877.1	2e-8	33/33	
21	116	-2.3	-1.5	-0.3	0.1	-1.0	NR: type 1 metallothionein mRNA, <i>Lablab purpureus</i>	AB176566.1	3e-20	77/85	
22	106	0.0	-1.2	0.3	0.1	-0.2	EST: Gm-c1050-3151, <i>Glycine max</i> NR: no homology	BI470292.1	e-13	80/92	
23	106	-2.7	-1.4	-3.6	-7.9	-3.9	NR: photosystem II reaction center mRNA, <i>Retama raetam</i>	AF439283.1	2e-37	100/106	
24	74	1.4	0.4	7.8	4.9	3.6	EST: PV_GEa0015B_E07.b1, <i>Phaseolus vulgaris</i> NR: lipid transfer protein II mRNA, <i>Vigna radiata</i>	CV530922.1 AY300807.1	7e-8 e-162	56/64 524/649	
25	70	18.2	43.3	95.3	34.9	47.9	No homology	-	-	-	
26	118	-61.9	-13.8	-39.7	-173.4	-72.2	NR: LHCII type III chlorophyll a/b binding protein mRNA, <i>Vigna radiata</i>	AF139465.2	5e-41	107/112	
27	95	-4.1	-7.1	-16.1	-11.3	-9.6	EST: Gm-c1057-4632, <i>Glycine max</i> NR: thioredoxin F isoform mRNA, <i>Pisum sativum</i>	BM524221.1 X63537.1	8e-27 5e-120	76/80 355/423	
28	107	-6.5	-12.8	-22.0	-37.9	-19.8	EST: Gm-c1054-5624, <i>Glycine max</i> NR: oxygen-evolving enhancer protein 3 precursor mRNA, <i>Pisum sativum</i>	BQ295912.1 AY292531.1	e-26 3e-130	70/72 375/450	
29	105	1.4	0.4	1.3	0.9	1.0	NR: ADP-ribosylation factor mRNA, <i>Vigna unguiculata</i>	AF022389.1	2e-40	96/99	
30	117	0.9	0.2	0.5	0.3	0.5	EST: Gm_ck24678, <i>Glycine max</i> NR: armadillo/beta-catenin repeat family protein mRNA, <i>Arabidopsis thaliana</i>	CD402153.1 NM_130135.3	e-13 3e-37	55/58 324/474	
31	95	-33.6	-8.6	-64.8	-47.3	-38.6	EST: PV_GEa0014A_H10.b1, <i>Phaseolus vulgaris</i> NR: light-harvesting complex II protein Lhcb5 mRNA, <i>Populus trichocarpa</i>	CV530559.1 XM_002329156	e-22 8 ^e -139	69/72 430/529	
32	90	-13.5	-3.8	-20.5	-140.9	-44.7	NR: LHCII type II chlorophyll a/b binding protein mRNA, <i>Vigna radiata</i>	AF279248.1	5e-31	80/83	
33	94	-3.4	-0.3	-0.5	-0.2	-1.1	NR: ribosomal protein L41 mRNA, <i>Pisum sativum</i>	L47967.1	5e-34	79/80	
34	93	217.1	16.3	69.8	130.5	108.4	EST: PCSC10644, <i>Phaseolus coccineus</i> NR: endo-1,4-beta-mannanase mRNA, <i>Glycine max</i>	CA913729.1 DQ812101.1	e-13 3e-19	71/78 59/60	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
35	58	-1.6	-1.2	-3.2	-6.5	-3.1	EST: PvEST0398, <i>Phaseolus vulgaris</i> NR: photosystem I subunit PsaD mRNA, <i>Glycine max</i>	EG588007.1 EF628505.1	3e-9 0	40/42 554/639	
36	70	37.5	103.6	61.1	160.6	90.7	EST: LR0968, <i>Phaseolus acutifolius</i> NR: desiccation protective protein LEA5 mRNA, <i>Glycine max</i>	EC911772.1 U66316.1	7e-11 e-84	40/41 319/414	
37	85	-3.2	-3.0	-22.3	-28.7	-14.3	EST: LVS_037_B10, <i>Phaseolus vulgaris</i> NR: photosystem I subunit X precursor mRNA, <i>Retama raetam</i>	CV533255.1 AF439284.1	9e-11 e-85	50/53 219/245	
38	86	-2.1	-2.9	-4.2	-4.5	-3.4	EST: PCS03519F, <i>Phaseolus coccineus</i> NR: PSI reaction center subunit III, <i>Citrus sinensis</i>	CA901755.1 AB002095.1	e-19 2e-19	52/52 241/302	
39	108	4.9	3.4	16.9	12.0	9.3	EST: RTS_112_B12, <i>Phaseolus vulgaris</i> NR: auxin-repressed protein mRNA, <i>Sesbania drummondii</i>	EH791060.1 EF564346.1	7e-6 6e-112	45/49 443/578	
40	96	-2.4	-3.5	-6.7	-10.2	-5.7	EST: LVS_043_D02, <i>Phaseolus vulgaris</i> NR: photosystem I reaction center subunit XI mRNA, <i>Zea mays</i>	CV533777.1 EU956354.1	8e-24 5e-51	65/67 202/259	
41	92	0.4	0.6	0.4	0.3	0.4	EST: TC-T7-CC1_A01_008, <i>Vigna unguiculata</i> NR: SUI1 homolog mRNA, <i>Salix bakko</i>	CK151408.1 AB003378.1	8e-24 2e-95	80/87 290/350	
42	77	-48.0	-43.3	-4.9	-299.0	-98.8	EST: NOD_235_F06, <i>Phaseolus vulgaris</i> NR: no homology	CV536924.1 -	e-9 -	41/43 -	
43	72	8.7	7.9	10.3	7.4	8.6	No homology	-	-	-	
44	84	-5.2	-1.4	-6.1	-14.5	-6.8	NR: ferredoxin I mRNA, <i>Trifolium pratense</i>	AY340639.1	e-9	41/43	
45	115	-3.1	-2.2	-1.0	-2.7	-2.2	EST: PCS03310, <i>Phaseolus coccineus</i> NR: psaA-psbB fragment, <i>Jasminum subhumile</i>	CA911942.1 DQ673259.1	9e-15 2e-45	92/109 128/143	
46	90	2.1	2.6	9.1	5.2	4.7	No homology	-	-	-	
47	104	-0.2	0.3	-0.4	0.9	0.1	EST: RTS_140_B02, <i>Phaseolus vulgaris</i> NR: hypothetical protein mRNA, <i>Cicer arietinum</i>	CV543835.1 AJ012688.1	6e-22 4e-65	71/76 237/297	
48	117	-1.8	-1.9	-8.1	-6.6	-4.6	EST: Gm-c1069-7273, <i>Glycine max</i> NR: precursor for 23-kDa protein of photosystem II mRNA, <i>Pisum sativum</i>	CA953443.1 D13296.1	2e-15 e-64	61/65 215/246	
49	96	-5.3	-2.9	-7.6	-13.5	-7.3	EST: PVEPSE2030E04.g, <i>Phaseolus vulgaris</i> NR: glyceraldehyde-3-phosphate dehydrogenase A subunit mRNA, <i>Glycine max</i>	CB544161.1 DQ224370.1	8e-24 3e-78	68/71 194/211	
50	72	0.3	0.0	0.0	0.3	0.1	NR: putative aquaporin-1 mRNA, <i>Phaseolus vulgaris</i>	U97023.1	5e-21	66/70	
51	102	2.8	3.1	113.1	58.1	44.3	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
52	110	27.3	13.8	115.6	219.2	94.0	EST: UCRVU01_J3_T7, <i>Vigna unguiculata</i> NR: seed maturation protein LEA 4 mRNA, <i>Glycine canescens</i>	CK151425.1 AY044271.1	2e-9 5e-65	50/55 193/227	
53	101	-3.9	-2.7	-15.3	-30.4	-13.1	EST: 25CA145, <i>Phaseolus vulgaris</i> NR: photosystem I light-harvesting chlorophyll a/b-binding protein, <i>Nicotiana tabacum</i>	EC997017.1 X64198.1	3e-20 e-10	83/93 64/79	
54	90	-13.8	-3.7	-135.9	-55.8	-52.3	NR: carbonic anhydrase mRNA, <i>Vigna radiata</i>	AF139464.2	2e-21	55/55	
55	98	-4.8	-2.6	-12.4	-13.3	-8.3	NR: chlorophyll a/b binding protein CP29 mRNA, <i>Vigna radiata</i>	AF139466.2	2e-42	96/98	
56	62	-2.8	-3.8	-24.0	-41.5	-18.0	No homology	-	-	-	
57	113	-7.0	-4.6	-34.7	-85.6	-33.0	NR: 23S ribosomal RNA <i>Vigna radiata</i>	AF322910.1	e-28	73/75	
58	99	-2.4	-0.4	-5.1	-3.0	-2.8	EST: Gm-c1049-9338, <i>Glycine max</i> NR: precursor for 33-kDa protein of photosystem II mRNA, <i>Pisum sativum</i>	BU083437.1 D13297.1	2e-6 e-90	33/34 248/287	
59	99	-0.8	-2.5	-1.2	-0.2	-1.2	No homology	-	-	-	
60	65	-0.6	-0.5	-0.5	-0.3	-0.5	NR: histone H3 (H3) mRNA, <i>Robinia pseudoacacia</i>	DQ917751.1	e-5	37/40	
61	90	-2.1	-1.5	-6.8	-11.1	-5.4	EST: Gm-c1049-7126, <i>Glycine max</i> NR: type II chlorophyll a/b binding protein mRNA, <i>Pisum sativum</i>	BU084452.1 X81962.1	2e-8 e-117	71/82 336/398	
62	60	0.3	-1.4	-0.9	-0.2	-0.5	No homology	-	-	-	
63	92	-0.8	0.1	0.7	4.5	1.1	No homology	-	-	-	
64	101	-0.6	0.9	1.4	-0.4	0.3	NR: cinnamyl alcohol dehydrogenase-like protein gene, <i>Lotus corniculatus</i>	AY028929.1	7e-15	71/80	
65	97	-0.8	-1.1	0.7	0.0	-0.3	EST: PCS05506_3', <i>Phaseolus coccineus</i> NR: initiation factor 5A(2) mRNA, <i>Nicotiana plumbaginifolia</i>	CA905615.1 X63542.1	4e-32 e-18	91/97 71/81	
66	95	0.5	0.0	-0.4	-0.6	-0.1	NR: catalase mRNA, <i>Vigna radiata</i>	D13557.1	7e-18	87/98	
67	96	-1.5	-1.6	-9.3	-12.9	-6.3	EST: PVEPSE3028A16.g, <i>Phaseolus vulgaris</i> NR: No homology	CB556075.1	3e-20	74/81	
68	81	94.4	15.8	110.1	64.8	71.3	No homology	-	-	-	
69	108	-2.2	-1.1	-10.3	-4.4	-4.5	EST: Gm-c1048-5950, <i>Glycine max</i> NR: peroxisomal glycolate oxidase	CA937570.1 AB333790.1	5e-7 e-108	31/31 222/222	
70	104	-2.3	-0.5	-0.1	0.4	-0.6	EST: UWA111, <i>Lupinus angustifolius</i> NR: ubiquitin-like protein mRNA, <i>Arabidopsis thaliana</i>	DT454412.1 AK220599.1	3e-5 6e-57	28/28 187/227	
71	83	2.8	1.9	3.6	5.4	3.4	EST: LVS_041_A04, <i>Phaseolus vulgaris</i> NR: dehydrin mRNA, <i>Phaseolus vulgaris</i>	CV533584.1 U54703.1	8e-5 9e-128	45/51 276/286	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT	Mean				
72	91	-2.0	-0.9	-1.5	-5.6	-2.5	EST: HDI_J_11.esd, <i>Macrotyloma uniflorum</i> NR: putative Cys2-His2 zinc finger transcription factor mRNA, <i>Juglans regia</i> NR: 4-coumaryl-CoA ligase mRNA, <i>Plantago major</i> NR: phosphoenolpyruvate carboxykinase, <i>Penaeus vannamei</i>	DR989417.1 AJ937310.1 AM159093.1 AJ250829.1	6e-6 e-25 e-25 4e-25	39/41 75/78 77/81 74/77
73	108	-0.5	-2.0	-0.9	0.2	-0.8	No homology	-	-	-
74	128	0.9	0.4	0.5	-0.6	0.3	No homology	-	-	-
75	89	-1.3	-2.1	-8.0	-16.9	-7.1	NR: chlorophyll a/b-binding protein CP24 precursor mRNA, <i>Vigna radiata</i>	AF139470.2	e-37	82/82
76	102	-2.6	-6.1	-9.1	-13.0	-7.7	EST: PCS04409, <i>Phaseolus coccineus</i> NR: no homology	CA901345.1 -	e-16 -	80/90 -
77	98	0.6	1.1	0.0	0.0	0.4	NR: ubiquitin-conjugation enzyme gene, <i>Glycine max</i>	AF532622.1	2e-18	68/74
78	135	7.0	21.0	84.6	11.8	31.1	NR: putative phosphatase mRNA, <i>Phaseolus vulgaris</i>	AJ518838.1	6e-7	63/72
79	119	-0.4	-0.3	-1.5	0.5	-0.4	EST: RTS_144_E01, <i>Phaseolus vulgaris</i> NR: hypothetical protein mRNA, <i>Zea mays</i>	CV544139.1 EU952715.1	3e-24 2e-32	92/100 152/198
80	114	1.1	1.8	8.2	9.5	5.1	No homology	-	-	-
81	49	21.0	60.4	96.7	103.1	70.3	No homology	-	-	-
82	101	29.9	13.0	65.2	82.8	47.7	No homology	-	-	-
83	95	-2.1	-0.2	-0.7	-7.9	-2.7	EST: POD_014_E08, <i>Phaseolus vulgaris</i> NR: no homology	CV539111.1 -	9e-36 -	91/95 -
84	99	-1.5	0.1	0.1	1.0	-0.1	NR: CYP1 mRNA, <i>Vigna radiata</i>	AB020612.1	e-31	84/88
85	88	-1.2	-0.2	-4.8	-4.5	-2.7	EST: Gm-c1026-2027, <i>Glycine max</i> NR: type II chlorophyll a/b binding protein mRNA, <i>Pisum sativum</i>	AW396080.1 X81962.1	7e-24 5e-62	69/71 224/277
86	91	-0.2	0.6	0.2	3.0	0.9	EST: BE-2994, <i>Phaseolus vulgaris</i> NR: cysteine proteinase mRNA, <i>Glycine max</i>	EX305098.1 AY383240.1	2e-9 8e-138	79/91 404/480
87	112	-1.4	-0.4	0.0	0.1	-0.4	EST: POD_012_G03, <i>Phaseolus vulgaris</i> NR: hypothetical protein mRNA, <i>Fragaria x ananassa</i>	CV538963.1 AY695666.1	8e-43 e-26	106/111 89/101
88	99	-1.5	-2.3	-3.3	-5.5	-3.2	EST: HDI_H_93.esd <i>Macrotyloma uniflorum</i> NR: Rieske iron-sulphur protein precursor mRNA, <i>Glycine max</i>	DR989271.1 AM498291.1	3e-17 4e-147	76/84 340/370
89	94	0.1	0.6	0.6	1.7	0.8	EST: BE-3285, <i>Phaseolus vulgaris</i> NR: DnaJ-like protein mRNA, <i>Lycopersicon esculentum</i>	EX305351.1 AF124139.1	8e-21 6e-30	69/74 131/166
91	96	0.3	-0.6	0.2	0.9	0.2	EST: PCS03549F, <i>Phaseolus coccineus</i> NR: acyl-CoA-binding protein mRNA, <i>Jatropha curcas</i>	CA900070.1 DQ452088.1	e-13 4e-66	81/93 213/259

