

Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes

Y. Lokko · J. V. Anderson · S. Rudd · A. Raji ·
D. Horvath · M. A. Mikel · R. Kim · L. Liu ·
A. Hernandez · A. G. O. Dixon · I. L. Ingelbrecht

Received: 11 February 2007 / Revised: 3 May 2007 / Accepted: 4 May 2007 / Published online: 31 May 2007
© Springer-Verlag 2007

Abstract Cassava (*Manihot esculenta* Crantz) is a staple food for over 600 million people in the tropics and subtropics and is increasingly used as an industrial crop for starch production. Cassava has a high growth rate under optimal conditions but also performs well in drought-prone areas and on marginal soils. To increase the tools for understanding and manipulating drought tolerance in cas-

sava, we generated expressed sequence tags (ESTs) from normalized cDNA libraries prepared from dehydration-stressed and control well-watered tissues. Analysis of a total of 18,166 ESTs resulted in the identification of 8,577 unique gene clusters (5,383 singletons and 3,194 clusters). Functional categories could be assigned to 63% of the unigenes, while another ~11% were homologous to hypothetical genes with unclear functions. The remaining ~26% were not significantly homologous to sequences in public databases suggesting that some may be novel and putatively specific to cassava. The dehydration-stressed library uncovered numerous ESTs with recognized roles in drought-responses, including those that encode late-embryogenesis-abundant proteins thought to confer osmoprotective functions during water stress, transcription factors, heat-shock proteins as well as proteins involved in signal transduction and oxidative stress. The unigene clusters were screened for short tandem repeats for further development as microsatellite markers. A total of 592 clusters contained 646 repeats, representing 3.3% of the ESTs queried. The ESTs presented here are the first dehydration stress transcriptome of cassava and can be utilized for the development of microarrays and gene-derived molecular markers to further dissect the molecular basis of drought tolerance in cassava.

Communicated by R. Reski.

Y. Lokko · A. Raji · A. G. O. Dixon · I. L. Ingelbrecht (✉)
Central Biotechnology Laboratory, International
Institute of Tropical Agriculture (IITA),
Oyo Road, Ibadan, Nigeria
e-mail: i.ingelbrecht@cgiar.org

J. V. Anderson · D. Horvath
Biosciences Research Laboratory, USDA-Agricultural Research
Service, 1605 Albrecht Blvd., Fargo, ND 58105-5674, USA

S. Rudd
Centre for Biotechnology, Tykistökatu 6, 20521 Turku, Finland

M. A. Mikel
Department of Crop Sciences and Roy J. Carver Biotechnology
Center, 2610 Institute for Genomic Biology,
University of Illinois, 1206 W. Gregory Drive,
Urbana, IL 61801, USA

R. Kim · L. Liu · A. Hernandez
W.M. Keck Center for Comparative and Functional Genomics,
Edward R. Madigan Laboratory, University of Illinois,
1201 W. Gregory Drive, Urbana, IL 61801, USA

Present Address:

Y. Lokko
Plant Breeding and Genetics Section, International Atomic
Energy Agency, Wagramer Strasse 5, PO Box 200,
1400 Vienna, Austria

Introduction

Drought is one of the major abiotic stresses limiting crop productivity worldwide. As water resources and arable land become limiting, the development of drought tolerant crops and use of marginal lands for agriculture will become increasingly important. Cassava (*Manihot esculenta* Crantz), also known as manioc, is an important food crop

in the tropics and subtropics where it is mainly grown for its starchy tuberous roots. Cassava ranks fourth as the most important source of food globally, after rice, sugarcane, and maize, and is also utilized as animal feed and for the industrial production of non-grain starch and ethanol (Balagopalan 2002). Cassava can be cultivated under conditions considered marginal for most other crops, including low fertility soils and areas that face sporadic or seasonal drought (Sakai et al. 1994).

In plants, drought triggers a wide variety of responses including alterations in gene expression, the synthesis of specific proteins (e.g., proteins that scavenge oxygen radicals, chaperone proteins, etc.), and the accumulation of metabolites or osmotically active compounds. The phytohormone abscisic acid (ABA) plays a key role in mediating responses to abiotic stress and promotes characteristic developmental changes that help plants cope with water deficit, such as the restriction of shoot growth and leaf area expansion (Lecoeur et al. 1995), the stimulation of root extension (Sharp et al. 1994), the accumulation of osmotically active solutes (Larosa et al. 1987), and the closing of stomata. Physiologically, tolerance to drought is a complex phenomenon involving drought escape, dehydration avoidance, dehydration tolerance, and desiccation tolerance mechanisms (Blum 1998). Genetically, tolerance to dehydration stress is a multifactorial trait, which makes breeding for drought tolerance arduous. Breeding plant varieties for sustainable production under moisture stress is also challenging because field trial climatic factors, such as temperature and drought, are often unpredictable (Blum 1998). Low fertility and the highly heterozygous nature of cassava render traditional breeding approaches difficult and make it a prime candidate for genomics-assisted breeding or genetic engineering.

In the past decade, excellent progress has been made in unraveling abiotic stress pathways at the molecular level in plants. This knowledge has been applied to the production of plants tolerant to drought, cold, and salt through genetic transformation (Zhang et al. 2004; Umezawa et al. 2006). Considerable progress has also been made in the genetic mapping of abiotic stress traits in plants, especially the cereals, for marker- or genomics-assisted breeding. For example, various genetic effects regulating drought, salt, and cold stress have been assembled on a single chromosome map in the *Triticeae* (Cattivelli et al. 2002).

Cassava can withstand short dry spells as well as prolonged periods of drought up to 4–6 months. Cassava responds to drought episodes mainly through dehydration avoidance by rapidly closing stomata to reduce transpiration and maintain high water potential (El Sharkawy 2004). The extent of osmotic adjustment and accumulation

of osmoprotective proteins such as dehydrins, appears small (Alves and Setter 2004a, b). Leaf growth and photosynthesis are decreased to near zero during episodes of water deficit but recover rapidly after rainfall. Thus, cassava responds to drought by arresting photosynthesis and growth, which are resumed after the recovery of water status. This behavior is appropriate for environments that face periodic water shortages or prolonged drought and contrasts with that of other drought tolerant plant species, such as sorghum, which respond to drought by partially closing stomata while adjusting osmotically and maintaining modest rates of photosynthesis. Drought avoidance mechanisms in cassava include root characteristics, such as early bulking and deep rooting (El Sharkawy 2004). In a comparative study of nine improved cassava varieties, Okogbenin et al. (2003) found that the adaptation response to drought stress was significantly influenced by genotype, suggesting a strong genetic basis for drought tolerance in cassava.

One approach for analyzing changes in gene expression under stress conditions is to compare expressed sequence tags (ESTs) from normal and stressed tissues. EST sequences provide a robust approximation of the expressed gene content of the parental genome under given sampling conditions (Satou et al. 2003; Alba et al. 2004). For example, several genes responsive to dehydration stress were identified from ESTs generated from ABA-treated and desiccated moss cDNA libraries (Machuka et al. 1999; Wood et al. 1999).

To reduce redundancies in EST collections and enrich for low-abundance transcripts, a cDNA library normalization approach can be followed (Bonaldo et al. 1996). Indeed, ESTs generated from normalized cDNA libraries have been used for the discovery of novel genes and have provided comprehensive analyses of genes expressed in *Arabidopsis* (Asamizu et al. 2000a), *Lotus japonicus* (Asamizu et al. 2000b), and *Triticum aestivum* (Ali et al. 2000).