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
92	96	0.1	1.9	0.7	1.7	1.1	EST: PCSC12178, <i>Phaseolus coccineus</i> NR: putative eukaryotic translation initiation factor SUI1 mRNA, <i>Arabidopsis thaliana</i>	CA905622.1 NM_180861.2	4e-13 2e-74	65/73 246/304	
93	87	-9.8	-4.4	-2.3	-24.8	-10.3	NR: pathogenesis-related protein PR4.2 mRNA, <i>Vigna unguiculata</i>	X98608.1	3e-7	68/77	
94	77	0.2	0.7	-0.3	0.8	0.3	EST: GLMAK40TF, <i>Glycine max</i> NR: unknown protein, <i>Arabidopsis thaliana</i>	EV269531.1 AY142583.1	e-9 3e-18	56/63 134/183	
95	83	-7.6	-4.0	-20.9	-65.1	-24.4	NR: ultraviolet-B-repressible protein mRNA, <i>Trifolium pratense</i>	AY340642.1	e-9	59/67	
97	94	1.0	2.0	2.2	15.6	5.2	No homology	-	-	-	
98	109	2.1	1.8	2.5	2.1	2.1	NR: ARG10 mRNA, <i>Vigna radiata</i>	AB012110.1	5e-16	85/98	
99	96	-19.0	-11.1	-7.3	-83.5	-30.2	EST: Gm-c1050-4247, <i>Glycine max</i> NR: putative caffeic acid methyl transferase mRNA, <i>Arachis hypogaea</i>	BI468602.1 AF479308.1	8e-24 5e-50	88/94 190/240	
100	97	-3.2	-1.7	-6.7	-15.9	-6.9	EST: Gm-c1062-7028, <i>Glycine max</i> NR: precursor for 33-kDa protein of photosystem II mRNA, <i>Pisum sativum</i>	CA801196.1 D13297.1	3e-20 3-150	62/65 437/529	
101	100	9.1	6.1	100.9	28.9	36.3	No homology	-	-	-	
102	91	47.2	12.6	85.7	82.8	57.1	EST: RTS_114_F06, <i>Phaseolus vulgaris</i> NR: CPRD14 protein mRNA, <i>Vigna unguiculata</i>	CV542032.1 D83971.1	4e-13 3e-144	71/81 388/452	
103	95	0.1	-0.2	0.3	0.1	0.1	No homology	-	-	-	
104	112	-1.7	-2.6	-4.9	-6.4	-3.9	No homology	-	-	-	
105	75	-13.8	-6.8	-64.8	-54.2	-34.9	EST: PV_GEa0012B_D06.b1, <i>Phaseolus vulgaris</i> NR: phosphoribulokinase mRNA, <i>Pisum sativum</i>	CV529964.1 Y11248.1	2e-8 2e-178	62/69 477/559	
106	92	39.9	3.7	43.8	145.0	58.1	NR: asparagine synthetase type-I mRNA, <i>Phaseolus vulgaris</i>	AJ133522.1	5e-6	57/65	
107	79	1.8	2.7	2.9	4.2	2.9	EST: Gm-c1051-5342, <i>Glycine max</i> NR: no homology	BG652352.1 -	e-15 -	57/61 -	
108	111	29.4	52.4	10.0	97.2	47.2	No homology	-	-	-	
109	95	-3.8	-3.9	-16.3	-28.9	-13.2	EST: POD_013_B05, <i>Phaseolus vulgaris</i> NR: putative desaturase-like protein mRNA, <i>Trifolium repens</i>	CV538991.1 AM282585.1	3e-29 0	88/94 557/658	
110	104	1.0	-2.2	-0.7	-0.1	-0.5	No homology	-	-	-	
111	48	26.8	10.1	54.0	56.5	36.8	No homology	-	-	-	
112	103	33.6	21.7	25.1	49.3	32.4	EST: Gc01_02e03, <i>Glycine clandestina</i> NR: low temperature and salt responsive protein, <i>Solanum tuberosum</i>	BG838378.1 AB061265.1	5e-16 e-27	58/62 142/187	
113	104	-5.1	-3.7	-1.6	-1.4	-2.9	EST: PvL175 <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Medicago truncatula</i>	EE253621.1 BT051404.1	3e-11 3e-16	53/58 113/151	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
114	136	4.6	10.7	52.0	18.9	21.5	EST : Gm-c1066-2825, <i>Glycine max</i> NR: putative chloroplast-targeted beta-amylase mRNA, <i>Brassica napus</i>	BM094466.1 AF319168.1	2e-16 3e-74	63/67 314/415	
115	107	2.0	4.4	2.8	2.3	2.9	EST : PCSC21603, <i>Phaseolus coccineus</i> NR: no homology	CA905718.1 -	5 ^e -10 -	45/48 -	
116	82	0.0	-0.8	-0.9	-2.6	-1.1	No homology	-	-	-	
117	77	-3.1	-1.2	-3.3	-20.7	-7.1	EST : PV_Gea0012D_H10.b1, <i>Phaseolus vulgaris</i> NR : putative RNA-binding protein mRNA, <i>Arabidopsis thaliana</i>	CV530181.1 AY070022.1	e-18 e-111	59/62 390/498	
118	112	69.3	36.5	21.9	38.8	41.6	No homology	-	-	-	
119	115	-0.3	-0.1	0.5	-0.5	-0.1	NR: tRNA-Cys (trnC) gene, <i>Panax assamicus</i>	AY275917.1	2e-28	82/87	
120	92	0.3	0.4	-0.3	0.1	0.1	EST : GLMC521TF, <i>Glycine max</i> NR: unknown mRNA, <i>Bassia scoparia</i>	EV274318.1 AY617078.1	3e-17 3e-55	64/69 201/252	
121	101	2.2	4.8	9.2	18.3	8.6	No homology	-	-	-	
122	105	0.1	0.4	6.4	10.3	4.3	No homology	-	-	-	
123	103	13.9	23.4	10.9	14.5	15.7	No homology	-	-	-	
124	100	0.7	1.3	1.9	n.d.	1.0	No homology	-	-	-	
125	96	-14.7	-2.8	-8.3	-32.5	-14.6	NR: acid phosphatase gene, <i>Phaseolus vulgaris</i>	AB116720.1	6e-6	68/81	
126	88	27.3	54.7	49.8	49.3	45.3	EST : Gm-c1071-4330, <i>Glycine max</i> NR: no homology	CA935204.1 -	2 ^e -18 -	68/74 -	
127	97	-4.7	-2.0	-5.9	-6.7	-4.8	No homology	-	-	-	
128	110	0.0	-0.5	-1.0	-0.7	-0.6	EST : Gm-c1068-6210, <i>Glycine max</i> NR: putative plastid-lipid associated protein mRNA, <i>Arabidopsis thaliana</i>	BM886786.1 NM_118350.2	2e-6 2e-26	42/46 153/205	
129	96	-1.9	0.5	0.2	0.6	-0.1	NR: ribosomal protein S28 mRNA, <i>Prunus persica</i>	AJ012653.1	2e-8	69/81	
130	93	0.3	0.1	0.1	2.1	0.7	EST: SSH-37, <i>Phaseolus vulgaris</i> NR: glutathione peroxidase mRNA, <i>Phaseolus vulgaris</i>	EG594330.1 DQ455600.1	e-7 5e-76	50/56 177/187	
131	89	0.2	0.8	2.8	2.5	1.6	NR: cysteine protease mRNA, <i>Vigna mungo</i>	AB038598.1	5 ^e -37	87/89	
132	107	-0.5	-0.2	-0.1	-0.3	-0.3	EST : PCSC21328, <i>Phaseolus coccineus</i> NR: CHAPERONIN 20; calmodulin binding mRNA, <i>Arabidopsis thaliana</i>	CA905919.1 NM_180714.2	8e-15 e-51	84/94 240/320	
133	98	2.5	9.2	39.8	28.9	20.1	NR: hypothetical protein mRNA, <i>Trifolium pratense</i>	AB236782.1	2e-8	57/65	
134	99	-4.5	-2.1	-1.8	-1.4	-2.4	EST: BE-3380, <i>Phaseolus vulgaris</i> NR: stable protein 1-related mRNA, <i>Arabidopsis thaliana</i>	EX305429.1 NM_112598.3	2e-24 e-51	87/96 237/317	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
135	60	-0.5	1.3	-2.6	-0.3	-0.5	EST: gmrtDrNS01_26-C_M13R_D12_090.s4, <i>Glycine max</i> NR: putative chloroplast 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase mRNA, <i>Populus trichocarpa</i>	CX707633.1 EU693025.1	4e-9 0	37/38 627/767	
136	120	-1.9	0.2	-1.1	-0.5	-0.8	EST: PCS04220, <i>Phaseolus coccineus</i> NR: ribosomal protein 16 gene, <i>Phaseolus vulgaris</i>	CA903667.1 AY077897.1	e-23 e-158	65/67 321/324	
137	69	2.8	6.5	34.9	4.3	12.1	No homology	-	-	-	
138	102	-1.0	-4.9	-29.4	-5.8	-10.3	EST: PCSC16633, <i>Phaseolus coccineus</i> NR: PSI reaction center subunit III mRNA, <i>Citrus sinensis</i>	CA901756.1 AB002095.1	2e-27 e-78	80/85 266/331	
139	98	-1.0	-0.1	0.6	0.4	0.0	EST: PCSC12265, <i>Phaseolus coccineus</i> NR: acireductone dioxygenase [iron(II)-requiring]/metal ion binding mRNA, <i>Arabidopsis thaliana</i>	CA910184.1 NM_117554.3	3e-23 2e-63	79/86 212/258	
140	107	0.0	-5.9	-40.2	-12.0	-14.5	EST: Gm-c1071-4520, <i>Glycine max</i> NR: photosystem I reaction center V mRNA, <i>Camellia sinensis</i>	CA935234.1 AY724779.1	e-22 5e-70	76/81 243/304	
141	93	-31.0	-3.5	-26.4	-44.5	-26.3	No homology	-	-	-	
142	84	-5.4	-3.2	-11.5	-1.1	-5.3	EST: PCS05630, <i>Phaseolus coccineus</i> NR: thioredoxin h2 mRNA, <i>Medicago truncatula</i>	CA901892.1 DQ121443.1	3e-18 9e-87	62/66 260/312	
143	95	-4.5	-1.8	-10.2	-23.9	-10.1	NR: CDPK-related protein kinase mRNA, <i>Vigna radiata</i>	AY551333.1	7e-27	79/84	
144	119	-8.9	-1.5	-6.6	-4.4	-5.3	EST: LVS_008_E04, <i>Phaseolus vulgaris</i> NR: serine hydroxymethyltransferase mRNA, <i>Pisum sativum</i>	CV531857.1 M87649.1	7e-28 3e-60	104/119 299/398	
145	116	-2.5	-0.1	-3.9	-2.1	-2.2	No homology	-	-	-	
146	126	0.9	0.1	0.1	-0.1	0.2	EST: PV_GEa0013B_F01.b1, <i>Phaseolus vulgaris</i> NR: no homology	CV530315.1 -	4e-17 -	75/84 -	
147	106	-6.9	-3.1	-34.2	-4.4	-12.2	No homology	-	-	-	
148	55	-14.0	-3.9	-29.3	-59.6	-26.7	NR: P42-1 putative NADH-dependent hydroxypyruvate reductase mRNA, <i>Glycine max</i>	AF503360.1	3e-6	48/53	
149	58	-6.9	0.0	-1.3	-1.6	-2.4	EST: Gm-c1080-5853, <i>Glycine max</i> NR: alanine aminotransferase 2 mRNA, <i>Glycine max</i>	BU763092.1 EU165372.1	e-4 0	39/42 453/473	
150	40	-6.2	-5.7	-12.3	-10.2	-8.6	NR: photosystem I chlorophyll a/b binding protein mRNA, <i>Pisum sativum</i>	EF208907.1	4e-7	33/34	
151	117	0.7	1.2	0.1	1.9	1.0	EST: MPMGp1174B0763Q, <i>Paracentrotus lividus</i> NR: orthodonticle-related protein mRNA, <i>Paracentrotus lividus</i>	AM597042.1 NM_214588.3	5e-7 e-124	31/31 428/544	
152	100	33.6	13.2	87.7	6.7	35.3	EST: gmrtDrNS01_11-B_M13R_G05_035.s4, <i>Glycine max</i> NR: myo-inositol oxygenase 5 mRNA, <i>Arabidopsis thaliana</i>	CX702986.1 NM_125047.2	3e-5 5e-92	52/60 401/536	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
153	100	7.4	19.5	80.5	20.5	32.0	No homology	-	-	-	
154	60	0.3	0.4	0.6	27.7	7.3	No homology	-	-	-	
155	88	-2.6	-0.7	-2.1	-4.4	-2.4	NR: photosystem I psaH protein mRNA, <i>Arachis hypogaea</i>	DQ887080.1	e-9	75/88	
156	72	0.2	3.2	0.0	1.1	1.1	EST: LVS_040_A02, <i>Phaseolus vulgaris</i> NR: no homology	CV533499.1	7e-11	58/65	
157	90	0.0	0.9	2.3	0.6	0.9	NR: isoprenylated protein mRNA, <i>Glycine max</i>	U13179.1	2e-17	72/80	
158	99	0.1	2.4	14.3	13.4	7.5	NR: heat shock protein 22 mRNA, <i>Glycine max</i>	X07188.1	4e-13	84/97	
159	114	0.3	0.9	0.4	1.0	0.6	EST: Gm-c1013-2397, <i>Glycine max</i> NR: catalytic subunit of protein phosphatase 1 mRNA, <i>Vicia faba</i>	AW132339.1 AB254851.1	e-4 0	61/72 545/623	
160	109	1.1	2.2	3.4	9.5	4.0	No homology	-	-	-	
161	95	0.2	1.3	0.0	0.2	0.4	NR: ribosomal protein S27 mRNA, <i>Arabidopsis thaliana</i>	NM_124167.3	4e-19	78/87	
162	94	-3.0	-0.7	-20.1	-35.8	-14.9	EST: Gm-c1069-7733, <i>Glycine max</i> NR: chloroplast phosphoglycerate kinase mRNA, <i>Populus nigra</i>	CB063640.1 AB018412.1	2e-24 2e-74	75/80 225/269	
163	101	11.6	20.6	10.7	34.4	19.3	No homology	-	-	-	
164	96	0.5	5.1	0.6	1.0	1.8	EST: PV_GEa0014B_B03.b1, <i>Phaseolus vulgaris</i> NR: no homology	CV530563.1	2e-24	82/88	
165	97	-1.6	-0.6	0.5	-0.5	-0.6	EST: RTS_146_H10, <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Medicago truncatula</i>	CV544302.1 BT051437.1	2e-39 7e-87	91/93 341/440	
166	93	10.0	7.8	10.5	42.1	17.6	EST: PCSC20002, <i>Phaseolus coccineus</i> NR: putative invertase inhibitor mRNA, <i>Cicer arietinum</i>	CA900450.1 AJ487472.1	2e-12 3 ^e -50	66/74 178/220	
167	88	-2.8	-7.1	-6.6	-6.8	-5.8	EST: Gm-c1057-4385, <i>Glycine soja</i> NR: UDP-glucose glucosyltransferase mRNA, <i>Catharanthus roseus</i>	BM523855.1 AB159213.1	3e-17 4e-15	66/72 80/100	
168	84	2.6	1.7	5.5	-0.5	2.3	EST: GLL061_B06_023, <i>Cyamopsis tetragonoloba</i> NR: no homology	EG979126.1	5e-6	32/33	
169	111	3.7	0.3	-0.8	0.1	0.8	No homology	-	-	-	
171	78	-4.5	-5.5	-34.3	-18.5	-15.7	No homology	-	-	-	
172	62	-4.0	-15.7	-70.4	-33.7	-30.9	NR: extensin gene, <i>Glycine max</i>	U44838.1	3e-6	32/33	
173	97	22.1	7.5	17.9	42.1	22.4	EST: Gm-c1068-2697, <i>Glycine max</i> NR: no homology	BI893889.1	2e-9	74/87	
174	102	-0.9	-0.3	-9.1	-13.9	-6.1	EST: PCEP03699, <i>Phaseolus coccineus</i> NR: no homology	CA899410.1	4e-29	92/100	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
175	91	-2.9	0.2	-0.6	-0.6	-1.0	EST: HDI_H_55.esd, <i>Macrotyloma uniflorum</i> NR: 4-coumaryl-CoA ligase mRNA, <i>Plantago major</i> NR: serine/threonine protein kinase mRNA, <i>Fagus sylvatica</i> NR: monodehydroascorbate reductase mRNA, <i>Plantago major</i>	DR989250.1 AM159093.1 AJ606472.1 AM158910.1	7e-18 2e-24 7e-24 2e-23	58/61 76/79 83/92 74/77	
177	108	-1.3	0.3	0.0	-1.1	-0.5	EST: AB020045, <i>Phaseolus vulgaris</i> NR: no homology	AB020045.1 -	4e-20 -	65/69 -	
178	104	4.4	1.4	22.4	14.5	10.7	No homology	-	-	-	
179	94	-1.6	0.0	-0.7	0.1	-0.5	NR: elongation factor EF-1a gene, <i>Glycine max</i>	X56856.1	5e-40	92/94	
180	92	-1.9	-1.4	-1.4	-3.5	-2.0	EST: PCSC08999, <i>Phaseolus coccineus</i> NR: no homology	CA915098.1 -	2e-9 -	59/66 -	
181	92	-0.4	0.0	0.7	-0.4	0.0	NR: cysteine proteinase inhibitor mRNA, <i>Vigna unguiculata</i>	Z21954.1	2e-27	88/95	
182	92	36.8	17.9	42.8	25.3	30.7	NR: putative cytidine or deoxycytidylate deaminase mRNA, <i>Cicer arietinum</i>	AJ006764.1	e-22	81/89	
184	101	7.4	-0.6	-0.7	-2.2	1.0	EST: POD_029_G03, <i>Phaseolus vulgaris</i> NR: no homology	CV540299.1 -	7e-6 -	29/29 -	
185	106	-20.9	-23.1	-37.1	-31.5	-28.2	NR: 1-aminocyclopropane-1-carboxylic acid oxidase mRNA, <i>Phaseolus vulgaris</i>	AF053354.1	e-13	59/64	
186	92	-1.9	-0.8	-1.1	-3.3	-1.8	EST: POD_029_F04, <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	CV540289.1 EF146791.1	3e-29 4e-46	87/92 201/262	
187	73	-16.0	-15.7	-28.3	-58.6	-29.7	EST: LVS_047_F11, <i>Phaseolus vulgaris</i> NR: H-protein of glycine decarboxylase mRNA, <i>Pisum sativum</i>	CV534122.1 J05164.1	2e-17 e3-144	60/64 417/503	
188	96	-4.7	-1.2	-27.4	-31.5	-16.2	NR: plastid-located glutamine synthetase gene, <i>Phaseolus vulgaris</i>	X61292.1	3e-14	84/98	
189	96	1.3	0.1	-4.2	-4.1	-1.7	EST: gmrhRww24-04-SP6_E10_1_072, <i>Glycine max</i> NR: tRNA-Val gene, <i>Cyanea pilosa</i>	CF923150.1 DQ285213.1	8e-27 2e-30	82/88 93/103	
190	94	1.3	-2.3	-0.2	-0.3	-0.4	EST: GLLBI59TF, <i>Glycine max</i> NR: no homology	EV267407.1 -	2e-5 -	37/40 -	
191	94	-1.4	-0.1	-0.7	8.4	1.5	EST: Gm-c1056-5710, <i>Glycine soja</i> NR: no homology	CA783981.1 -	e-16 -	77/87 -	
192	107	-1.6	-0.1	-1.8	-0.4	-1.0	NR: ribosomal protein L21mRNA, <i>Arabidopsis thaliana</i>	NM_100831.3	3e-5	37/40	
193	90	9.0	2.3	6.3	1.8	4.8	No homology	-	-	-	
194	88	-1.2	0.0	-1.1	-1.5	-0.9	EST: Gm-c1069-7293, <i>Glycine max</i> NR: histone H1-like protein mRNA, <i>Camellia sinensis</i>	CA953449.1 EU716314.1	7e-27 3e-60	79/84 354/499	
195	114	45.1	23.6	46.9	n.d.	28.9	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
196	85	0.7	1.3	0.7	1.4	1.0	EST: PCSC09493, <i>Phaseolus coccineus</i> NR: no homology	CA912998.1	4e-19	66/71	
197	114	-0.5	0.3	-0.1	-0.2	-0.1	NR: arg2 mRNA, <i>Vigna radiata</i>	D14411.1	4e-23	86/94	
198	95	-0.2	-0.1	1.3	-0.1	0.2	NR: ubiquitin-conjugating enzyme 2 mRNA, <i>Arachis hypogaea</i>	DQ097727.1	4e-10	73/86	
199	114	7.7	13.8	35.7	43.3	25.1	No homology	-	-	-	
200	101	-0.4	0.4	0.9	-0.3	0.1	NR: ubiquitin extension protein mRNA, <i>Gossypium hirsutum</i>	DQ116441.1	2e-9	62/71	
201	93	0.0	-0.5	0.0	0.2	-0.1	NR ubiquitin-like protein SMT3 mRNA, <i>Phaseolus vulgaris</i>	AF451278.1	6e-9	34/34	
202	87	-0.5	-0.2	-0.1	0.4	-0.1	NR: ribosomal protein L30 mRNA, <i>Zea mays</i>	AF034949.1	5e-12	45/47	
203	78	-1.0	0.0	0.0	0.7	-0.1	EST: GLNAY94TF, <i>Glycine max</i> NR: heat- and acid-stable phosphoprotein mRNA, <i>Zea mays</i>	EV280924.1	2e-11	50/54	
204	91	48.3	1.0	45.9	2.9	24.5	No homology	-	-	-	
205	105	-3.1	-13.1	-13.2	-32.6	-15.5	EST: LVS_023_B10, <i>Phaseolus vulgaris</i> NR: no homology	CV532295.1	2e-6	78/94	
206	93	-1.0	-0.4	0.2	-0.2	-0.3	NR: ribosomal protein L33 mRNA, <i>Castanea sativa</i>	AF334840.1	e-4	71/86	
207	52	0.3	-1.4	-0.3	-0.6	-0.5	EST: POD_024_H09, <i>Phaseolus vulgaris</i> NR: Mg chelatase subunit mRNA, <i>Glycine max</i>	CV539907.1	e-4	29/30	
208	111	45.1	4.0	13.8	8.6	17.9	No homology	-	-	-	
209	115	1.2	0.5	1.0	2.3	1.3	EST: PCSC09156, <i>Phaseolus coccineus</i> NR: HSP associated protein like mRNA, <i>Arabidopsis thaliana</i>	CA906266.1	1e-32	99/106	
210	111	-2.4	0.3	-0.7	-0.4	-0.8	NR: HMG-1 like protein mRNA, <i>Glycine max</i>	AY059803.1	5e-45	182/231	
211	105	-2.7	-0.2	-6.5	-9.0	-4.6	EST: LVS_045_A06, <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	X58245.1	9e-24	95/106	
212	98	0.4	-0.5	0.5	-1.6	-0.3	EST: Gm-c1071-3566, <i>Glycine max</i> NR: DNA-directed RNA polymerases I, II, and III 7.3 kDa polypeptide mRNA, <i>Zea mays</i>	CV533912.1	e-35	88/91	
213	87	-0.8	-0.5	1.0	0.4	0.0	NR: putative eukaryotic translation initiation factor 5a-2 mRNA, <i>Oryza sativa</i>	EF144236.1	5e-18	85/105	
214	71	0.2	-1.6	-4.6	-11.7	-4.4	EST: UCRVU02_D05_T3, <i>Vigna unguiculata</i> NR: cationic peroxidase 2 mRNA, <i>Glycine max</i>	CA802795.1	e-41	95/97	
216	74	-9.0	-1.5	-21.6	n.d.	-8.0	EST: PEG002-C-109480-501, <i>Aplysia californica</i> NR: aspartic acid-rich protein aspolin2 mRNA, <i>Theragra chalcogramma</i>	EU970473.1	5e-32	130/162	
217	100	-0.3	0.0	-0.1	-2.8	-0.8	NR: cyclin-dependent kinases regulatory subunit mRNA, <i>Glycine max</i>	EB226998.1	5e-6	29/29	
218	101	2.6	2.1	6.1	3.4	3.6	No homology	AB117518.1	e-41	158/197	
								AY439094.1	4e-7	49/55	
								-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
219	95	-3.1	0.2	-4.1	-1.4	-2.1	No homology	-	-	-	
220	133	-2.0	-3.0	-14.7	-59.6	-19.8	EST: PV_GEa0011D_A09.b1, <i>Phaseolus vulgaris</i> NR: unknown mRNA	CV529784.1 AF098661.1	e-5 7e-31	50/57 122/148	
221	104	-3.8	-4.1	-9.3	-10.8	-7.0	EST: BE-1361, <i>Phaseolus vulgaris</i> NR: GA-like protein mRNA, <i>Cicer arietinum</i>	EX303743.1 EF375953.1	2e-15 7e-75	60/65 318/421	
222	103	3.0	13.2	20.4	13.4	12.5	No homology	-	-	-	
223	90	-12.4	-31.5	-16.6	-36.9	-24.4	EST: PCSC20280, <i>Phaseolus coccineus</i> NR: ferredoxin I gene, <i>Pisum sativum</i>	CA901455.1 M31713.1	4e-10 e-45	52/58 168/209	
224	99	26.3	21.7	35.9	21.7	26.4	NR: beta-amylase 1 mRNA, <i>Glycine max</i>	AJ871580.1	e-7	67/76	
225	106	0.0	0.0	-1.2	-0.1	-0.3	NR: S-adenosylmethionine synthetase mRNA, <i>Phaseolus lunatus</i>	AB062358.1	6e-34	97/104	
226	131	-0.6	0.2	-0.8	-3.2	-1.1	NR: early light-induced protein mRNA, <i>Glycine max</i>	U82810.1	2e-22	105/121	
227	91	14.2	3.2	12.8	33.3	15.8	EST: ABWZ1092, <i>Glycine max</i> NR: transcription factor MYB139, <i>Glycine max</i>	EH220099.1 DQ822919.1	3e-14 5e-13	71/79 262/400	
228	101	1.4	2.4	1.2	38.5	10.9	NR: glutathione S-transferase GST 10 mRNA, <i>Glycine max</i>	AF243365.1	2e-5	43/48	
229	90	-5.0	-0.8	-3.7	-23.9	-8.3	NR: H+-transporting ATP synthase mRNA, <i>Hyacinthus orientalis</i>	AY389668.1	e-6	47/50	
230	81	0.4	0.5	-1.6	-4.4	-1.3	EST: BE-1775, <i>Phaseolus vulgaris</i> NR: photosystem II core complex protein psbY mRNA, <i>Zea mays</i>	EX304470.1 EU955191.1	2e-20 7e-25	74/81 235/345	
231	107	0.3	-0.8	0.9	0.9	0.3	EST: PCSC19514, <i>Phaseolus coccineus</i> NR: vacuolar membrane ATPase subunit G mRNA, <i>Citrus limon</i>	CB280531.1 AF184068.1	4e-38 2e-69	102/107 261/335	
232	86	-9.0	-4.9	-37.1	-7.4	-14.6	No homology	-	-	-	
233	51	-0.5	0.1	-2.1	-3.5	-1.5	EST: BE-3464, <i>Phaseolus vulgaris</i> NR: CBS1 mRNA, <i>Hyacinthus orientalis</i>	EX305501.1 AY389652.1	7e-13 8e-23	46/48 111/142	
234	88	24.2	12.6	42.9	20.0	24.9	EST: Gm-c1068-1529, <i>Glycine max</i> NR: TSJT1-like protein mRNA, <i>Solanum tuberosum</i>	BF424584.1 DQ191656.1	2e-36 5e-38	86/88 217/300	
235	113	-7.6	-2.9	-6.2	-11.7	-7.1	No homology	-	-	-	
236	111	0.1	1.0	0.9	1.2	0.8	EST: UCRVU02_G05_T7, <i>Vigna unguiculata</i> NR: no homology	DR068416.1 -	6e-28 -	99/110 -	
237	102	-0.9	-0.3	-0.8	-4.4	-1.6	EST: PCS02877, <i>Phaseolus coccineus</i> NR: ribosomal protein SocL12, <i>Spinacia oleracea</i>	CA903663.1 X13153.1	8e-12 2e-75	42/43 305/401	
238	54	38.8	24.0	4.1	8.9	19.0	EST: GmUV-B_220, <i>Glycine max</i> NR: vacuolar processing enzyme 2 mRNA, <i>Glycine max</i>	EV555079.1 AY062213.1	e-5 3e-35	40/44 109/120	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
239	90	0.4	0.6	3.4	0.8	1.3	EST: UCRVU02_MA_02_T3, <i>Vigna unguiculata</i> NR: alcohol dehydrogenase 1 mRNA, <i>Glycine max</i>	EG594280.1 AF532629.1	8e-33 0	86/90 731/838	
240	86	-4.9	-1.9	-2.9	-29.8	-9.9	No homology	-	-	-	
241	102	-0.4	-0.2	1.7	0.6	0.4	No homology	-	-	-	
242	87	-1.9	-2.0	-9.5	-17.4	-7.7	EST: GmUV-B_118, <i>Glycine max</i> NR: glycine cleavage complex P protein, <i>Pisum sativum</i>	EV554998.1 X59773.1	4e-22 0	62/64 678/784	
243	83	1.3	1.2	3.3	15.6	5.4	NR: Snakin-like cysteine rich protein mRNA, <i>Phaseolus vulgaris</i>	AM158277.1	5e-15	59/64	
244	91	0.5	1.8	0.5	6.4	2.3	No homology	-	-	-	
245	82	-1.1	-1.4	-3.6	n.d.	-1.5	No homology	-	-	-	
246	103	-0.2	-0.1	0.6	-0.5	-0.1	EST: RTS_126_E08, <i>Phaseolus vulgaris</i> NR: no homology	CV542845.1 -	2e-6 -	39/42 -	
247	92	0.3	-0.3	0.9	0.7	0.4	EST: NOD_233_B07, <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	CV536723.1 EF147532.1	2e-5 5e-35	34/36 206/285	
248	90	0.4	1.0	0.4	0.4	0.5	NR: glycine-rich RNA-binding protein RGP-1c mRNA, <i>Nicotiana sylvestris</i>	AY485991.1	6e-12	54/59	
249	99	21.0	8.1	24.5	32.5	21.5	No homology	-	-	-	
250	92	5.3	0.4	5.8	12.8	6.1	EST: PVEPSE2025H12, <i>Phaseolus vulgaris</i> NR: putative lipid transfer protein mRNA, <i>Cicer arietinum</i>	CB540903.1 AJ630658.1	6e-6 6e-38	57/65 145/180	
251	109	1.1	2.2	0.0	0.5	1.0	EST: PCS05756, <i>Phaseolus coccineus</i> NR: no homology	CA916079.1 -	2e-28 -	95/104 -	
252	98	2.0	0.7	3.6	1.5	2.0	EST: WS02042_B19, <i>Populus trichocarpa</i> x <i>Populus nigra</i> NR: glutaredoxin family protein mRNA, <i>Arabidopsis thaliana</i>	DT524263.1 NM_128606.2	4e-7 e-42	40/43 206/276	
253	114	-0.7	-0.4	-1.1	-7.1	-2.3	EST: POD_011_E06, <i>Phaseolus vulgaris</i> NR: ultraviolet-B-repressible protein mRNA, <i>Gossypium hirsutum</i>	CV538873.1 AY551823.1	7e-28 e-20	103/113 147/204	
254	91	1.1	1.7	5.6	6.2	3.7	EST: RTS_123_A05, <i>Phaseolus vulgaris</i> NR: 19-3 aldo/keto reductase mRNA, <i>Malus x domestica</i>	CV542630.1 AY347812.1	8e-33 6e-39	86/90 221/304	
255	69	-3.1	-2.5	-5.9	-28.2	-9.9	No homology	-	-	-	
256	101	0.6	-0.2	0.4	0.0	0.2	NR: heme oxygenase 1 mRNA, <i>Glycine max</i>	AF320024.1	8e-24	89/98	
257	85	-0.4	-0.3	-1.1	3.4	0.4	EST: PCSC17187, <i>Phaseolus coccineus</i> NR: ribosomal protein S14 mRNA, <i>Lupinus luteus</i>	CA902863.1 AF026079.1	9e-8 9e-86	44/48 242/281	
258	54	-2.1	-0.8	-1.0	-0.1	-1.0	No homology	-	-	-	
259	89	20.0	25.5	16.3	26.5	22.1	No homology	-	-	-	
260	94	-2.6	-2.3	-9.8	-10.8	-6.4	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
261	98	4.2	3.7	5.8	28.9	10.7	NR: ferritin mRNA, <i>Phaseolus vulgaris</i>	X58274.1	3e-20	86/97	
262	88	-0.3	0.7	-1.2	-12.6	-3.3	EST: PVEPSE3011D12, <i>Phaseolus vulgaris</i> NR: no homology	CB555983.1 -	3e-20 -	71/77 -	
263	109	1.1	2.1	n.d.	18.9	5.5	No homology	-	-	-	
264	97	-8.5	-0.8	-5.2	-4.4	-4.8	NR: precursor of photosystem II 22 kDa protein mRNA, <i>Nicotiana tabacum</i>	X84225.1	4e-10	78/92	
265	102	1.5	5.8	25.9	2.3	8.9	EST: RTS_113_E02, <i>Phaseolus vulgaris</i> NR: no homology	CV541965.1 -	e-7 -	38/40 -	
266	90	-0.3	-1.5	-0.8	-0.1	-0.7	EST: PCEP00562, <i>Phaseolus coccineus</i> NR: no homology	CA899331.1 -	2e-30 -	85/90 -	
267	97	-0.3	1.7	1.5	4.2	1.8	EST: RTS_102_E12, <i>Phaseolus vulgaris</i> NR: putative hydrophobic protein mRNA, <i>Arabidopsis thaliana</i>	CV541250.1 NM_119212.4	8e-18 e-14	64/69 136/194	
268	105	-1.3	-0.4	1.0	1.2	0.2	EST: PCSC20738, <i>Phaseolus coccineus</i> NR: ubiquinol-cytochrome c reductase complex 14 kDa protein mRNA, <i>Zea mays</i>	CA901941.1 EU973896.1	9e-24 9e-28	74/79 117/146	
269	94	8.8	18.8	10.2	25.3	15.8	EST: BE-2204, <i>Phaseolus vulgaris</i> NR: nodule-enhanced protein phosphatase type 2C mRNA, <i>Lotus japonicus</i>	EX304066.1 AF092431.1	4e-13 0	66/71 505/600	
270	90	-1.0	-0.6	-2.4	1.9	-0.5	EST: GLMB442TF, <i>Glycine max</i> NR: elongation factor EF-2, <i>Pisum sativum</i>	EV271196.1 AB082376.1	4e-7 5e-64	34/35 187/218	
271	101	-0.3	-1.0	-3.4	-2.0	-1.7	EST: PVEPSE2003H07, <i>Phaseolus vulgaris</i> NR: no homology	CB539331.1 -	4e-29 -	92/100 -	
272	94	-0.5	0.4	0.4	0.5	0.2	EST: NOD_238_F05, <i>Phaseolus vulgaris</i> NR: lipoic acid synthase-like protein mRNA, <i>Brassica rapa</i>	CV537162.1 EU186340.1	e-28 2e-21	79/83 143/195	
273	81	12.6	15.1	23.5	33.7	21.2	No homology	-	-	-	
274	85	-1.7	-2.1	-2.3	-0.5	-1.7	No homology	-	-	-	
275	93	3.4	-0.1	0.7	0.8	1.2	EST: POD_011_C01, <i>Phaseolus vulgaris</i> NR: ras-related GTP-binding protein mRNA, <i>Pisum sativum</i>	CV538848.1 X65650.1	3e-14 e-23	71/79 170/233	
276	103	11.6	5.8	27.9	40.9	21.5	EST: RTS_125_B03, <i>Phaseolus vulgaris</i> NR: no homology	CV542772.1 -	5e-22 -	80/88 -	
277	102	-0.3	0.3	-1.4	-1.0	-0.6	No homology	-	-	-	
278	78	-1.2	-2.5	-6.5	-13.5	-5.9	EST: Gm_ck37440, <i>Glycine max</i> NR: no homology	CD409952.1 -	2e-5 -	52/60 -	
279	88	0.1	0.3	0.7	0.7	0.5	NR: profilin mRNA, <i>Phaseolus vulgaris</i>	X81982.1	2e-24	66/68	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
280	102	6.9	20.6	7.7	31.3	16.6	No homology	-	-	-	
281	102	0.6	-0.8	3.3	0.1	0.8	EST: BE-252, <i>Phaseolus vulgaris</i> NR: SKP1-like protein 1 mRNA, <i>Malus x domestica</i>	EG563016.1 EU586193.1	7e-46 2e-88	99/100 253/298	
282	95	-2.5	-0.6	-2.9	-3.8	-2.5	NR: enzymatic resistance protein mRNA, <i>Glycine max</i>	DQ167250.1	6e-43	94/95	
283	107	-1.1	0.1	6.6	0.5	1.5	NR: cytosolic ascorbate peroxidase 2 mRNA, <i>Glycine max</i>	AB082932.1	e-22	95/105	
284	106	-1.7	-0.4	-0.8	0.5	-0.6	EST: BE-1039, <i>Phaseolus vulgaris</i> NR: ribosomal protein L34 mRNA, <i>Nicotiana tabacum</i>	EG588019.1 L27107.1	2e-37 e-98	100/106 305/372	
285	106	-17.0	-7.8	-17.6	-22.8	-16.3	EST: PCEP00684, <i>Phaseolus coccineus</i> NR: hypothetical protein, <i>Gossypium hirsutum</i>	CA896844.1 X54092.1	5e-7 e-34	52/59 238/333	
286	110	-0.1	1.7	2.9	6.2	2.7	No homology	-	-	-	
287	92	0.2	-1.0	-0.4	-0.8	-0.5	NR: cytochrome b5 isoform mRNA, <i>Vernicia fordii</i>	AY578728.1	e-16	65/71	
288	109	21.0	10.3	27.9	18.9	19.5	EST: HDI_J_73.esd, <i>Macrotyloma uniflorum</i> NR: rhabdomic opsin mRNA, <i>Platynereis dumerilii</i> NR: insulin-like growth factor binding protein, <i>Oncorhynchus mykiss</i> NR: 4-coumaryl-CoA ligase mRNA, <i>Plantago major</i>	DR989445.1 AJ316544.1 DQ146965.2 AM159093.1	5e-10 2e-22 2e-22 2e-22	57/64 69/72 73/77 72/75	
289	66	1.8	3.1	1.9	2.7	2.4	EST: Gm-c1055-2925, <i>Glycine max</i> NR: no homology	BG650004.1 -	6e-5 -	54/63 -	
290	98	-0.9	-0.8	-0.6	3.1	0.2	NR: ribosomal protein L36a/L44 mRNA, <i>Arabidopsis thaliana</i>	NM_117509.2	e-19	64/68	
291	86	0.3	-1.5	-0.5	0.4	-0.3	No homology	-	-	-	
292	98	0.4	1.0	-0.2	0.1	0.3	EST: BE-206,1 <i>Phaseolus vulgaris</i> NR: ubiquitin conjugating enzyme 2 mRNA, <i>Malus x domestica</i>	EX303949.1 EU586200.1	3e-14 6e-141	78/87 383/450	
294	87	-14.3	-2.3	-10.8	-12.0	-9.8	NR: photosystem I-N subunit mRNA, <i>Phaseolus vulgaris</i>	AF492824.1	e-12	55/60	
295	84	0.4	0.0	0.6	1.0	0.5	EST: BE-2929, <i>Phaseolus vulgaris</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	EX305036.1 EF145271.1	4e-16 5e-84	58/62 342/451	
296	96	0.8	26.3	4.6	18.2	12.5	No homology	-	-	-	
297	107	-13.6	-5.2	-0.3	-5.5	-6.1	No homology	-	-	-	
298	90	-1.9	0.0	0.1	0.7	-0.3	NR: ribosomal protein S20 mRNA, <i>Arabidopsis thaliana</i>	NM_114372.4	2e-8	63/73	
299	106	-0.3	0.9	-0.3	0.1	0.1	EST: ACBU6518, <i>Glycine max</i> NR: no homology	EH262715.1	e-22	95/104	
300	94	3.7	1.8	3.6	1.8	2.7	EST: PVEPSE3005B06, <i>Phaseolus vulgaris</i> NR: polyubiquitin 2 mRNA, <i>Phaseolus vulgaris</i>	CB541873.1 AF527442.1	2e-27 3e-70	86/93 169/181	
301	90	-0.6	-0.3	-1.0	-3.1	-1.2	NR: RNP1 chloroplast RNA binding protein, <i>Phaseolus vulgaris</i>	X82030.1	6e-18	82/91	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
302	116	-0.4	-1.2	-0.1	-0.1	-0.4	No homology	-	-	-	
303	96	-1.5	1.0	0.9	0.4	0.2	EST: POD_004_E06, <i>Phaseolus vulgaris</i> NR: immunophilin mRNA, <i>Vicia faba</i>	CV538352.1 U96925.1	6e-6 e-111	41/45 314/370	
304	88	-2.8	-0.9	-2.3	-0.7	-1.7	NR: ribosomal protein S21, <i>Zea mays</i>	X98656.1	e-4	26/26	
305	43	-1.5	-0.9	0.3	-0.5	-0.6	No homology	-	-	-	
306	90	4.5	-0.5	1.0	6.2	2.8	NR: chloroplast-localized ISCA-like protein mRNA, <i>Arabidopsis thaliana</i>	NM_100924.3	9e-14	66/74	
307	93	-1.4	-0.8	1.6	-0.8	-0.3	EST: Gm-c1036-12572, <i>Glycine max</i> NR: cytochrome-c oxidase mRNA, <i>Brassica rapa</i>	BU964870.1 EF110926.1	e-31 e-34	87/92 143/181	
308	96	7.9	4.4	7.2	12.3	7.9	No homology	-	-	-	
309	105	-11.0	3.5	-17.6	-25.0	-12.5	EST: Gm-c1049-344, <i>Glycine max</i> NR: fructose-bisphosphate aldolase mRNA, <i>Glycine max</i>	BE611237.1 AY492006.1	7e-6 6e-55	72/85 123/123	
310	59	-3.3	-1.5	-14.7	-1.1	-5.2	No homology	-	-	-	
311	118	-6.6	-16.6	-18.6	-20.7	-15.6	EST: BE-2353, <i>Phaseolus vulgaris</i> NR: putative PSII-P protein mRNA, <i>Trifolium pratense</i>	EX304193.1 AB236819.1	8e-37 7e-60	112/122 174/199	
312	87	-6.6	-0.2	-1.1	-12.6	-5.1	EST: LVS_035_C02, <i>Phaseolus vulgaris</i> NR: chloroplast small heat shock protein gene, <i>Epilobium amurense</i>	CV533104.1 EF467641.1	6e-6 e-41	38/41 288/410	