Expressed sequence tags can be used for the production of microarrays, which are rapidly becoming a preferred tool to identify and dissect complex genetic networks that underlie physiological and developmental processes (Richmond and Somerville 2000). Gene-derived molecular markers are an inexpensive byproduct of EST datasets (Kota et al. 2003; Thiel et al. 2003).

Although drought tolerance in cassava is genotype-dependent, the genetic basis underlying drought tolerance in cassava is unknown at present. Also, genomics or molecular approaches to investigate the response of cassava to drought have not been reported. We intended to develop a genomics resource for cassava as a tool for gene discovery, genetic studies and expression profiling, specifically targeting dehydration stress.

Materials and methods

Plant material and growth conditions

In vitro plantlets of the cassava accession TME117 (local name *Isunikankiyan*, Nigeria) were derived from meristem cultures at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, and shipped to the USDA Biosciences Research Laboratory in Fargo, ND, USA. Plants were propagated and maintained as previously described (Anderson et al. 2004). One year after planting, dehydration stress on the plants was initiated by withholding water for a period of 6–14 days. Control plants were regularly watered. Leaf, stem, petiole, and meristem tissue was obtained from plants grown in 11" wide × 9.5" tall pots filled with a mixture of one part sunshine mix #1 and 2 parts sandy loam. Collection of leaf, stem, petiole, and meristem tissue subjected to dehydration stress occurred 6 days after withholding water, at which point the mature leaves showed visible wilting. Storage root with peel, and root tissue was obtained from plants grown in 22.75" wide × 18.25" tall pots filled with a mixture of one part sunshine mix #1 and 2 parts sandy loam. Collection of storage root with peel, and root tissue occurred 14 days after withholding water, at which point the mature leaves showed visible wilting. Tissues were collected from plants grown in three separate pots. All tissue samples were individually ground in liquid N₂ prior to storage at –80°C.

cDNA library construction and normalization

Total RNA was extracted as described by Anderson and Horvath (2001) and then pooled in equal amounts before isolation of mRNA. Poly(A)⁺mRNA was isolated twice from total RNA of each tissue sample using the Oligotex Direct mRNA kit (Qiagen, Valencia, CA, USA). Reverse transcription of mRNA into double-stranded cDNA was accomplished using the SuperScript™ Choice System modified with *NotI*/oligo(dT)₁₈ primers with an identifying 5-bp tag embedded between the *NotI* cloning site and the oligo(dT)₁₈. The modified oligo(dT)₁₈ primers were as follows: (*NotI*)TCCGA(dT)₁₈ for the control mRNA and (*NotI*)TCGCA(dT)₁₈ for the dehydration-stressed mRNA. *EcoRI* adapters were ligated onto the blunt-ended, double-stranded cDNAs (more than 450 bp). After digestion with *NotI*, the cDNAs were directionally cloned into *EcoRI*–*NotI* digested pBSII SK(+) phagemid vector (Stratagene, La Jolla, CA, USA). Purified plasmid DNA from the primary libraries was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cloned cDNA population, were used as a driver for normalization.

Hybridization between the single-stranded library and the PCR products was carried out for 44 h at 30°C. Unhybridized single-stranded DNA circles were separated from hybridized DNA, rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library.

Nucleotide sequencing and sequence data analysis

Randomly selected clones were partially sequenced from the 5' end on an ABI 3730. Base calling was done using PhredPhrap (Ewing et al. 1998). Assessment of high quality sequence and calculation of plate-wise success rate was done using QualTrim (bases with score 20 and above were considered high quality). Vector sequences and bases having Phred scores lower than 20 were removed. Clean sequences with a length of 200 or more bases after trimming were masked of repeat and low complexity sequences using RepeatMasker. The final clean sequences were then screened for unwanted sequences such as *Escherichia coli* genome, vector, mitochondrial DNA, ribosomal RNA, and viral DNA using BLASTN. All EST data are publicly available through the National Center for Biotechnology Information (NCBI, Bethesda, MD, USA; GenBank dbEST accession nos. DV440840—DV459005).

Annotation of filtered EST sequences

Expressed sequence tag sequence collections were annotated within the Sputnik framework for comparative plant genomics as described (Rudd 2005). Briefly, ESTs were clustered using the Hashed Position Tree 2 algorithm (Biomax Informatics, Martinsried, Germany) and the clusters assembled into unigenes using the CAP3 algorithm. Unigene sequences were assigned a unique id. Peptide sequences were derived for all unigenes using the ESTScan application (Iseli et al. 1999). Prior to ESTScan prediction, a *M. esculenta* species-specific ESTScan model was created by training with open reading frames identified through BLASTX of the unigenes against the SWISS-PROT database with the results filtered using the expectation value of 1e-10. The sequences were further annotated for structural and functional attributes using InterPro domains (Mulder et al. 2003), and the complete sequence collections were summarized using the MIPS catalogue of functionally annotated proteins (Funcat) and Gene Ontology terms (Mewes et al. 2002; Harris et al. 2004).

Statistical analysis

The data for Figs. 2 and 3 was analyzed using a *t*-test to determine whether there were significant differences in terms of number of ESTs in the various functional

categories for the dehydration stress treatment versus control, well-watered treatment. The Statistical Analysis System Package (SAS Institute Version 9.1 2004) was used.

Marker prediction

Each unigene EST set was searched for simple sequence repeat (SSR) markers as described (Rudd et al. 2005). SSR markers were identified by scoring for repeats. If repeats at the same locus and of different length were found, the candidate SSR was labeled as a probable SSR.

Results

Characterization of the cDNA libraries and EST sequences

Two normalized cDNA libraries were constructed using mRNA isolated from dehydration-stressed and control well-watered cassava tissues. The control library contained 5×10^5 recombinant colony forming units (cfu) and the dehydration-stressed library contained 1×10^6 cfu. Inserts ranged from 0.5 to 2.5 kb. In total, 8,956 randomly collected, high quality ESTs were generated from the normalized control cDNA library and 9,210 from the normalized dehydration-stressed library, yielding a total of 18,166 ESTs with an average read length of 586 nucleotides (Table 1). These 18,166 cDNAs produced a total of 8,577 unigenes, which consisted of 5,383 singletons and 3,194 multi-member clusters with an average size of 4.05.

The unigene sequences were annotated for structural and functional characteristics using a selection of bioinformatics tools that are relevant to comparative genomics and biological understanding as described (Rudd 2005). Approximately 63% of the unigenes were assigned func-

tions (Fig. 1). The largest category (25.71%) contained EST sequences with no similarities to previously sequenced genes, which indicates the presence of putative novel genes that are reported here for the first time and may be specific to cassava. This was followed by sequences with unclear function (11.24%). Most of the identified transcripts appeared to be from genes related to metabolism (10.19%) and cellular organization (8.69%).

A comparison between the functional categories of the dehydration-stressed and the control library is shown (Fig. 2). Overall, the numbers of unigenes in the different functional groups were similar for the two libraries. A further breakdown of the ESTs in the ‘cell rescue, defense, cell death, and ageing’ category is presented (Fig. 3). A *t*-test was performed on the data in Figs. 2 and 3 and showed that no significant differences were found in terms of number of ESTs in the different functional categories between the two treatments ($t = 0.09$ for the data in Fig. 2 and $t = 0.22$ for the data in Fig. 3).

Identification of putative dehydration stress-responsive genes unique to dehydration-stressed tissues

The EST profiles of the transcriptome from the dehydration and control libraries may provide clues for the identification of dehydration stress-responsive genes. To assess whether the dehydration stress treatment was successful in enrichment of drought responsive genes, EST that were unique to the dehydration stressed library were identified in silico for the functional category ‘cell rescue, defense, cell death, and ageing’. These are listed in Table 2. The type member of the EST cluster is indicated in this table as well as the EST copy number within each cluster. As shown, the dehydration-stressed library contained numerous unique ESTs that encode proteins with recognized roles in drought responses. Examples include the precursor to Early Responsive to Dehydration ERD1 (Kiyosue et al. 1994), the cysteine proteinase Response to Dehydration RD19 (Yamagushi-Shinozaki and Shinozaki 1993), dehydration-responsive protein RD22 precursor, drought-induced protein Di19-like (Gosti et al. 1995) and various high and low molecular weight heat-shock proteins (HSPs) (Vierling 1991). Transcription factors that were unique to dehydration-stressed tissues included a heat-stress transcription factor A3, heat-shock transcription factor 21, and a homeobox-leucine zipper protein ATHB12, known to mediate growth response to water deficit (Olsson et al. 2004). Molecular chaperonins unique to the dehydration-stressed tissue were DnaK- and DnaJ-like proteins. DnaJ protein is involved as a cochaperonin in the function of HSP70s in protein folding and stabilization (Netzer and Hartl 1998). Transcripts implicated in oxidative stress responses and/or protection of cellular membranes that were

Table 1 Summary report for cassava ESTs

Feature	Statistic
Total number of clean ESTs	18,166
EST total nucleotides (nt)	10,747,309
Cluster count	8,577
Singleton unigenes	5,383
Multi-member unigenes	3,194
Average EST length (nt)	586
Average unigene length (nt)	681
Average cluster size (singletons excluded)	4.05
Average cluster size (singletons included)	2.13
Protein coding potential (%)	77.53
Peptide total amino acids	1,510,481

Fig. 1 Functional classification of 18,166 cassava ESTs from the combined dehydration-stressed and control cDNA libraries

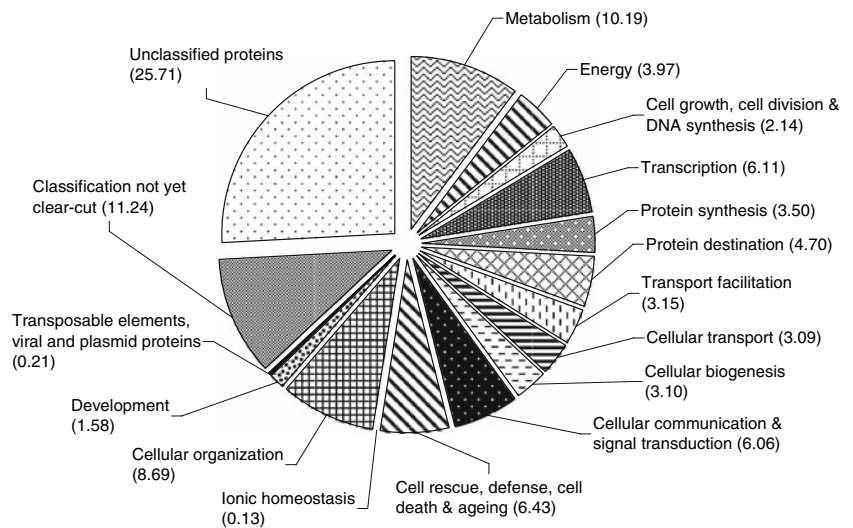


Fig. 2 Comparison of functional classification between unigene sequences of the dehydration-stressed and control cDNA library

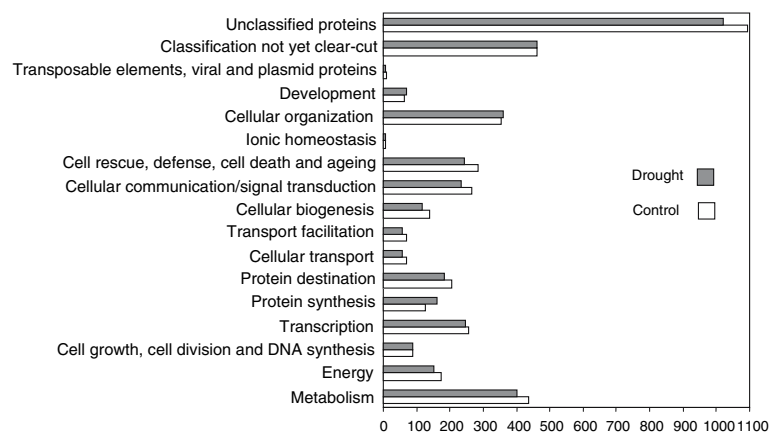
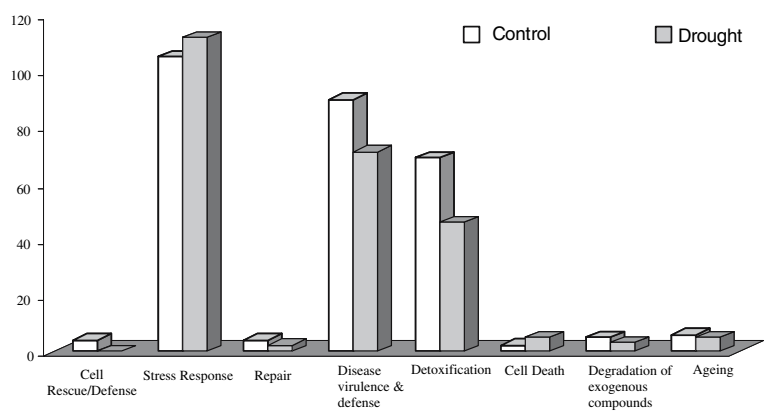


Fig. 3 Comparison of unigene sequences in the ‘cell rescue, defense, death, and ageing’ functional category for the dehydration-stressed and control cDNA library



unique to the dehydration-stressed library included monodehydroascorbate reductase and phospholipid hydroperoxide glutathione peroxidase. Other dehydration-stressed unique transcripts included phospholipase D and phosphoinositide-specific phospholipase C (PI-PLC). Phospholipid signaling has recently been implicated in the responses of plants to various environmental cues, both

biotic and abiotic (Meijer and Munnik 2003). Phospholipase D catalyzes the hydrolysis of structural phospholipids and produces phosphatidic acid, which acts as a second messenger in signal transduction pathways. PI-PLC is involved in the synthesis of inositol 1,4,5-triphosphate, which stimulates the release of Ca^{2+} from intracellular stores and as such is also involved in signal transduction. In