313	91	0.1	1.3	2.3	0.1	1.0	EST: Gm-c1036-2780, <i>Glycine max</i> NR: ethylene responsive protein (EREB) mRNA, <i>Glycine max</i>	BM086625.1 AF537220.1	6e-12 0	60/67 470/487	
314	103	17.9	20.7	19.9	13.4	18.0	No homology	-	-	-	
315	102	-3.1	-2.2	-1.2	-0.4	-1.7	EST: PCSC19358, <i>Phaseolus coccineus</i> NR: no homology	CA916614.1 -	4e-7 -	63/71 -	
316	95	-0.9	0.5	-0.2	2.7	0.5	NR: QM-like protein mRNA, <i>Camellia sinensis</i>	AY641733.2	9e-5	33/35	
317	100	-0.3	0.2	0.0	-0.1	0.0	NR: putative histone H3 mRNA, <i>Arabidopsis thaliana</i>	NM_101207.2	7e-15	77/88	
318	102	-1.9	0.1	0.4	-0.4	-0.4	EST: PVEPLE1001B11, <i>Phaseolus vulgaris</i> NR: no homology	CB539105.1 -	e-28 -	88/95 -	
319	59	-1.6	-1.1	-0.1	0.0	-0.7	EST: PCSC20674, <i>Phaseolus coccineus</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	CA913383.1 EF147758.1	9e-7 6e-17	42/46 100/131	
320	61	-5.4	-1.0	-0.3	-0.1	-1.7	EST: Gm-c1080-2816, <i>Glycine max</i> NR: nucleoside diphosphate kinase mRNA, <i>Glycine max</i>	BU764764.1 AY157740.1	6e-8 0	48/52 413/416	
321	95	-0.4	0.4	0.5	0.9	0.4	NR: 70 kDa heat shock cognate protein 1 mRNA, <i>Vigna radiata</i>	AY485986.1	6e-40	92/94	
322	83	0.6	0.3	1.0	2.9	1.2	NR: fiber protein Fb19 mRNA, <i>Gossypium barbadense</i>	AY429440.1	e-4	47/54	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
323	92	1.4	-1.0	-1.9	-1.3	-0.7	EST: Gm-c1073-3127, <i>Glycine max</i> NR: ribulose-5-phosphate-3-epimerase mRNA, <i>Pisum sativum</i>	BQ298142.1 AF369887.1	7e-18 4e-40	74/81 119/134	
324	99	-0.9	-0.2	-0.3	0.2	-0.3	EST: PCS03780, <i>Phaseolus coccineus</i> NR: no homology	CA904765.1 -	8e-24 -	77/83 -	
325	122	-0.2	-0.1	0.6	-4.9	-1.2	EST: LR0938, <i>Phaseolus acutifolius</i> NR: lectin mRNA, <i>Glycine max</i>	CX129865.1 AJ010265.1	2e-9 e-67	56/63 268/342	
326	91	-0.1	0.1	2.3	4.5	1.7	No homology	-	-	-	
327	103	-0.1	-0.8	-0.1	0.1	-0.2	No homology	-	-	-	
328	115	-8.5	-12.2	-20.5	-22.8	-16.0	EST: PCEP00684, <i>Phaseolus coccineus</i> NR: hypothetical protein, <i>Gossypium hirsutum</i>	CA896844.1 X54092.1	e-16 e-34	59/63 238/333	
329	162	-2.8	-4.3	-14.7	-19.6	-10.3	NR: aldolase gene, <i>Pisum sativum</i>	M97476.1	3e-15	69/77	
330	99	1.0	0.1	0.8	0.7	0.6	EST: PVEPSE3030L18, <i>Phaseolus vulgaris</i> NR: no homology	CB543656.1 -	2e-30 -	82/85 -	
331	107	9.5	19.5	7.7	3.4	10.0	No homology	-	-	-	
332	95	0.9	-1.2	-0.8	-0.2	-0.3	EST: Gm-r1089-476, <i>Glycine max</i> NR: no homology	CO985565.1 -	3e-17 -	75/84 -	
333	98	-1.9	0.5	0.1	0.8	-0.1	No homology	-	-	-	
334	113	-0.3	0.5	1.2	-0.3	0.3	EST: PCSC18417, <i>Phaseolus coccineus</i> NR: snRNP core Sm protein Sm-X5-like protein mRNA, <i>Zea mays</i>	CA914183.1 EU975726.1	6e-25 e-46	79/85 210/278	
335	119	7.9	14.9	12.0	8.6	10.8	No homology	-	-	-	
336	92	1.0	0.1	-0.4	-4.9	-1.0	EST: PV_GEa0015B_D12.b1, <i>Phaseolus vulgaris</i> NR: ho homology	CV530916.1 -	2e-27 -	71/73 -	
337	97	-1.2	-12.2	-27.5	-9.9	-12.7	EST: T7_A05_047, <i>Vigna unguiculata</i> NR: putative tonoplast intrinsic protein mRNA, <i>Pisum sativum</i>	DR068287.1 AJ243309.1	3e-11 e-98	84/99 368/469	
338	74	21.0	3.2	9.2	6.0	9.8	No homology	-	-	-	
339	49	-2.8	-1.6	-3.9	-3.3	-2.9	EST: PCS03731, <i>Phaseolus coccineus</i> NR: eukaryotic translation initiation factor 5A isoform III mRNA, <i>Hevea brasiliensis</i>	CA905597.1 AF516352.1	7e-10 8e-80	35/35 231/271	
340	79	0.2	0.0	0.2	0.7	0.3	No homology	-	-	-	
341	89	-41.0	n.d.	6.0	-0.8	-9.0	No homology	-	-	-	
342	53	-1.4	-0.3	0.0	-0.5	-0.5	No homology	-	-	-	
343	121	-0.9	0.4	1.8	1.0	0.6	NR: ribosomal protein S15A D mRNA, <i>Arabidopsis thaliana</i>	NM_114473.5	3e-8	48/53	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
344	88	1.1	0.2	3.6	0.6	1.4	No homology	-	-	-	
345	93	-18.1	-4.9	-7.8	-10.8	-10.4	EST: HDI_K_35.esd, <i>Macrotyloma uniflorum</i> NR: rhabdomic opsin mRNA, <i>Platynereis dumerilii</i> NR: putative tetrahydrofolate synthase mRNA, <i>Trifolium pratense</i> NR: sinapyl alcohol dehydrogenase mRNA, <i>Plantago major</i>	DR989514.1 AJ316544.1 AB236821.1 AM159094.1	e-31 6e-17 7e-16 7e-16	84/88 67/73 68/76 68/74	
346	97	-0.9	-0.6	0.6	0.9	0.0	EST: PCSC18771, <i>Phaseolus coccineus</i> NR: hydroxyproline rich glycoprotein mRNA, <i>Pisum sativum</i>	CA904065.1 U78952.1	e-25 2e-76	80/86 244/296	
347	83	-2.1	0.2	0.0	-0.8	-0.7	NR: ribosomal protein S29 mRNA, <i>Oryza sativa</i>	EF576076.1	6e-24	77/83	
348	86	-8.1	-0.1	-0.3	-1.1	-2.4	NR: putative ribosomal protein S27 mRNA, <i>Hordeum vulgare</i>	AJ314595.1	3e-13	62/69	
349	108	-0.2	-16.5	-2.6	-12.6	-8.0	NR: glyoxysomal malate dehydrogenase mRNA, <i>Medicago sativa</i>	AF527542.1	6e-34	100/108	
350	94	-2.8	-0.4	-1.2	-1.5	-1.5	EST: PV_GEA0012C_E11.b1, <i>Phaseolus vulgaris</i> NR: unknown protein mRNA, <i>Cicer arietinum</i>	CV530067.1 EF375947.1	7e-18 3e-44	80/90 264/375	
351	76	-0.1	1.3	-1.2	0.0	0.0	NR: putative 23S rRNA pseudouridine synthase mRNA, <i>Vigna unguiculata</i>	AM748496.1	4e-9	47/50	
352	52	0.6	3.0	0.6	2.6	1.7	No homology	-	-	-	
353	99	-1.7	-0.6	-0.3	0.3	-0.6	EST: PCSC21589, <i>Phaseolus coccineus</i> NR: ribosomal protein L35 mRNA, <i>Euphorbia esula</i>	CA904760.1 AF227980.1	4e-10 2e-96	60/68 299/365	
354	129	-1.1	-0.5	-2.6	1.2	-0.8	EST: PCSC21049, <i>Phaseolus coccineus</i> NR: allantoinase mRNA, <i>Robinia pseudoacacia</i>	CA911778.1 AY466437.1	e-14 3e-124	65/72 426/535	
355	101	-6.6	0.3	1.2	0.1	-1.3	NR: snRNP-related protein mRNA, <i>Medicago sativa</i>	X63376.1	8e-27	91/100	
356	118	-1.9	-1.3	-5.4	-4.1	-3.2	EST: ABWZ1314, <i>Glycine max</i> NR: DNAJ heat shock protein, <i>Arabidopsis thaliana</i>	EH220440.1 NM_117003.2	7e-28 2e-84	102/114 337/443	
357	76	-3.5	0.7	0.3	0.7	-0.5	NR: small GTP-binding protein mRNA, <i>Lotus japonicus</i>	Z73960.1	5e-18	61/65	
358	60	-0.1	-1.4	-2.3	-0.6	-1.1	No homology	-	-	-	
359	101	-2.5	0.9	-0.7	-18.9	-5.3	NR: accd gene, <i>Nicotiana glauca</i>	AM286777.1	e-34	92/97	
360	62	-1.5	-0.9	-2.3	-2.6	-1.8	EST: Gm-c1049-6380 <i>Glycine max</i> NR: no homology	BU081263.1 -	3e-10 -	49/52 -	
361	104	-0.7	0.8	0.0	-2.6	-0.6	NR: 26S ribosomal RNA, <i>Glycyrrhiza uralensis</i>	EF571299.1	6e-43	101/103	
362	100	-0.9	0.6	4.8	7.4	3.0	EST: GmO3_198, <i>Glycine max</i> NR: no homology	EG702111.1 -	7e-6 -	39/41 -	
363	86	3.7	14.9	3.1	23.3	11.3	No homology	-	-	-	
364	73	-20.0	-2.8	-12.7	-12.0	-11.9	No homology	-	-	-	
365	58	-2.5	0.6	-1.5	1.9	-0.3	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
366	56	1.0	-1.5	-21.5	-39.1	-15.3	EST: POD_019_F11, <i>Phaseolus vulgaris</i> NR: proline-rich protein mRNA, <i>Glycine max</i>	CV539499.1 AF248055.1	8e-7 3e-177	45/50 514/613	
367	96	-4.7	-0.5	-0.7	-3.5	-2.4	EST: HDI_B_27.esd, <i>Macrotyloma uniflorum</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	DR988689.1 EF145605.1	8e-24 3e-54	85/93 172/205	
368	107	-1.9	-1.8	-10.8	-26.1	-10.1	No homology	-	-	-	
369	104	0.7	0.4	0.0	-0.6	0.1	EST: PCSC19657, <i>Phaseolus coccineus</i> NR: DNA-directed RNA polymerase II 8.2 kDa polypeptide mRNA, <i>Zea mays</i>	CA902226.1 EU958689.1	3e-17 3e-47	78/88 175/219	
370	89	1.4	1.8	0.4	2.0	1.4	No homology	-	-	-	
371	95	4.8	17.9	20.4	8.6	12.9	No homology	-	-	-	
372	102	7.4	15.1	16.9	20.5	15.0	No homology	-	-	-	
373	80	-3.3	-3.2	-5.9	-20.7	-8.2	EST: Gm-c1056-2496, <i>Glycine soja</i> NR: ATP synthase delta subunit mRNA, <i>Pisum sativum</i>	CA785102.1 M94558.1	6e-24 3e-68	68/71 348/480	
374	110	-2.2	-2.3	1.0	0.7	-0.7	EST: LR0981, <i>Phaseolus acutifolius</i> NR: ribosomal protein S17 mRNA, <i>Arabidopsis thaliana</i>	CX129904.1 NM_180434.2	4e-26 6e-78	87/95 279/354	
375	58	7.4	43.4	4.1	n.d.	13.7	No homology	-	-	-	
376	102	-12.4	-6.9	-2.9	-8.0	-7.6	No homology	-	-	-	
377	84	1.3	2.2	0.2	0.9	1.2	EST: POD_016_G07, <i>Phaseolus vulgaris</i> NR: RNA Binding Protein 47 mRNA, <i>Nicotiana plumbaginifolia</i>	CV539282.1 AJ292768.1	3e-29 2e-65	80/84 295/395	
378	71	1.7	4.7	0.4	-2.0	1.2	No homology	-	-	-	
379	119	-15.0	-8.2	-12.7	-6.6	-10.6	No homology	-	-	-	
380	91	-1.5	-1.9	-3.9	-1.3	-2.1	NR: putative pyridoxine biosynthetic enzyme mRNA, <i>Phaseolus vulgaris</i>	AY007525.1	e-19	61/64	
381	60	10.5	5.5	10.0	8.9	8.7	No homology	-	-	-	
382	130	0.3	0.0	1.4	0.7	0.6	EST: PCSC15061, <i>Phaseolus coccineus</i> NR: drought-responsive family protein mRNA, <i>Arabidopsis thaliana</i>	CA909578.1 NM_122581.2	6e-41 2e-38	119/127 188/251	
383	93	-13.3	-0.5	-1.2	-0.8	-3.9	EST: BE-1337 <i>Phaseolus vulgaris</i> NR: ribosomal protein L30 mRNA, <i>Lupinus luteus</i>	EX303723.1 AJ223316.1	4e-13 6e-120	48/49 305/346	
384	107	-0.5	-0.8	-0.4	3.1	0.3	NR: ribosomal protein L39 mRNA, <i>Triticum aestivum</i>	AY846827.1	2e-20	65/69	
385	80	0.6	-1.0	-1.0	0.5	-0.2	EST: RTS_129_D10, <i>Phaseolus vulgaris</i> NR: no homology	CV543031.1 -	4e-19 -	57/59 -	
386	93	1.4	-0.3	0.0	1.2	0.6	NR: hypothetical protein mRNA, <i>Cicer arietinum</i>	AJ012688.1	e-13	54/58	
387	105	-0.9	-0.2	0.5	-2.4	-0.7	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
388	106	-0.9	0.4	-0.1	-12.0	-3.2	EST: Gm-r1089-6430, <i>Glycine max</i> NR: unknown mRNA, <i>Populus trichocarpa</i> x <i>Populus deltoides</i>	CO983124.1 EF148649.1	5e-10 2e-114	89/106 332/398	
389	125	-0.9	-1.4	0.4	0.1	-0.5	NR: copper chaperone mRNA, <i>Arabidopsis thaliana</i>	NM_115482.3	e-13	63/70	
390	97	5.3	7.0	1.8	9.5	5.9	No homology	-	-	-	
391	100	11.6	18.3	13.0	17.0	15.0	EST: GLLAK58TF, <i>Glycine max</i> NR: no homology	EV264613.1 -	e-25 -	89/98 -	
392	95	15.8	11.3	13.3	15.8	14.0	EST: gmrtDrNS01_37-D_M13R_E07_055.s4, <i>Glycine max</i> NR: no homology	CX711734.1 -	3e-26 -	87/95 -	
393	83	2.1	5.1	2.1	15.6	6.2	EST: RTS_106_G11, <i>Phaseolus vulgaris</i> NR: transcription factor bZIP33 mRNA, <i>Glycine max</i>	CV541547.1 DQ787063.1	5e-6 e-180	35/37 551/678	
394	94	-0.9	-0.4	-2.9	-0.6	-1.2	No homology	-	-	-	
395	103	-2.3	-4.6	-1.2	-2.6	-2.7	EST: PVEPSE2028C07, <i>Phaseolus vulgaris</i> NR: no homology	CB541056.1 -	e-28 -	94/103 -	
396	85	-1.6	0.2	-0.2	-4.4	-1.5	EST: PCSC09373, <i>Phaseolus coccineus</i> NR: ribosomal protein L12 mRNA, <i>Prunus armeniaca</i>	CA903972.1 U93168.1	4e-13 7e-69	56/61 244/307	
397	133	0.0	6.1	1.5	2.9	2.6	No homology	-	-	-	
398	107	-0.9	-0.3	-6.4	-14.2	-5.4	EST: BE-18, <i>Phaseolus vulgaris</i> NR: glyceraldehyde 3-phosphate dehydrogenase C subunit mRNA, <i>Glycine max</i>	EE743279.1 DQ192668.1	2e-24 4e-129	74/78 437/547	
399	51	-9.0	-5.6	-13.7	-2.9	-7.8	No homology	-	-	-	
400	74	-0.9	-0.8	-4.9	-9.8	-4.1	EST: PCSC18500, <i>Phaseolus coccineus</i> NR: geranylgeranyl hydrogenase mRNA, <i>Glycine max</i>	CA914429.1 AF068686.3	e-4 e-173	59/70 452/517	
401	122	-0.3	-0.1	0.2	-1.1	-0.3	EST: NOD_246_G01, <i>Phaseolus vulgaris</i> NR: ATP-dependent Clp protease adaptor protein containing protein mRNA, <i>Zea mays</i>	CV537808.1 EU958506.1	2e-6 7e-39	63/73 245/346	
403	83	-0.7	-1.0	-1.6	-3.1	-1.6	EST: PVEPSE3016E08, <i>Phaseolus vulgaris</i> NR: no homology	CB541610.1 -	2e-17 -	54/56 -	
404	101	2.1	2.2	6.0	21.7	8.0	EST: LR0851, <i>Phaseolus acutifolius</i> NR: heat shock protein 70 gene, <i>Glycine max</i>	CX129789.1 X62799.1	7e-9 5e-67	58/65 213/263	
405	118	-3.8	-0.7	-0.3	0.4	-1.1	EST: PCSC14767, <i>Phaseolus coccineus</i> NR: ribosomal protein S8 mRNA, <i>Arabidopsis thaliana</i>	CA903592.1 NM_122036.4	e-10 6e-15	40/41 57/62	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
406	99	1.4	-0.5	-0.1	15.8	4.1	EST: PCSC14456, <i>Phaseolus coccineus</i> NR: no homology	CA908721.1	3e-14	86/99	
407	93	-5.0	0.6	1.2	-0.4	-0.9	NR: putative ribosomal protein S9 mRNA, <i>Artemisia annua</i>	EF549580.1	3e-17	81/92	
408	102	0.2	0.6	1.0	1.6	0.8	EST: PVEPSE3028M17, <i>Phaseolus vulgaris</i> NR: no homology	CB543245.1	4e-26	76/79	
409	106	-0.1	-0.5	0.0	-3.1	-0.9	EST: PV_GEa0015A_C11.b1, <i>Phaseolus vulgaris</i> NR: dynein light chain mRNA, <i>Vicia faba</i>	CV530842.1 AB215106.1	e-10 6e-14	52/57 82/101	
410	97	-1.6	-0.9	1.0	4.0	0.6	NR: low molecular weight heat shock protein 23.9 mRNA, <i>Glycine max</i>	U21722.1	2e-27	89/97	
411	106	-1.4	-5.2	-8.8	-5.8	-5.3	NR: glyceraldehyde-3-phosphate dehydrogenase B subunit mRNA, <i>Glycine max</i>	DQ224371.1	5e-13	62/69	
412	87	0.7	0.3	-0.7	-1.4	-0.3	NR: 14-.3.3 protein mRNA, <i>Glycine max</i>	AF532628.1	2e-24	73/76	
413	94	0.3	1.4	3.6	1.2	1.6	NR: protease inhibitor mRNA, <i>Glycine max</i>	U12150.1	e-22	84/93	
414	112	-1.3	-0.9	1.3	n.d.	-0.2	EST: PCSC19237, <i>Phaseolus coccineus</i> NR: putative L24 ribosomal protein mRNA, <i>Arachis hypogaea</i>	CA904425.1 DQ889568.1	2e-18 5e-128	74/82 406/505	
415	134	-4.7	0.9	-1.7	-6.2	-2.9	EST: VVL133C09, <i>Vitis vinifera</i> NR: photosystem II CP47 chlorophyll apoprotein mRNA, <i>Leea quineensis</i>	DT037915.1 EU002427.1	e-20 0	94/106 591/623	
416	97	-1.7	-0.5	-2.9	-0.5	-1.4	No homology	-	-	-	
417	100	-11.0	-1.1	-2.9	-2.6	-4.4	EST: RTS_129_B11, <i>Phaseolus vulgaris</i> NR: no homology	CV543011.1	3e-20	84/93	
418	88	2.6	1.1	1.2	n.d.	1.2	No homology	-	-	-	
419	96	0.3	-0.8	-0.5	-0.4	-0.3	NR: ribosomal protein S19 mRNA, <i>Cicer arietinum</i>	AJ012684.1	2e-9	80/95	
420	109	4.2	1.3	2.6	1.4	2.4	No homology	-	-	-	
421	94	-1.9	-0.1	0.3	-0.8	-0.6	NR: putative ribosomal protein s12 mRNA, <i>Fragaria x ananassa</i>	U19940.1	2e-24	84/92	
422	109	15.8	7.5	5.0	27.7	14.0	No homology	-	-	-	
423	103	0.2	3.9	-1.9	0.3	0.6	No homology	-	-	-	
424	86	0.5	0.1	5.1	13.4	4.8	EST: WHAGA031_F07, <i>Aphis gossypii</i> NR: no homology	DR391648.1	9e-5	30/31	
425	101	5.3	6.6	19.9	21.7	13.4	No homology	-	-	-	
426	101	-0.9	0.3	-1.9	0.1	-0.6	EST: PCSC13066, <i>Phaseolus coccineus</i> NR: putative epsilon subunit of mitochondrial F1-ATPase, <i>Cicer arietinum</i>	CA901296.1 AJ487471.1	2e-12 3e-85	89/101 343/442	
427	44	-8.5	-2.1	-1.3	-4.4	-4.1	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
428	92	-0.1	-1.6	-11.7	-4.9	-4.6	EST: ALV_012A_F07, <i>Phaseolus vulgaris</i> NR: photosystem II reaction center, <i>Retama raetam</i>	CV529905.1 AF439283.1	4e-10 3e-93	39/40 252/290	
429	103	-1.1	-3.4	-0.5	-0.8	-1.4	EST: PCSC17141, <i>Phaseolus coccineus</i> NR: ribosomal protein l22-2 mRNA, <i>Oryza sativa</i>	CA904332.1 EF576241.1	4e-32 e-78	96/103 257/316	
430	89	-3.3	-1.5	-0.4	-1.3	-1.6	NR: ribosomal protein small subunit 28 mRNA, <i>Helianthus annuus</i>	AF522185.1	9e-11	70/81	
431	118	-0.9	0.3	-0.3	0.5	-0.1	EST: PCSC18992, <i>Phaseolus coccineus</i> NR: no homology	CA906839.1 -	2e-15 -	45/45 -	
432	96	3.9	3.5	-0.7	3.1	2.4	EST: PCEP05722, <i>Phaseolus coccineus</i> NR: ATP sulfurylase mRNA, <i>Glycine max</i>	CA896660.1 AF452454.2	3-13 e-133	51/54 334/372	
433	48	1.3	1.1	1.0	1.8	1.3	EST: Gm-c1087-5370, <i>Glycine max</i> NR: ribosomal protein L10A mRNA, <i>Arabidopsis thaliana</i>	BQ611699.1 NM 100709.4	2e-6 e-31	29/29 133/168	
434	144	-0.9	-0.1	0.6	3.8	0.9	EST: RTS_121_A07, <i>Phaseolus vulgaris</i> NR: putative DIM-like protein mRNA, <i>Glycine max</i>	CV542488.1 AY323127.1	7e-7 e-147	38/39 356/397	
435	95	-0.6	-0.9	0.0	-1.3	-0.7	NR: RING-box protein mRNA, <i>Arachis hypogaea</i>	DQ294622.1	3e-26	87/95	
436	87	-2.1	-1.6	-0.4	3.4	-0.2	NR: calmodulin mRNA, <i>Glycine max</i>	L01430.1	8e-8	44/47	
437	88	-1.1	0.1	0.0	-3.5	-1.1	No homology	-	-	-	
438	67	0.8	2.1	-0.3	4.0	1.6	NR: salt-tolerance protein mRNA, <i>Glycine max</i>	DQ234265.1	6e-8	38/40	
439	92	0.4	-14.8	-0.1	-1.7	-4.1	NR: histone H3 mRNA, <i>Robinia pseudoacacia</i>	DQ917751.1	4e-10	60/68	
440	72	-1.1	-4.3	0.5	-0.8	-1.4	No homology	-	-	-	
441	98	-4.7	0.0	-1.6	-2.2	-2.1	EST: PCSC18275, <i>Phaseolus coccineus</i> NR: acidic ribosomal protein P1a-like mRNA, <i>Solanum tuberosum</i>	CA903786.1 DQ235177.1	e-28 e-33	88/95 216/300	
442	104	0.9	1.3	2.4	n.d.	1.2	EST: RTS_108_F03, <i>Phaseolus vulgaris</i> NR: no homology	CV541668.1 -	2e-28 -	83/87 -	
443	102	-0.4	0.6	0.2	0.4	0.2	NR: ribosomal protein L17 mRNA, <i>Castanea sativa</i>	AF334838.1	e-22	81/89	
444	101	-1.1	-1.0	-0.3	-0.6	-0.8	EST: GLMA194TF, <i>Glycine max</i> NR: hypothetical protein mRNA, <i>Arabidopsis thaliana</i>	EV268043.1 AK226354.1	3e-5 2e-25	50/56 141/188	
445	96	-0.6	-0.1	-1.3	-0.5	-0.6	No homology	-	-	-	
446	105	-0.2	-1.4	-0.5	0.1	-0.5	EST: PCSC18504, <i>Phaseolus coccineus</i> NR: actin 2-like mRNA, <i>Phaseolus vulgaris</i>	CA907968.1 EU369189.1	4e-32 5e-78	97/105 189/204	
447	103	3.2	1.3	17.9	13.4	8.9	EST: Gm_ck44037, <i>Glycine max</i> NR: no homology	CD412848.1 -	2e-6 -	39/42 -	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
448	106	1.8	2.7	1.0	1.6	1.8	EST: POD_015_A09, <i>Phaseolus vulgaris</i> NR: no homology	CV539154.1 -	3e-20 -	72/77 -	
449	97	2.8	0.5	13.0	1.2	4.4	No homology	-	-	-	
450	112	-0.6	-1.6	-10.7	-2.6	-3.9	EST: LVS_023_C11, <i>Phaseolus vulgaris</i> NR: fibrillin gene, <i>Pisum fulvum</i>	CV532305.1 EU270992.1	e-50 2e-25	110/112 90/104	
451	98	-0.3	0.0	-0.3	6.2	1.4	NR: heat shock protein mRNA, <i>Hevea brasiliensis</i>	AF521007.1	6e-6	35/37	
452	111	2.1	0.1	-0.2	-0.2	0.5	NR: progesterone 5-beta-reductase gene, <i>Digitalis cariensis</i>	DQ213016.1	7e-6	38/41	
453	99	-8.0	-0.8	-7.8	-8.0	-6.2	EST: PCSC16115, <i>Phaseolus coccineus</i> NR: inorganic pyrophosphatase mRNA, <i>Zea mays</i>	CA900434.1 EU955065.1	5e-19 e-45	84/94 162/199	
454	94	-2.1	-4.6	-3.4	n.d.	-2.5	EST: RTS_128_D03, <i>Phaseolus vulgaris</i> NR: copper/zinc-superoxide dismutase mRNA, <i>Pisum sativum</i>	CV542947.1 J04087.1	2e-24 3e-98	79/84 272/317	
455	50	3.2	7.0	13.0	14.6	9.4	No homology	-	-	-	
456	53	-0.1	-0.8	-1.0	6.7	1.2	EST: PCSC12203, <i>Phaseolus coccineus</i> NR: endothelial differentiation-related factor 1 mRNA, <i>Zea mays</i>	CA902615.1 EU961228.1	e-17 2e-36	51/52 121/142	
457	133	-0.1	-1.3	-2.4	-4.4	-2.1	EST: PCS05094, <i>Phaseolus coccineus</i> NR: myo-inositol 1-phosphate synthase mRNA, <i>Phaseolus vulgaris</i>	CA900507.1 AJ853494.2	e-4 2e-65	37/39 145/147	
458	72	0.2	0.3	-1.7	n.d.	-0.3	No homology	-	-	-	
459	95	-1.9	-0.2	-1.2	1.2	-0.5	EST: ABWZ4766, <i>Glycine max</i> NR: small GTP-binding protein, <i>Lotus japonicus</i>	EH224100.1 Z73938.1	7e-15 0	50/52 503/564	
460	96	0.6	3.1	0.7	-0.5	1.0	No homology	-	-	-	
461	105	-10.4	-0.5	-9.8	-17.4	-9.5	No homology	-	-	-	
462	97	-2.5	-0.1	2.6	-1.5	-0.4	EST: Gm-c1074-7486, <i>Glycine max</i> NR: cysteine synthase mRNA, <i>Glycine max</i>	BU577355.1 EF433420.1	3e-5 8e-116	46/52 285/314	
463	94	0.5	0.3	-2.9	6.2	1.0	NR: 26S ribosomal RNA, <i>Glycyrrhiza uralensis</i>	EF571299.1	9e-45	94/94	
464	45	-5.7	-0.2	-11.8	-4.9	-5.6	No homology	-	-	-	
465	94	1.1	2.8	-0.5	4.5	2.0	No homology	-	-	-	
466	92	8.4	3.0	3.1	17.0	7.9	EST: Gm-c1073-2431, <i>Glycine max</i> NR: no homology	BQ298693.1 -	3e-8 -	81/93 -	
467	101	1.1	-3.4	0.3	0.8	-0.3	EST: PCSC21498, <i>Phaseolus coccineus</i> NR: no homology	CA911022.1 -	5e-13 -	68/77 -	
468	46	-1.2	-0.4	-0.2	0.5	-0.3	No homology	-	-	-	
469	91	1.7	0.8	2.1	-1.1	0.9	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
470	97	-4.2	-3.4	-5.4	-8.7	-5.4	EST: ALV_013D_D10, <i>Phaseolus vulgaris</i> NR: thylakoid membranephosphoprotein of 14 kDa mRNA, <i>Arabidopsis thaliana</i>	CV530468.1 NM_130248.4	9e-27 3e-50	88/95 219/288	
471	93	1.8	0.7	-0.7	10.1	3.0	EST: Gm-c1073-5579, <i>Glycine max</i> NR: Ser/Thr protein kinase mRNA, <i>Medicago truncatula</i>	BQ628013.1 DQ984942.1	e-7 5e-25	53/60 102/123	
472	101	0.5	5.8	1.5	-2.2	1.4	EST: Gm-c1052-2421, <i>Glycine max</i> NR: carboxyl-terminal-processing protease precursor mRNA, <i>Zea mays</i>	BQ094115.1 EU960627.1	8e-18 9e-23	83/92 118/154	
473	107	2.9	2.4	5.0	10.1	5.1	No homology	-	-	-	
474	86	4.8	7.5	7.7	7.4	6.8	NR: unknown gene, <i>Davidia involucrata</i>	AF448811.1	2e-8	57/68	
475	87	-0.9	0.0	-1.0	1.2	-0.2	NR: ribosomal protein PETRP-like mRNA, <i>Solanum tuberosum</i>	DQ207869.1	9e-17	65/71	
476	104	-7.0	-0.6	-11.7	-9.8	-7.3	EST: GLMAY90TF, <i>Glycine max</i> NR: heat shock factor-binding protein 1 mRNA, <i>Zea mays</i>	EV270725.1 EU965664.1	e-13 5e-26	57/62 114/143	
477	44	-0.9	-1.0	-2.9	-0.6	-1.3	EST: PCSC20808, <i>Phaseolus coccineus</i> NR: no homology	CA908885.1	e-7	37/39	
478	76	-7.6	-5.2	-1.0	0.1	-3.4	No homology	-	-	-	
479	101	-2.8	0.4	-1.0	-4.0	-1.8	EST: PV_GEa0122p23f, <i>Phaseolus vulgaris</i> NR: CONSTANS-like 2b mRNA, <i>Glycine max</i>	BQ481806.1 DQ371244.1	e-10 4e-157	47/49 513/637	
480	134	-4.0	-6.0	-9.8	-22.8	-10.7	No homology	-	-	-	
481	71	-0.5	3.5	0.0	5.6	2.2	No homology	-	-	-	
482	100	-0.9	-3.8	-1.9	-2.6	-2.3	NR: ribosomal protein L37a mRNA, <i>Gossypium hirsutum</i>	AF127042.1	2e-24	90/100	
483	110	-0.3	-0.3	-1.5	2.3	0.1	EST: PCSC17987, <i>Phaseolus coccineus</i> NR: unknown mRNA, <i>Populus trichocarpa</i>	CA901238.1 EF147836.1	9e-15 3e-21	79/93 106/136	
484	106	11.6	13.2	13.0	6.2	11.0	No homology	-	-	-	
485	94	7.4	24.0	8.2	0.7	10.1	EST: slv1c.pk001.h3, <i>Glycine max</i> NR: no homology	EE400004.1	2e-6	51/58	
486	130	-0.7	-3.1	-1.6	-4.4	-2.5	NR: 14-3-3 related protein mRNA, <i>Glycine max</i>	U70536.1	3e-30	109/118	
487	115	-1.9	0.6	-0.6	1.2	-0.2	NR: endopeptidase gene, <i>Phaseolus vulgaris</i>	X63102.1	2e-23	76/82	
488	95	0.3	0.8	-1.0	0.5	0.2	EST: 51-PTE-4192, <i>Glycine soja</i> NR: no homology	DT083632.1	4e-10	42/44	
489	98	-0.9	1.3	-2.6	1.0	-0.3	EST: PCSC15494, <i>Phaseolus coccineus</i> NR: no homology	CA916471.1	e-10	68/77	
490	105	-3.8	-1.1	-9.8	-15.3	-7.5	No homology	-	-	-	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
491	88	-10.0	0.9	0.6	-3.5	-3.0	No homology	-	-	-	
492	78	0.3	5.6	1.3	1.2	2.1	EST: PVEPSE3030L15, <i>Phaseolus vulgaris</i> NR: no homology	CB543655.1	3e-13	50/53	
493	100	1.5	5.3	1.4	6.2	3.6	EST: Gm-c1068-3960 <i>Glycine max</i> NR: putative zinc dependent protease mRNA, <i>Trifolium pratense</i>	BI974577.1 AB236774.1	e-10 9e-67	75/78 258/337	
494	46	1.4	0.1	0.5	-1.3	0.2	No homology	-	-	-	
495	110	0.5	0.7	0.4	1.2	0.7	No homology	-	-	-	
496	95	-1.5	2.4	1.0	-3.3	-0.3	NR: S-adenosylmethionine decarboxylase gene, <i>Phaseolus vulgaris</i>	AY327898.1	e-16	71/79	
497	104	5.3	1.8	0.3	18.2	6.4	EST: PCS05419, <i>Phaseolus coccineus</i> NR: tetraspanin 8 mRNA, <i>Arabidopsis thaliana</i>	CA910584.1 NM_179714.4	e-4 e-28	32/34 216/310	
498	96	0.3	4.7	7.0	15.8	6.9	No homology	-	-	-	
499	112	5.3	2.4	2.7	2.9	3.3	EST: RTS_123_A09, <i>Phaseolus vulgaris</i> NR: no homology	CV542633.1	2e-6	72/85	
500	91	4.2	10.4	10.0	12.2	9.2	No homology	-	-	-	
501	108	-12.0	-11.5	-6.9	-1.7	-8.0	No homology	-	-	-	
502	95	-5.7	-8.7	1.5	-3.1	-4.0	NR: sedoheptulose-1,7-bisphosphatase mRNA, <i>Fragaria x ananassa</i>	AY679611.1	4e-7	58/67	
503	88	-1.9	-1.9	-1.2	-3.3	-2.1	EST: PVEPSE2004D06, <i>Phaseolus vulgaris</i> NR: no homology	CB539368.1	9e-5	30/31	
504	93	0.0	1.3	0.6	-3.3	-0.3	EST: PVEPSE3014D09, <i>Phaseolus vulgaris</i> NR: no homology	CB555991.1	5e-19	60/63	
505	96	-0.1	0.0	1.0	2.9	0.9	EST: POD_039_F08, <i>Phaseolus vulgaris</i> NR: no homology	CV541011.1	4e-38	92/95	
506	102	10.5	6.6	15.9	8.6	10.4	EST: TC-T7-ST3_E06_020, <i>Vigna unguiculata</i> NR: seed albumin mRNA, <i>Vigna radiata</i>	CK151447.1 X70671.1	3e-17 e-163	57/60 493/614	
507	105	-1.9	-0.2	-1.2	-0.4	-0.9	EST: Gm_ck35973, <i>Glycine max</i> NR: no homology	CD409129.1	3e-5	30/31	
508	81	0.7	-0.8	15.3	8.6	5.9	EST: NOD_210_G11, <i>Phaseolus vulgaris</i> NR: no homology	CV534958.1	3e-7	34/35	
509	92	-6.0	-0.8	-3.9	-8.0	-4.7	EST: Gm-c1076-3391, <i>Glycine max</i> NR: no homology	BQ740755.1	3e-14	67/75	
510	91	-1.9	-1.6	0.1	0.5	-0.7	EST: 89-PTE-5161, <i>Glycine soja</i> NR: no homology	DT084311.1	4e-16	70/78	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
511	100	0.9	0.9	4.1	1.6	1.9	EST: CB555936.1, <i>Phaseolus vulgaris</i> NR: no homology	CB555936.1	2e-21	87/97	
512	95	-0.9	-1.9	-1.9	-14.2	-4.7	EST: PV_GEa0017A_E10.b1, <i>Phaseolus vulgaris</i> NR: no homology	CV531415.1	e-34	86/89	
513	55	-1.4	-1.1	0.4	-0.4	-0.6	NR: U6 small nuclear RNA gene, <i>Lycopersicon esculentum</i>	X51447.1	3e-19	54/55	
514	55	-0.7	-7.8	0.0	0.1	-2.1	No homology	-	-	-	
515	94	1.1	0.7	1.9	2.7	1.6	EST: Gm-c1073-5081, <i>Glycine max</i> NR: no homology	BQ454183.1	e-25	86/94	
516	109	-1.5	0.3	1.7	-0.8	-0.1	NR: ribosomal protein S23 mRNA, <i>Hyacinthus orientalis</i>	AY389732.1	4e-32	97/105	
517	111	-11.0	0.5	0.3	0.1	-2.5	EST: PVEPSE3028P10, <i>Phaseolus vulgaris</i> NR: no homology	CB543275.1	e-16	88/100	
518	92	4.2	1.8	14.9	5.6	6.7	EST: BE-3484, <i>Phaseolus vulgaris</i> NR: no homology	EX305519.1	3e-26	78/83	
519	71	0.0	0.7	2.4	0.7	1.0	No homology	-	-	-	
520	108	0.3	1.1	-0.6	0.4	0.3	No homology	-	-	-	
521	115	4.8	1.7	-0.2	1.2	1.9	No homology	-	-	-	
522	94	-4.0	-10.4	-16.6	-3.5	-8.6	EST: GLMAW18TF, <i>Glycine max</i> NR: no homology	EV270495.1	5e-16	58/62	
523	56	0.0	-8.7	-9.8	-19.6	-9.5	No homology	-	-	-	
524	50	1.4	12.6	7.2	10.1	7.8	No homology	-	-	-	
525	118	-0.1	-0.8	1.4	0.4	0.2	NR: 14-3-3 protein mRNA, <i>Vigna angularis</i>	AB042299.1	8e-52	115/118	
526	81	-1.5	-1.4	0.0	-1.4	-1.1	EST: PVEPSE2007D02, <i>Phaseolus vulgaris</i> NR: ribulose-5-phosphate-3-epimerase, <i>Pisum sativum</i>	CB539575.1 AF369887.1	4e-22 8e-94	71/75 372/484	
527	107	7.4	5.8	15.9	8.9	9.5	No homology	-	-	-	
528	97	-2.2	-0.9	-4.9	-0.5	-2.1	NR: 20 kDa protein of CP24 mRNA, <i>Spinacea oleracea</i>	Z25886.1	2e-12	67/76	
529	129	-1.9	-0.1	4.0	-0.5	0.4	NR: SKP1-like b mRNA, <i>Medicago truncatula</i>	DQ641946.1	6e-50	121/127	
530	92	-0.2	-2.8	0.8	5.6	0.9	No homology	-	-	-	
531	58	0.6	1.8	0.4	0.4	0.8	No homology	-	-	-	
532	97	0.5	0.1	-0.3	4.5	1.2	No homology	-	-	-	
533	99	3.2	4.1	13.9	6.2	6.9	No homology	-	-	-	
534	106	0.0	-0.1	0.2	6.7	1.7	EST: LR-0238, <i>Phaseolus acutifolius</i> NR: putative synaptobrevin protein mRNA, <i>Arabidopsis thaliana</i>	DV643070.1 AK118318.1	2e-24 7e-45	93/104 227/307	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
535	48	9.5	1.3	5.6	2.3	4.7	No homology	-	-	-	
536	94	-1.9	-0.2	-0.2	-1.3	-0.9	NR: putative chloroplast thioredoxin m2 mRNA, <i>Pisum sativum</i>	AJ316577.1	2e-18	83/94	
537	44	12.6	5.6	13.0	8.6	9.9	No homology	-	-	-	
538	87	-5.7	-0.8	-4.9	-6.2	-4.4	EST: PVEPSE3003H12, <i>Phaseolus vulgaris</i> NR: Medicago truncatula rps16-1 mRNA for 30S ribosomal protein S16	CB541816.1	2e-6	45/50	
539	107	0.3	0.9	0.0	0.0	0.3	EST: PVEPSE3027E06, <i>Phaseolus vulgaris</i> NR: no homology	CB542967.1	9e-18	91/100	
540	104	2.1	0.1	2.7	n.d.	1.3	EST: SSH-70, <i>Phaseolus vulgaris</i> NR: no homology	EG594364.1	e-4	86/105	
541	113	-4.7	2.8	-4.9	-0.8	-1.9	EST: PCSC18407, <i>Phaseolus coccineus</i> NR: no homology	CA904438.1	3e-30	94/102	
542	94	-11.0	n.d.	n.d.	-5.3	-4.1	EST: GLNA382TF, <i>Glycine max</i> NR: no homology	EV278183.1	e-7	47/51	
544	85	-4.7	-1.6	-5.4	-5.3	-4.3	No homology	-	-	-	
545	91	-2.8	-0.4	-0.5	-1.3	-1.2	No homology	-	-	-	
546	123	-0.6	-1.6	0.0	2.3	0.0	No homology	-	-	-	
547	80	-0.2	-0.5	0.0	-0.5	-0.3	EST: PCSC11446, <i>Phaseolus coccineus</i> NR: no homology	CA909546.1	e-12	49/52	
548	81	-8.5	0.0	0.7	7.4	-0.1	No homology	-	-	-	
549	101	2.7	-0.1	-2.3	-0.5	0.0	EST: PVEPSE3010B04, <i>Phaseolus vulgaris</i> NR: putative nitrilase-associated protein mRNA, <i>Arachis hypogaea</i>	CB555970.1 DQ889507.1	2e-34 6e-25	92/97 84/95	
550	122	-8.5	2.0	2.1	-0.2	-1.2	EST: PCSC12287, <i>Phaseolus coccineus</i> NR: ubiquitin-like protein 5 mRNA, <i>Hordeum vulgare</i>	CA906871.1 EF143989.1	9e-9 2e-75	86/102 203/231	
551	91	1.4	1.3	3.1	0.2	1.5	NR: selenium binding protein mRNA, <i>Lotus japonicus</i>	AJ401227.1	2e-14	49/51	
552	97	-1.6	-2.5	-3.9	-3.3	-2.8	EST: BE-1556, <i>Phaseolus vulgaris</i> NR: no homology	EX304299.1	e-22	84/92	
553	96	-0.6	-2.5	-1.9	-9.0	-3.5	No homology	-	-	-	
554	96	0.0	9.2	9.2	18.2	9.2	No homology	-	-	-	
555	111	0.3	0.1	-1.9	-9.8	-2.8	NR: transcription factor MYB123 mRNA, <i>Glycine max</i>	DQ822934.1	6e-31	96/103	
556	82	0.7	-0.3	1.7	7.4	2.4	EST: ABWZ3760, <i>Glycine max</i> NR: no homology	EH223134.1	e-6	48/53	