Table 2 The putative identity of ESTs unique to the dehydration-stressed library

Putative identity	EST cluster id	Number of ESTs in cluster	Accession number ^a
<i>Nicotiana tabacum</i> lesion-inducing like	CV02010A2D10.f1	7	Q8LBI4
ATP-dependent Clp protease ATP-binding subunit clpA homolog, chloroplast	CV02009B2C02.f1	5	P35100
DnaJ-like protein	CV02017A1H05.f1	4	Q9SDN0
Unknown protein (At5g25170)	CV02028A1B08.f1	3	Q6NQ19
DNA-binding protein	CV02007B2A12.f1	3	Q43427
Ubiquitin/ribosomal protein CEP52	CV02009A2F04.f1	3	S28420
Homeobox-leucine zipper protein ATHB-12	CV02010B1D10.f1	3	Q9M276
Germin-like protein subfamily 3 member 1 precursor (AtGLP1)	CV02018A1H09.f1	3	P94040
Putative monodehydroascorbate reductase, cytoplasmic isoform 2	CV02020A1G02.f1	3	Q9LK94
Homeobox-leucine zipper protein ATHB-12	CV02023B2F06.f1	3	Q9M276
DnaK-type molecular chaperone hsc70-3 <i>Lycopersicon esculentum</i>	CV02002X1B10.f1	2	Q40151
Heat-shock protein (Hsp70)	CV02003B2F02.f1	2	Q84QJ3
Stress-induced protein STI1-like protein	CV02004A1H05.f1	2	Q9STH1
Homeotic protein HAHB-1	CV02005B2F04.f1	2	Q39941
Class-1 LMW heat shock protein	CV02008A2G07.f1	2	Q8H288
Calcineurin B-like protein	CV02010B2E10.f1	2	Q7FRS8
Heat-shock protein HSP23.6 mitochondrial precursor	CV02011B1F11.f1	2	Q96331
Phospholipid hydroperoxide glutathione peroxidase	CV02013B1C03.f1	2	Q8W259
Dehydration-responsive protein RD22 precursor	CV02017A1H03.f1	2	Q08298
17.3 kDa class I heat shock protein (HSP 17.3)	CV02018A1C06.f1	2	P02519
RuBisCO subunit binding-protein alpha subunit (60 kDa chaperonin alpha subunit) (CPN-60 alpha)	CV02024B2G12.f1	2	P08824
Probable monodehydroascorbate reductase, cytoplasmic isoform 2	CV02027B1D05.f1	2	Q9LK94
Cold acclimation protein	CV02028B1F01.f1	2	Q9FUW4
<i>Nicotiana tabacum</i> ORF	CV02004A1G03.f1	1	P93384
Hypothetical protein AT4g25680	CV02013B1E08.f1	1	Q9SZZ6
Germin-like protein subfamily 3 member 2 precursor	CV02001X1B05.f1	1	Q9SR72
Blue copper-binding protein F9D12.16	CV02003B1C11.f1	1	O81500
Heat shock (ATP-binding chain SB100)	CV02003B1H12.f1	1	Q39889
Blue copper-binding protein, 19K	CV02003B2F11.f1	1	Q39131
Heat shock cognate protein	CV02004A2H10.f1	1	O22329
Expressed protein (DNAJ protein-like-At2g35795)	CV02006A1C07.f1	1	Q8RV04
Heat stress transcription factor A3	CV02007B2B06.f1	1	Q9LLF2
AT4g28240/F26K10_120 (Putative wound-induced protein)	CV02008B2E03.f1	1	Q94A22
Hypothetical protein At1g14890 F10B6.30	CV02009A1E10.f1	1	Q9LQU1
DnaJ-like protein	CV02009A2D06.f1	1	Q9M2L3
Salt-inducible protein homolog	CV02010A2E06.f1	1	O23484
Cysteine proteinase RD19	CV02010B1D02.f1	1	P43296
Polyubiquitin 4 UBQ4	CV02010B2F06.f1	1	P59263
Disease resistance RPP5 like protein	CV02011B2F12.f1	1	O23530
Phospholipase D (AtPLDdelta)	CV02013B2D04.f1	1	Q9C5Y0
Phosphoinositide-specific phospholipase C-line (MZN1.13)	CV02014A2A12.f1	1	Q944C2
RuBisCO subunit binding-protein, chloroplast precursor (60 kDa chaperonin beta subunit) (CPN-60 beta)	CV02014B1A05.f1	1	P21241

Table 2 continued

Putative identity	EST cluster id	Number of ESTs in cluster	Accession number ^a
Ripening-related protein-like (DC1.2 homologue)	CV02014B1H07.f1	1	Q9LVA3
Heat shock protein 81-2 (HSP81-2)	CV02015A2G03.f1	1	P55737
Heat shock cognate 70 kDa protein 1	CV02015A2G05.f1	1	P22953
Centrin	CV02015B1A04.f1	1	O82659
Germin-like protein (GLP3b)	CV02015B1D04.f1	1	P94072
Hypothetical protein (Induced upon wounding stress) (At4g24220)	CV02016B1A12.f1	1	Q9STX2
Monodehydroascorbate reductase, cytoplasmic isoform 3 (NADH)-like protein	CV02017A1F06.f1	1	Q9LFA3
Regulator of chromosome condensation like (Cell cycle regulatory protein)	CV02017B1F08.f1	1	Q8GYL6
Major latex-like protein	CV02017B2C12.f1	1	Q93VR4
Drought-induced protein Di19-like protein	CV02018A1G11.f1	1	Q9FJ17
DbjIBAA90612.1 Hypothetical protein At5g62390 (KED-like protein)	CV02018B1H03.f1	1	Q9LVA0
Erd1 protein precursor	CV02019A1F02.f1	1	P42762
Cysteine proteinase-like protein	CV02019B1C04.f1	1	Q9SUL1
Shock protein SRC2-like	CV02020A1G10.f1	1	Q9LZI7
UV-damaged DNA binding factor-like protein	CV02021A2F07.f1	1	Q9M0V3
KED-like protein	CV02022A2B11.f1	1	Q9LVH1
Major latex protein homolog-like	CV02023A1H04.f1	1	Q93VR4
Stellacyanin (uclacyanin 3)-like predicted GPI-anchored protein	CV02023A1H06.f1	1	Q9LY37
Heat-shock transcription factor 21 (AtHSF21)	CV02025A1B05.f1	1	O49403
Heat-shock protein 81-2 (HSP81-2)	CV02025A1D03.f1	1	P55737
Heat-shock cognate 70 kDa protein 2 (Hsc70.2)	CV02026A1A02.f1	1	P22954

^a GenBank accession number of the most similar sequence as identified by BlastX alignment

Arabidopsis, a transcript encoding PI-PLC was significantly induced under various environmental stresses, such as dehydration, salinity, and low temperature (Hirayama et al. 1995). A calcineurin B-like (CBL) protein was also detected in our dehydration-stressed library (Kudla et al. 1999). CBL proteins are Ca²⁺ sensors that regulate the activity of a specific group of kinases. In Arabidopsis, CBL1 is a critical calcium sensor in abiotic stress responses (Cheong et al. 2003). Overall, at least 28 of the 63 cassava EST clusters in Table 2 showed significant homology to known drought responsive genes in other plant species, illustrating that the dehydration stress treatment was effective for enriching ESTs involved in drought responses.

In addition, several transcripts induced by other environmental stresses or by biotic stresses were also unique to the dehydration library. These included ESTs encoding a salt-inducible protein homolog, a cold acclimation protein, a stress-induced STI1-like protein, various blue copper-binding (BCB) and germin-like proteins (GLP) (Zimmerman et al. 2006), and the disease resistance RPP5-like protein.

Digital Northern

Gorantla et al. (2007) recently examined expression profiles of highly expressed genes through digital Northern of a normalized EST library constructed from drought-stressed seedlings to identify putative stress-responsive genes in rice. Here, highly prevalent ESTs clusters with ten or more copies in the dehydration stressed library were considered to detect genes that are potentially involved in dehydration stress. This analysis was conducted for the functional category ‘cell rescue, defense, cell death, and ageing’ and Table 3 lists EST clusters that have at least ten copies in the dehydration stressed library. The EST copy number in the control library is also indicated. As shown, all highly prevalent ESTs in the dehydration stressed library are also represented in the control library. Metallothionein (MT)-like genes were the most abundant class in the combined normalized libraries (120 total hits for MT-1 and 116 total hits for MT-3). MT-like proteins are involved in metal detoxification and are efficient scavengers of free hydroxyl radicals (Palmiter 1998). In

Table 3 EST copy number of highly prevalent ESTs (≥ 10 copies) in dehydration-stressed library (St) compared to the control (Co)

Putative identity	EST cluster id	Number of ESTs in cluster			Accession number ^a
		St	Co	Total	
Metallothionein-like protein 1 (MT-1)	CV02028A2A07.f1	84	36	120	P20238
Hypothetical protein	CV02024B1B05.f1	68	25	93	Q05349
Chlorophyll A-B binding protein 40	CV02023A1E09.f1	66	18	84	P27495
Ribulose biphosphate carboxylase (RuBisCO small subunit)	CV02009B2C08.f1	51	43	94	Q42915
Photosystem II 22K protein precursor (CP22)	CV02018A1B07.f1	46	12	58	Q02060
Metallothionein-like protein type 3 (MT-3)	CV02028A2H07.f1	41	75	116	Q96386
Late-embryogenesis protein lea5	CV02007A2E03.f1	36	8	44	Q39644
Putative ADP-ribosylation	CV02007B2D10.f1	35	15	50	P51823
Glycine-rich RNA-binding protein	CV02007A2F07.f1	32	9	41	Q8VX74
Light regulated protein precursor	CV02005A1A06.f1	32	51	83	Q8S2S6
Ribulose biphosphate carboxylase small chain (RuBisCO small subunit)	CV01020A1B01.f1	31	22	53	Q42915
CP12 chloroplast protein precursor	CV02016B1F01.f1	31	25	56	O24136
Photosystem I reaction center subunit V, chloroplast precursor (PSI-G)	CV02026A2F08.f1	30	13	43	Q9S7N7
Photosystem II oxygen-evolving complex protein 3	CV02005B1C12.f1	29	8	37	Q9XFT3
Chlorophyll a-b binding protein 8, chloroplast precursor (LHCI type III CAB-8)	CV02019B2D03.f1	28	7	35	P27522
40S ribosomal protein S21	CV02014B2C08.f1	27	4	31	Q41852
Lipid transfer protein (LTP)	CV02020B2B08.f1	27	4	31	Q8RYA8
Hypothetical protein	CV02028A2G10.f1	25	14	39	Q84VZ7
Lipid transfer protein (LTP)	CV02025B2D03.f1	24	6	30	Q9M6B6
Translationally controlled tumor protein homolog (TCTP)	CV02023B1B01.f1	23	6	29	Q9ZSW9
60S acidic ribosomal protein P1	CV02025B2A02.f1	22	5	27	P52855
Oxygen-evolving enhancer protein II (OEE2)	CV01019B2B06.f2	21	6	27	Q9SLQ8
Chlorophyll a/b-binding protein CP26	CV02023A1D06.f1	21	7	28	Q39286
Metallothionein-like protein (MT-3)	CV01016B2F09.f1	20	59	79	Q96386
60S ribosomal protein L10a-1	CV02012A2C02.f1	17	2	19	Q8VZB9
Expressed protein (Hypothetical protein At2g36930)	CV02026A2C06.f1	17	3	20	Q9SJL4
Chlorophyll a-b binding protein 151, chloroplast precursor (LHCII type II CAB-151)	CV01020B1A10.f1	17	6	23	P27518
Chlorophyll a-b binding protein 6A, chloroplast precursor (LHCI type I CAB-6A)	CV02009B2H05.f1	17	9	26	P12360
Photosystem II 10 (PII10)	CV02018A2C10.f1	17	9	26	Q40519
DnaJ protein homolog (DNAJ-1)	CV02015A1F03.f1	15	6	21	Q04960
Photosystem I subunit O precursor	CV01009B1A12.f1	15	9	24	Q949Q5
Glutaredoxin	CV02005B2G07.f1	15	11	26	P55143
Lipid transfer precursor protein	CV02005B1H10.f1	14	1	15	Q8RYA8
40S ribosomal protein S28	CV02016A2G01.f1	14	2	16	P34789
Photosystem I chain XI precursor (PSI-L)	CV01020A1B04.f1	14	7	21	Q9SUI4
Plastocyanin, chloroplast precursor	CV01027A2A06.f1	14	8	22	P17340
Chlorophyll a/b-binding protein precursor (LHCII type I CAB-21)	CV02018B2C08.f1	14	8	22	P27493
Chlorophyll a/b-binding protein (LHCII type II CAB-151)	CV02017B2E04.f1	13	1	14	P27518
Photosystem II reaction center W	CV02013B2G11.f1	13	4	17	Q41387
Chlorophyll a/b binding protein CP29	CV02006B2F04.f1	13	5	18	Q9XQB2
Hypothetical 41.5 kDa protein in YCF6-CHLB intergenic region	CV01026B1A06.f1	13	6	19	P51277
Carbonic anhydrase, chloroplast precursor	CV02003A2A03.f1	13	6	19	P27141

Table 3 EST copy number of highly prevalent ESTs (≥ 10 copies) in dehydration-stressed library (St) compared to the control (Co)

Putative identity	EST cluster id	Number of ESTs in cluster			Accession number ^a
		St	Co	Total	
Ubiquitin-conjugating enzyme (Ubiquitin carrier protein 14)	CV02026A1H04.f1	13	7	20	S46656
Catalase	CV02017B1D06.f1	13	8	21	Q7XTK8
Putative glycogen synthase	CV02023B2E04.f1	13	8	21	Q9MAC8
Thiamin biosynthesis protein thi1-	CV01009A2C11.f1	13	11	24	O23787
Elongation factor 1B alpha-subunit.	CV02006B2B12.f1	12	1	13	Q9SCX3
Actin-depolymerizing factor 2 (ADF 2)	CV01028B1F03.f1	12	2	14	Q9FV11
Dehydration-responsive protein RD22 precursor (At5g25610)	CV01016B2A08.f1	12	3	15	Q08298.
Catalase isozyme 1	CV02027A2G03.f1	12	6	18	P17598
40S ribosomal protein S8	CV01015B2G11.f1	11	2	13	Q08069
Photosystem II protein (PSII-T)	CV02024A1F06.f1	11	3	14	P31336
Photosystem I psaH protein	CV02025B2H11.f1	11	6	17	Q7XA96
Putative binding precursor	CV01009A1H08.f1	11	7	18	Q9T0N9
Chlorophyll a-b binding protein, chloroplast precursor (LHCI type II CAB)	CV02016A2B12.f1	11	8	19	P13869
Metallothionein-like protein 1 (MT-1)	CV01023A2B06.f1	11	21	32	P20238
Nuclear transport factor 2	CV02026B2G05.f1	10	3	13	B86405
Hypothetical protein	CV01014A1A12.f1	10	4	14	Q9XI42
Triose phosphate/phosphate translocator, chloroplast precursor (CTPT)	CV01013B2D02.f1	10	5	15	P21727
Ribulose biphosphate carboxylase/oxygenase activase, chloroplast precursor (RuBisCO activase)	CV02019A1D07.f1	10	5	15	P10896
Translationally controlled tumor protein homolog	CV02005A1D04.f1	10	7	17	Q9ZSW9
Acetone-cyanhydrin lyase	CV02023A1B02.f1	10	7	17	S45682
Cytochrome B6-F complex iron-sulfur subunit 2, chloroplast precursor (Rieske iron-sulfur protein)	CV02014A2G05.f1	10	9	19	Q02585
Chlorophyll a/b-binding protein precursor (LHCI type II CAB)	CV02017B1E06.f1	10	11	21	P13869