Table 17 (continued).

MPSS cluster no.	Length [nucleotides]	Change after water-deficit treatment					Mean	Homology to sequences in 'non-redundant' (NR) and 'est_others' (EST) BLAST nucleotide databases	Homology to Acc. no.	E-value	Nucleotide identity
		DipC	AS-17	Swazi Red	LunT						
557	104	2.5	3.7	12.0	11.0	7.3	EST: Gm-c1068-3576, <i>Glycine max</i> NR: no homology	BI893450.1 -	7e-6 -	41/45 -	
558	108	-0.9	-1.9	-4.9	0.4	-1.8	EST: PvEST0509, <i>Phaseolus vulgaris</i> NR: no homology	EH040310.1 -	9e-18 -	70/77 -	
559	61	-0.7	1.0	-1.9	0.1	-0.4	No homology	-	-	-	
560	62	-4.2	1.8	-1.2	-5.3	-2.2	No homology	-	-	-	
561	98	-0.1	-0.1	0.0	0.4	0.1	NR: calmodulin mRNA, <i>Vigna radiata</i>	L20691.1	e-34	93/97	
562	124	-1.9	-1.0	-1.3	-5.5	-2.4	EST: PVEPSE2015F01, <i>Phaseolus vulgaris</i> NR: malate dehydrogenase mRNA, <i>Vitis vinifera</i>	CB540154.1 L34836.1	e-20 6e-35	68/73 140/176	
563	104	-2.2	4.1	2.6	2.3	1.7	NR: glucosyltransferase-3 mRNA, <i>Vigna angularis</i>	AB070746.1	4e-41	100/104	
564	89	1.4	12.6	5.1	11.0	7.5	EST: LR0964, <i>Phaseolus acutifolius</i> NR: no homology	CX129889.1 -	6e-12 -	45/47 -	
565	94	-8.0	-4.3	-7.8	-15.2	-8.8	EST: GLMBK34TF, <i>Glycine max</i> NR: no homology	EV272471.1 -	e-31 -	87/92 -	
566	106	-0.9	0.9	0.7	1.2	0.5	EST: POD_007_B01, <i>Phaseolus vulgaris</i> NR: ribosomal protein S14 mRNA, <i>Lupinus luteus</i>	CV538547.1 AF026079.1	3e-5 4e-37	67/76 134/161	
567	104	0.0	-1.6	-9.8	-8.7	-5.0	EST: SP6_A04_032, <i>Vigna unguiculata</i> NR: no homology	DR068284.1 -	4e-29 -	80/84 -	
568	95	-4.7	-2.5	4.6	-2.6	-1.3	EST: PCEP04620, <i>Phaseolus coccineus</i> NR: no homology	CA896889.1 -	2e-21 -	67/71 -	
569	97	8.4	7.5	9.0	13.4	9.6	No homology	-	-	-	
570	92	1.1	0.7	2.4	0.9	1.3	NR: ubiquitin carrier protein mRNA, <i>Medicago sativa</i>	L06967.1	3e-7	67/79	