^a GenBank accession number of the most similar sequence as identified by BlastX alignment

rice, MT-like proteins were found to be the most abundant EST class in drought-stressed leaf tissue (Reddy et al. 2002) but they were also abundant in non-stressed leaf tissue (Cho et al. 2004) which is in agreement with our observations. ESTs encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit and the chlorophyll a/b binding protein, key enzymes in photosynthesis, were also found in abundance. With 36 copies, the late-embryogenesis-abundant (LEA) group 5 LEA transcript was among the most abundant EST in the dehydration-stressed library. ESTs encoding a lipid-transfer protein (LTP) were also highly prevalent in the dehydration-stressed library with 27 copies. The LEA proteins are a group of proteins commonly involved in the enhancement of stress tolerance with suggested roles in binding water, protein or membrane stabilization, and in ion sequestration (Chaves and Oliveira 2004). LTPs are ubiquitous in higher plants. A number of functions have been proposed for plant LTPs, including involvement in cuticle biosynthesis, surface wax production, or adaptation to environmental

stresses (Kader 1997; Treviño and O'Connell 1998). ESTs encoding the Response to Dehydration protein, RD22 (Yamagushi-Shinozaki and Shinozaki 1993), was also highly abundant in the dehydration-stressed library.

In silico identification of microsatellite sequences

To obtain gene-derived markers for further genetic studies of drought tolerance and other important traits in cassava, the EST clusters were screened for SSRs. Stretches of di-, tri-, and tetrameric nucleotide repeats were identified using parameters that would detect dimeric motifs with seven or more repeats, trimeric motifs with six or more repeats, and tetra- or pentameric motifs with four or more repeats. Perfect and near perfect repeats with slight repeat pattern deviations were scored. This allows for the possibility that there may be a perfect repeat pattern at a locus within different varieties and cultivars (Rudd et al. 2005). A total of 646 potential microsatellite loci were identified in 592 unigene sequences, representing 3.3% of the total number

of ESTs queried. This figure is similar to the rate of microsatellite discovery in other species such as grape, sugarcane, and switchgrass where the frequency of EST-derived SSRs was between 2.5 and 3.8% (Scott et al. 2000; Cordeiro et al. 2001; Tobias et al. 2005). We detected 186 perfect dinucleotides, 131 perfect trinucleotides, 1 perfect tetranucleotide, and 2 perfect pentanucleotides. In addition, there were 264 imperfect dinucleotides, 57 imperfect trinucleotides, 4 imperfect tetranucleotides, and 1 imperfect pentanucleotide. The different classes of SSRs are summarized in Table 4. Dinucleotide repeats represented 70% of the total number of microsatellites. Of these (TC)_n and (AT)_n were the most common. Trinucleotide repeats were also detected with (GAA)_n repeats being the most common class. Only one (imperfect) (GC)_n repeat was returned and tetra- and pentanucleotide repeats were also uncommon.

Discussion

Cassava is remarkable in its ability to withstand brief drought spells as well as extended periods of seasonal drought of 4–6 months. As such, drought tolerance in cassava shows several characteristics that cannot be studied in traditional model plants. In addition, cassava performs well in marginal soils where other crops fail. Considering that drought, together with soil degradation, represents a major cause of yield reduction or even crop failure in non-temperate regions, the molecular characterization of drought tolerance in cassava has implications for many aspects of crop improvement and utilization in cassava and other crops.

In this study we have reported the characterization of 18,166 ESTs generated from two normalized cDNA libraries, one prepared from dehydration-stressed tissues yielding 9,210 ESTs and one from well-watered control tissues yielding 8,956 ESTs. Dehydration stress was applied for brief periods (6–12 days) and both root and leaf tissues were sampled. We identified 8,577 unique sequences from the 18,166 ESTs and this supports the theory that normalization reduces redundancy. The number of functional genes in plants has been estimated to range from 26,000 to nearly 50,000 (Goff et al. 2002). Thus, the 8,577 unique sequences identified in this study likely represent one-third to one-sixth of the potentially expressed genes in cassava. Anderson et al. (2004) have previously reported on the development of an EST resource for cassava. These ESTs were primarily obtained from cDNA libraries targeted for starch and biotic stresses (cassava bacterial blight and cassava mosaic disease). The EST resource described here targets dehydration stress and therefore complements the existing cassava EST resource. Overall, the ESTs developed in this study double the number of publicly

Table 4 Frequency of simple sequence repeats in the EST collection

Class	Repeat	Frequency perfect repeats	Frequency imperfect repeats	
Di-	TC	82	89	
	AT	67	88	
	AG	25	72	
	TG	10	10	
	GC	0	1	
	AC	2	4	
	Tri-	CAG	13	3
		CTC	3	1
		CCA	5	3
		CAA	1	0
CAT		9	6	
GGC		2	0	
GAT		7	8	
GAA		22	13	
GCC		3	1	
GCT		6	2	
AGG		4	1	
ATA		6	3	
TCT		23	7	
Tetra-	TGG	5	4	
	TTA	21	4	
	TTG	1	0	
	CCT	0	1	
	CTTT	1	1	
	AAAG	0	2	
	TTAT	0	1	
	Penta-	TTTTG	1	0
		AGAAA	1	1

Assembled unigene clusters were screened with openSputnik to identify repeats equal to or greater than 14 bp for dinucleotides and equal to or greater than 18 bp for the other repeats

available EST sequences for cassava bringing the total to 36,120 sequences (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html; dbEST release 011907).

Annotation of the unique transcripts resulted in 17 different functional categories that were reasonably comparable to the functional categorization of ESTs from other higher plants: ‘unclassified proteins’ comprised the highest number of ESTs (25.71%) and the category ‘metabolism’ contains the largest number of classified ESTs (10.19%) (Ramírez et al. 2005). The number of EST clusters present in the 17 functional categories in the dehydration-stressed and the control, well-watered tissues was not significantly different as shown by *t*-tests. The functional category ‘cell rescue, defense, cell death, and ageing’ was further broken down into its eight subcategories. The number of ESTs present in these categories for the drought and control libraries was also not significantly different.

Although the control and the dehydration-stressed libraries may contain a similar number of classified genes, this does not mean that the genes within the different functional categories are identical. To assess qualitative differences between the dehydration stress and control libraries and investigate whether the drought stress treatment was effective in enriching for dehydration-responsive genes, ESTs unique to the dehydration-stressed library were identified *in silico* in the ‘cell rescue, defense, cell death, and ageing’ category. This analysis uncovered numerous ESTs with recognized roles in drought stress in other plant species, including those encoding proteins involved in oxidative stress and/or protection of cellular membranes (monodehydroascorbate reductase, catalase, ascorbate peroxidase, and phospholipid hydroperoxide glutathione peroxidase), signal transduction (phospholipase D, PI-PLC, and CBL proteins), as well as transcription factors (heat stress transcription factor A3, heat shock transcription factor 21, and ATHB12). ESTs induced by other abiotic stresses than drought or by biotic stresses were also unique to the dehydration-stressed tissues. These included ESTs encoding proteins involved in metal homeostasis and tolerance (BCB proteins, MT proteins, and stellacyanin), cold (cold acclimation protein), salt (salt-inducible protein), and host plant resistance (GLP and the disease resistance RPP5-like protein). There is considerable overlap between various abiotic stress signaling pathways in plants, especially drought, cold, and salt, since all these stresses require protection against cellular dehydration (Knight and Knight 2001). It is therefore not surprising to find cold and salt-stress induced genes in the dehydration-stressed library. Previous studies have also revealed links between biotic and abiotic pathways. Chini et al. (2004) recently demonstrated that an activation-tagged allele of activated disease resistance 1, previously shown to convey broad spectrum disease resistance, conferred significant drought tolerance, indicating that there may be significant overlap between signaling network(s) that establish disease resistance and drought tolerance.

The EST libraries used in this study have relatively low redundancy because they were normalized, but still contain many more copies of some transcripts than others. EST clusters with ten or more members in the ‘cell rescue, defense, cell death, and ageing’ category have also been studied. This analysis also uncovered genes that have recognized roles in drought stress responses in other plant species. The LEA proteins deserve special mention in view of previous physiological studies on drought in cassava. ESTs encoding LEA proteins were among the most abundant ESTs in the dehydration-stressed tissues. LEA proteins are thought to have a role in binding water and some of the LEA proteins can be considered compatible solutes involved in osmotic adjustment (Ingram and Bartels 1996).

Physiological studies in cassava have shown that the main mechanism for drought resistance is drought avoidance through the production of ABA with little or no change in leaf water potential indicating minimal osmotic adjustment (El Sharkawy 2004). In agreement with this, dehydrins, a subgroup of LEA proteins, could not be detected in cassava tissues following brief episodes (6 days) of water deficiency (Alves and Setter 2004a). Our analysis shows that ESTs encoding LEA proteins are highly abundant in dehydration stressed tissues compared to control tissues. It is tempting to speculate that specific LEA proteins, distinct from dehydrins, play a role in osmotic adjustment in cassava. Recently, specific types of LEA proteins have been expressed in transgenic wheat and rice resulting in improved drought tolerance (Sivamani et al. 2000; Rohila et al. 2002).

The EST data and analysis presented here provide a first overview of the transcripts that are expressed in cassava under dehydration stress. The annotation and comparative analysis of these ESTs have identified numerous transcripts with recognized roles in dehydration stress, many of which were unique to the dehydration-stressed library. Overall, the data indicates that the dehydration treatment and normalization procedure has been effective in capturing drought-responsive transcripts. It should be emphasized that EST approaches are suitable for qualitative rather than quantitative comparisons (Rodriguez Milla et al. 2002; Ramírez et al. 2005). In addition, EST copy number in normalized EST libraries will be further diluted due to the normalization procedure, which reduces redundancy. A comprehensive catalogue of dehydration-responsive genes along with a more accurate quantitative assessment of transcript levels will require the use of a more refined research tool for expression profiling, such as DNA microarrays. DNA microarray studies will also facilitate allele mining by screening cassava varieties with different levels of drought tolerance.

Several EST clusters identified in this study encode similar proteins. It is possible that these ESTs clustered separately due to partial sequence information or due to limitations of the clustering and annotation protocols. Alternatively, it is possible that some EST clusters represent distinct members of multigene families that show extensive sequence identity. For example, CBL proteins are encoded by a family of at least six genes in *Arabidopsis* that are differentially regulated by stress signals (Kudla et al. 1999). Similarly, the (non-specific) LTPs are encoded by multigene families with individual genes having time- and tissue-specific expression patterns and induced under a variety of conditions (Lindorff-Larsen et al. 2001). Other examples include the MT proteins, encoded by multigene families that are differentially regulated in rice (Hsieh et al. 1995; Kawasaki et al. 2001), PLD proteins, LEA proteins,

GLP, and BCB proteins which are all encoded by multi-gene families in other plant species. Whether these EST clusters indeed represent distinct members of multigene families in cassava, needs to be verified by expression profiling or other techniques that can discern different members of multigene families.

The cassava genotype used in this study, TME117, is a landrace grown in humid and subhumid conditions. While this genotype is not specifically adapted to (semi)arid environments, it should be noted that the (sub)humid agroecologies also experience periodic dry spells and an extended dry season. Also, phenotypic differences between plant varieties can often be attributed to differences in gene regulation rather than to novel gene sequences *per se*. The fact that numerous drought-responsive genes have been uncovered, supports the hypothesis that this genotype can be utilized to identify genes involved in drought tolerance and opens up new avenues to search for (superior) allelic variants.

The EST sequence resources for cassava we have generated should provide readily available sources of genes that can be used to discover and develop functional molecular markers for germplasm characterization and marker assisted breeding of cassava and other species from the *Euphorbiaceae* such as castor bean. Popular markers that can be developed from ESTs are single nucleotide polymorphisms, conserved orthologous sets and SSRs. In this study, we have identified a total of 646 microsatellite repeats. Okogbenin et al. (2006) recently constructed an SSR linkage map for cassava with SSRs developed from genomic and cDNA libraries. These authors found that 40% of the SSRs derived from cDNA were polymorphic. Their findings suggest that the EST-SSRs described in this paper will significantly increase the number of SSRs available for molecular genetic studies in cassava.

In summary, the EST collection described here is the first reported dehydration stress transcriptome of cassava and has already uncovered a wealth of putative drought-responsive genes. This EST resource is also a rich source of gene-derived molecular markers. In the short term, these tools can be applied in genetic studies and for expression profiling to further dissect drought tolerance mechanisms in cassava. Appropriate protocols have been established for cassava, which allow the introduction of new traits via modern gene transfer techniques (Taylor et al. 2004). Likewise, framework linkage maps have been developed for cassava and utilized for genetic mapping of genes controlling agronomic traits (Fregene and Puonti-Kaerlas 2002). In view of recent progress in manipulating drought tolerance in model plants using biotechnology tools, a better understanding of drought tolerance in cassava at the molecular level will, in the long run, also facilitate the genetic improvement of cassava for drought tolerance using a combination of biotechnological and conventional methods.

Acknowledgments The authors thank Dr. R. Okechukwu for help with statistical analysis, Mrs. J. Plancarte for preparing figures and bibliography, Mrs. F. Kolade for technical assistance, and Mrs. R. Umelo for critical reading of the manuscript. This research was partially supported by the US Agency for International Development (USAID). The opinions expressed herein are those of the authors and do not necessarily reflect the views of the USAID.