Table 18. Comparison of \log_2 ratios obtained through MPSS and microarray analyses for 132 genes in the 2006 CE water-deficit experiment. \log_2 ratios ≥ 2 and ≤ -2 are highlighted in red and green, respectively (n.d. = not detected). Genes marked by an asterisk showed a potential qualitative difference between DipC and LunT according to MPSS.

MPSS cluster no.	MPSS				Microarrays			
	DipC	AS-17	Swazi Red	LunT	DipC	AS-17	Swazi Red	LunT
1	-1.6	-0.7	-1.4	-0.2	-0.9	-1.4	-0.4	0.0
2	6.8	7.0	9.7	10.0	4.5	6.4	19.1	12.5
3	2.7	2.2	4.2	3.7	1.7	2.3	3.0	3.7
4	3.5	1.7	3.6	3.9	1.0	0.4	1.8	2.5
5	2.6	1.7	5.5	6.5	2.3	2.9	5.8	8.5
8	-3.0	-2.6	-5.1	-6.0	-3.3	-4.9	-4.6	-5.3
9	-4.0	-3.4	-5.0	-8.7	-4.0	-6.2	-6.3	-7.9
10	1.8	1.2	3.4	2.5	1.6	2.1	2.8	2.6
11	-2.1	-1.9	-2.2	-2.7	-2.4	-3.4	-3.0	-3.9
12	0.5	0.9	1.3	1.5	0.0	0.9	1.9	2.3
14	0.4	0.9	0.3	1.0	0.8	-0.3	-0.3	0.6
15	0.4	-0.2	1.2	0.8	0.1	-0.6	0.2	1.0
18	7.9	7.4	7.6	5.5	3.0	5.4	16.4	8.5
19	-2.9	-2.0	-4.4	-7.5	-3.6	-5.8	-4.8	-10.2
20*	-2.5	-4.5	1.7	0.6	-1.6	-2.6	-0.3	-1.8
21*	-1.7	-1.3	-0.4	0.1	-1.2	-0.7	-0.1	-0.4
24	1.3	0.5	3.1	2.6	1.2	1.4	3.1	4.5
25	4.3	5.5	6.6	5.2	2.1	0.8	3.0	4.4
28	-2.9	-3.8	-4.5	-5.3	-3.4	-5.2	-8.7	-16.4
29	1.3	0.4	1.2	0.9	0.4	0.1	-0.3	0.1
33	-2.1	-0.3	-0.6	-0.2	-0.2	-0.2	-0.8	-0.3
34	7.8	4.1	6.1	7.0	5.1	5.3	7.5	12.3
36	5.3	6.7	6.0	7.3	5.9	4.8	17.7	11.7
39	2.5	2.1	4.2	3.7	2.3	2.6	2.7	6.1
41	0.5	0.7	0.5	0.4	0.8	-1.1	-0.6	0.4
43	3.3	3.2	3.5	3.1	3.1	2.5	3.1	4.7
44	-2.6	-1.3	-2.8	-4.0	-2.5	-4.3	-5.2	-5.2
46	1.6	1.9	3.3	2.6	1.4	0.8	1.3	2.4
49	-2.7	-2.0	-3.1	-3.9	-3.4	-4.9	-5.3	-12.0
50	0.4	0.0	0.0	0.4	0.3	-0.6	-0.1	0.5
52	4.8	3.9	6.9	7.8	2.5	5.6	16.7	17.0
61	-1.6	-1.3	-3.0	-3.6	-2.1	-3.3	-3.1	-4.1
63	-0.9	0.2	0.8	2.5	-2.1	0.1	1.0	3.2
64	-0.7	0.9	1.2	-0.5	0.4	0.2	0.4	1.5
65	-0.8	-1.1	0.8	0.0	0.0	0.1	0.4	0.2
66	0.6	0.0	-0.5	-0.7	-0.3	-0.9	-0.7	-1.1
68	6.6	4.1	6.8	6.0	3.9	2.5	10.1	10.9
70*	-1.7	-0.5	-0.2	0.5	-0.4	-1.1	-0.3	-0.2
71	1.9	1.6	2.2	2.7	2.5	2.1	2.5	4.0
74	0.9	0.5	0.6	-0.7	0.5	-0.5	0.3	2.0
76	-1.9	-2.8	-3.3	-3.8	-3.7	-3.4	-10.4	-13.1
77	0.7	1.1	0.0	0.0	0.7	0.0	-0.5	-0.3
80	1.1	1.5	3.2	3.4	1.0	1.6	2.9	5.3
82	5.0	3.8	6.0	6.4	2.2	1.7	6.7	6.1
84*	-1.3	0.1	0.1	1.0	-0.3	-0.4	-0.1	-0.2

Table 18 (continued).

MPSS cluster no.	MPSS				Microarrays			
	DipC	AS-17	Swazi Red	LunT	DipC	AS-17	Swazi Red	LunT
86	-0.2	0.6	0.2	2.0	0.9	-0.4	-0.9	0.4
89	0.2	0.7	0.7	1.4	0.7	0.3	0.0	0.5
93	-3.4	-2.4	-1.7	-4.7	-2.7	-2.5	-1.8	-4.7
97	1.0	1.6	1.7	4.1	0.9	-0.7	11.8	12.1
98	1.6	1.5	1.8	1.6	0.8	-0.4	1.2	2.4
99	-4.3	-3.6	-3.1	-6.4	-5.3	-3.1	-4.4	-6.8
101	3.3	2.8	6.7	4.9	4.7	5.6	14.0	7.4
102	5.6	3.8	6.4	6.4	3.6	4.4	15.0	7.7
107	1.5	1.9	2.0	2.4	0.7	-0.1	1.3	1.6
110	1.0	-1.7	-0.8	-0.1	-3.7	-3.5	-5.0	-12.2
111	4.8	3.5	5.8	5.8	3.6	4.2	17.1	13.9
112	5.1	4.5	4.7	5.7	3.1	2.5	15.2	7.6
114	2.5	3.5	5.7	4.3	2.9	2.5	3.5	3.3
115	1.6	2.4	1.9	1.7	2.2	1.5	2.5	5.0
116	0.1	-0.8	-0.9	-1.9	-1.4	-2.5	-8.4	-4.2
117	-2.0	-1.1	-2.1	-4.4	-3.3	-3.2	-14.8	-14.2
124	0.8	1.2	1.5	n.d.	0.1	0.3	0.9	0.0
126	4.8	5.8	5.7	5.7	2.9	3.2	14.8	7.3
129*	-1.5	0.6	0.3	0.7	-0.1	-0.7	-0.6	-0.7
130	0.4	0.2	0.1	1.6	-0.3	-0.3	0.7	1.2
133	1.8	3.4	5.3	4.9	2.9	3.0	8.6	4.4
139*	-1.0	-0.2	0.7	0.4	-1.5	-1.1	0.1	-2.4
143	-2.5	-1.5	-3.5	-4.6	-3.2	-3.7	-5.3	-6.3
144	-3.3	-1.3	-2.9	-2.4	-2.4	-3.2	-6.0	-11.6
152	5.1	3.8	6.5	3.0	4.1	3.2	15.4	10.4
154	0.4	0.5	0.7	4.8	1.0	0.3	2.2	10.6
155	-1.8	-0.7	-1.6	-2.4	-1.1	-1.8	-2.4	-2.3
157	0.1	0.9	1.7	0.7	-0.3	0.2	-0.4	-0.9
158	0.1	1.8	3.9	3.8	0.1	0.2	3.1	4.5
159	0.4	0.9	0.5	1.0	-0.2	-0.4	-1.1	-1.6
161	0.2	1.2	-0.1	0.3	0.1	-0.6	-0.1	-0.5
162	-2.0	-0.7	-4.4	-5.2	-3.9	-4.9	-9.9	-15.8
166	3.5	3.1	3.5	5.4	1.4	1.4	14.4	12.1
167	-1.9	-3.0	-2.9	-3.0	-3.7	-1.5	-3.6	-4.2
168*	1.8	1.4	2.7	-0.6	0.6	-0.5	1.9	1.0
169	2.2	0.3	-0.9	0.1	-0.4	-0.1	-0.1	-1.1
174	-0.9	-0.3	-3.3	-3.9	-1.5	-1.1	-2.8	-2.6
179*	-1.4	0.0	-0.8	0.1	0.3	-0.4	-0.4	-0.6
182	5.2	4.2	5.5	4.7	4.4	3.5	16.3	15.2
185	-4.5	-4.6	-5.3	-5.0	-4.9	-5.0	-11.4	-15.7
188	-2.5	-1.1	-4.8	-5.0	-3.8	-4.6	-15.4	-15.6
189*	1.2	0.1	-2.4	-2.4	0.2	-0.3	-2.5	-1.1
194	-1.2	0.0	-1.1	-1.3	-1.3	-2.8	-0.7	-1.7
195	5.5	4.6	5.6	n.d.	4.1	3.7	12.8	15.6
209	1.2	0.6	1.0	1.7	0.8	1.2	2.4	2.3
212	0.5	-0.6	0.5	-1.4	1.8	2.9	3.7	5.8
217	-0.4	0.0	-0.1	-1.9	-0.2	-2.1	-1.1	-0.3
226	-0.7	0.3	-0.8	-2.1	-0.1	-0.6	-0.7	0.7

Table 18 (continued).

MPSS cluster no.	MPSS				Microarrays			
	DipC	AS-17	Swazi Red	LunT	DipC	AS-17	Swazi Red	LunT
227	3.9	2.1	3.8	5.1	1.7	-0.7	4.5	5.5
230*	0.5	0.6	-1.4	-2.4	0.4	-0.4	-0.8	0.5
239	0.4	0.7	2.1	0.8	0.4	1.1	2.7	2.8
243	1.2	1.1	2.1	4.1	0.5	0.0	2.8	2.6
261	2.4	2.2	2.8	4.9	2.9	3.1	8.1	5.6
262	-0.4	0.8	-1.1	-3.8	-1.5	-4.0	-3.7	-14.0
270*	-1.0	-0.7	-1.8	1.6	-1.1	-1.1	-0.8	-0.6
275	2.1	-0.1	0.8	0.8	-0.2	-0.2	2.1	2.5
300	2.2	1.5	2.2	1.5	1.1	1.0	3.5	3.6
307	-1.3	-0.8	1.4	-0.9	-1.0	-0.9	-0.7	-7.6
323*	1.2	-1.0	-1.6	-1.2	-1.9	-2.7	-3.3	-14.0
325	-0.3	-0.2	0.7	-2.6	-1.7	-1.2	-1.4	-10.0
336	1.0	0.2	-0.5	-2.6	0.3	0.4	0.1	1.1
349	-0.3	-4.1	-1.8	-3.8	-0.2	-1.4	-0.7	-0.4
355*	-2.9	0.4	1.1	0.1	-3.9	-3.6	-9.0	-14.4
357*	-2.2	0.8	0.3	0.7	-0.1	-0.5	0.1	-0.1
374	-1.7	-1.7	1.0	0.7	0.4	-0.5	-0.5	4.2
377	1.2	1.7	0.3	1.0	0.3	-0.1	-0.4	0.1
393	1.7	2.6	1.6	4.1	0.9	-0.6	0.7	1.0
398	-0.9	-0.4	-2.9	-3.9	-3.0	-4.0	-2.9	-4.1
412	0.7	0.3	-0.8	-1.3	-0.2	-0.8	-0.2	0.1
418	1.8	1.1	1.2	n.d.	1.6	-0.9	6.8	8.9
438	0.8	1.6	-0.4	2.3	0.3	-2.0	-2.9	1.0
442	1.0	1.2	1.8	n.d.	-0.5	-1.0	-0.2	-4.1
451	-0.4	0.1	-0.4	2.8	0.6	1.1	2.2	3.8
452*	1.7	0.2	-0.3	-0.3	-0.4	-0.2	0.5	-4.4
454	-1.6	-2.5	-2.1	n.d.	-3.1	-0.9	-3.7	-4.0
458	0.2	0.4	-1.4	n.d.	-0.3	-2.6	-0.2	1.3
472	0.6	2.8	1.3	-1.7	0.0	-0.7	1.5	-7.8
479	-1.9	0.4	-1.0	-2.3	-2.1	-3.2	-11.7	-9.2
494*	1.2	0.2	0.6	-1.2	1.8	0.4	1.8	3.2
496	-1.3	1.8	1.0	-2.1	0.2	0.5	0.5	1.5
551	1.3	1.2	2.0	0.3	0.6	0.5	2.3	1.0
555*	0.4	0.2	-1.6	-3.4	-0.4	-1.7	-0.7	-2.3
617	2.2	1.6	1.2	-0.4	1.6	1.3	4.0	2.1
766	2.7	0.9	1.4	-0.1	0.8	-0.6	2.0	2.0
811	1.9	2.5	3.2	-1.4	1.1	-0.2	1.1	1.6
823	3.2	3.4	2.0	1.3	0.6	1.5	12.7	8.8
1236	1.1	0.6	-2.0	-0.1	-2.4	-4.7	-10.0	-6.1

Danksagung

Mein herzlicher Dank gilt allen, die zur Entstehung dieser Arbeit beigetragen haben.

Vielen Dank an meinen Doktorvater Herrn Prof. Dr. Gerhard Wenzel für die freundliche Aufnahme am Lehrstuhl für Pflanzenzüchtung und die Überlassung dieses spannenden Themas.

Ebenso gilt mein Dank Frau Prof. Dr. Chris-Carolin Schön für die Möglichkeit, diese Arbeit ordentlich beenden zu können.

Ganz besonders möchte ich mich bei Herrn Dr. Volker Mohler für seine ständige Diskussionsbereitschaft, die vielen konstruktiven Vorschläge, die gründliche Durchsicht meiner Arbeit und ganz allgemein für seine unverkrampfte, aufmunternde Sicht der Dinge bedanken.

Für die Hilfe in Labor und Gewächshaus, sei es praktischer Art bei der DNA-Extraktion und Probenahme oder durch Weitergabe ihres Erfahrungsschatzes, sei Frau Christina Hartmann, Herrn Stefan Schwertfirm, Frau Amalie Fiedler, Frau Anna Maria Schwaiger, Frau Xia Dong und Frau Dr. Anna Uzarowska vielmals gedankt. Außerdem durfte ich von der langjährigen Bambara-Praxis von Herrn Dr. Klaus Fleißner und der Hilfe von Herrn Dr. Peter Westermeier bei zahlreichen technischen und EDV-Angelegenheiten profitieren – vielen Dank dafür.

Den Lehrstuhlsekretärinnen Frau Vera Bork, Frau Petra Rometsch und Frau Waltraud Deragisch bin ich für ihre Unterstützung bei vielen administrativen Aufgaben dankbar.

Besten Dank an alle Dürnaster – insbesondere Herrn Dr. Bernhard Hauser, Herrn Richard Dinkel, Herrn Florian Steinbacher und Frau Sabine Parsy – die mir bei der Kultur und physiologischen Betrachtung der oftmals widerspenstigen Erdnuss mit Rat und Tat zur Seite standen.

Ein Dankeschön geht auch an Frau Dr. Sabine von Tucher (Pflanzenernährung) und Frau Dr. Maren Olbrich (Helmholtz Zentrum), die mich Analysen außer Haus durchführen ließen.

Ganz maßgeblich am Gelingen dieser Arbeit waren zwei Unterauftragnehmer beteiligt. Herr Dr. Fritz Thümmeler (vertis Biotechnologie AG) gab den Anstoß zur 454-Sequenzierung, auf der vieles aufbaute, erstellte die dazu notwendigen cDNAs und gab mir wertvolle Tipps. Dafür möchte ich mich herzlich bedanken.

Likewise, very special thanks go to Dr. Andrzej Kilian and his team at Diversity Arrays Technology Pty Ltd for their affable cooperation and the great efforts and passion put into the development of the DArT chips.

I would also like to acknowledge my project partners. At Nottingham, Prof. Sayed Azam-Ali and Dr. Rakhi Basu provided a lot of support concerning various aspects of the BAMLINK project. Dr. Sean Mayes became like a supervisor to me and willingly answered my numerous questions at any time. I think we were able to solve many problems together, which I wouldn't have managed alone. Thank you very much indeed. Many thanks, too, to Mr. Søren Thorndal Jørgensen (Copenhagen), Mr. Ozie Molosiwa (Nottingham), Dr. Martha Kandawa-Schulz (Namibia) and Dr. Dominique Dumet (IITA) and for the provision of seeds or DNA samples and helpful discussions.

Last but not least möchte ich mich bei meiner Familie und meinen Freunden bedanken, die Verständnis für meine Hochs und Tiefs während der letzten drei Jahre hatten und mich aufgebaut und an mich geglaubt haben.

Lebenslauf

Persönliche Daten

Name: Florian Stadler
Geburtsdatum: 7. August 1979
Geburtsort: Ebersberg
Familienstand: ledig

Schulbildung

1986-1990 Grundschule Anzing
1990-1999 Franz-Marc-Gymnasium Markt Schwaben

Studium

2000-2001 Studium der Landschaftsökologie an der Westfälischen Wilhelms-Universität Münster
2001-2006 Studium der Gartenbauwissenschaften an der Technischen Universität München, Wissenschaftszentrum Weihenstephan
Diplomarbeit am Fachgebiet Obstbau
Abschluss als Dipl.-Ing. agr.

Zivildienst

1999-2000 GSF – Forschungszentrum für Umwelt und Gesundheit, Neuherberg,
Institut für Ökologische Chemie

Berufliche Tätigkeit

2006-2009 Wissenschaftlicher Mitarbeiter an der Technischen Universität München,
Lehrstuhl für Pflanzenzüchtung