References

- Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JKC, Martin G, Tanksley SD, Bouzayen M, Jahn MM, Giovannoni J (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* 39:697–714
- Ali S, Holloway B, Taylor WC (2000) Normalization of cereal endosperm EST libraries for structural and functional genomic analysis. *Plant Mol Biol Rep* 18:123–132
- Alves AA, Setter TL (2004a) Abscisic acid accumulation and osmotic adjustment in cassava under water deficit. *Environ Exp Bot* 51:259–271
- Alves AA, Setter TL (2004b) Response of cassava leaf area expansion to water deficit: cell proliferation, cell expansion and delayed development. *Ann Bot (Lond)* 94:605–613
- Anderson JV, Delseny M, Fregene MA, Jorge V, Mba C, Lopez C, Restrepo S, Soto M, Piegu B, Verdier V, Cooke R, Tohme J, Horvath DP (2004) An EST resource for cassava and other species of *Euphorbiaceae*. *Plant Mol Biol* 56:527–539
- Anderson JV, Horvath DP (2001) Random sequencing of cDNAs and identification of mRNAs. *Weed Sci* 49:581–589
- Asamizu E, Nakamura Y, Sato S, Tabata S (2000a) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res* 7:175–180
- Asamizu E, Nakamura Y, Sato S, Tabata S (2000b) Generation of 7,137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. *DNA Res* 7:127–130
- Balagopalan C (2002) Cassava Utilization in Food, Feed and Industry. In: Hillocks RI, Thresh JM (eds) *Cassava Biology, Production and Utilization*. CAB Intl, Wallingford, UK, pp 301–318
- Blum A (1998) *Plant breeding for stress environments*. CRC, Boca Raton, FL, pp 43–76
- Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res* 6:791–806
- Cattivelli L, Baldi P, Crosatti C, Di Fonzo N, Faccioli P, Grossi M, Mastrangelo AM, Pecchioni N, Stanca AM (2002) Chromosome regions and stress-related sequences involved in resistance to abiotic stress in *Triticeae*. *Plant Mol Biol* 6:649–665
- Chaves MM, Oliveira MM (2004) Mechanisms underlying plant resilience to water deficits: prospects for water-saving agriculture. *J Exp Bot* 55:2365–2384
- Chini A, Grant JJ, Seki M, Shinozaki K, Loake G (2004) Drought tolerance established by enhanced expression of the *CC-NB-LRR* gene, *ADR1*, requires salicylic acid, *EDS1* and *ABI1*. *Plant J* 38:810–822
- Cheong YH, Kim K-N, Pandey GK, Gupta R, Grant JJ, Luan S (2003) *CBL1*, a calcium sensor that differentially regulates salt, drought, and cold responses in *Arabidopsis*. *Plant Cell* 15:1833–1845
- Cho S, Ok S, Jeung J, Shim K, Jung K, You M, Kang K, Chung Y, Choi H, Moon H, Shin J (2004) Comparative analysis of 5,211 leaf ESTs of wild rice (*Oryza minuta*). *Plant Cell Rep* 22: 839–847

- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci* 160:1115–1123
- El Sharkawy MA (2004) Cassava biology and physiology. *Plant Mol Biol* 56:481–501
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Fregene M, Puonti-Kaerlas J (2002) Cassava Biotechnology. In: Hillocks RI, Thresh JM (eds) *Cassava Biology, Production and Utilization*. CAB Intl, Wallingford, UK, pp 179–207
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100
- Gorantla M, Babu PR, Reddy Lachagari VB, Reddy AMM, Wusirika R, Bennetzen JL, Reddy AR (2007) Identification of stress-responsive genes in *indica* rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings. *J Exp Bot* 58:253–265
- Gosti F, Bertauche N, Vartanian N, Giraudat J (1995) Abscisic acid-dependent and -independent regulation of gene expression by progressive drought in *Arabidopsis thaliana*. *Mol Gen Genet* 246:10–18
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la CN, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The gene ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258–D261
- Hirayama T, Ohto C, Mizoguchi T, Shinozaki K (1995) A gene encoding a phosphatidylinositol-specific phospholipase C is induced by dehydration and salt stress in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 92:3903–3907
- Hsieh HM, Liu WK, Huang PC (1995) A novel stress-inducible metallothionein-like gene from rice. *Plant Mol Biol* 28:381–389
- Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:377–403
- Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* 138–148
- Kader J-C (1997) Lipid-transfer proteins: a puzzling family of plant proteins. *Trends Plant Sci* 2:66–70
- Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13:889–906
- Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K (1994) Cloning of cDNA for genes that are early-responsive to dehydration-stress (ERDs) in *Arabidopsis thaliana* L.: identification of three ERDs as HSP cognate genes. *Plant Mol Biol* 25:791–798
- Knight H, Knight MR (2001) Abiotic stress signalling pathways: specificity and cross-talk. *Trends Plant Sci* 6:262–267
- Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Mol Genet Genomics* 270:24–33
- Kudla J, Xu Q, Harter K, Gruissem W, Luan S (1999) Genes for calcineurin B-like proteins in *Arabidopsis* are differentially regulated by stress. *Proc Natl Acad Sci* 96:4718–4723
- Larosa PC, Hasegawa PM, Rhodes D, Clithero JM, Watad AE, Bressan RA (1987) Abscisic acid stimulated osmotic adjustment and its involvement in adaptation of tobacco cells to NaCl. *Plant Physiol* 85:174–181
- Lecoœur J, Wery J, Turc O, Tardieu F (1995) Expansion of pea leaves subjected to short water deficit: cell number and cell size are sensitive to stress at different periods of leaf development. *J Exp Bot* 46:1093–1101
- Lindorff-Larsen K, Lerche HM, Poulsen FM, Roepstorff P, Winther JR (2001) Barley lipid transfer protein, LTP1, contains a new type of lipid-like post-translational modification. *J Biol Chem* 276:33547–33553
- Machuka J, Bashiardes S, Ruben E, Spooner K, Cuming A, Knight C, Cove D (1999) Sequence analysis of expressed sequence tags from an ABA-treated cDNA library identifies stress response genes in the moss *Physcomitrella patens*. *Plant Cell Physiol* 40:378–387
- Meijer HJ, Munnik T (2003) Phospholipid-based signaling in plants. *Annu Rev Plant Biol* 54:265–306
- Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, Rudd S, Weil B (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* 30:31–34
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, Bucher P, Copley RR, Courcelle E, Das U, Durbin R, Falquet L, Fleischmann W, Griffiths-Jones S, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lopez R, Letunic I, Lonsdale D, Silventoinen V, Orchard SE, Pagni M, Peyruc D, Ponting CP, Selengut JD, Servant F, Sigrist CJ, Vaughan R, Zdobnov EM (2003) The InterPro database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 31:315–318
- Netzer WJ, Hartl FU (1998) Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. *Trends Biochem Sci* 23:68–73
- Okogbenin E, Ekanayake IJ, Porto MCM (2003) Genotypic variability in adaptation responses of selected clones of cassava to drought stress in the Sudan Savanna Zone of Nigeria. *J Agric Crop Sci* 189:376–389
- Okogbenin E, Marin J, Fregene M (2006) An SSR-based molecular genetic map for cassava. *Euphytica* 147:433–440
- Olsson AS, Engström P, Söderman E (2004) The homeobox genes *ATHB12* and *ATHB7* encode potential regulators of growth in response to water deficit in *Arabidopsis*. *Plant Mol Biol* 55:663–677
- Palmiter RD (1998) The elusive function of metallothioneins. *Proc Natl Acad Sci USA* 95:8428–8430
- Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. *Plant Physiol* 137:1211–1227
- Reddy AR, Ramakrishna W, Chandra Sekhar A, Ithal N, Ravindra Babu P, Bonaldo MF, Soares MB, Bennetzen JL (2002) Novel genes are enriched in normalized cDNA libraries from drought-

- stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome* 45:204–211
- Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol* 3:108–116
- Rodriguez Milla MA, Butler E, Rodriguez Huete A, Wilson CF, Anderson O, Gustafson JP (2002) Expressed sequence tag-based gene expression analysis under Aluminum stress in Rye. *Plant Physiol* 130:1706–1716
- Rohila JS, Jain RK, Wu R (2002) Genetic improvement of Basmati rice for salt and drought tolerance by regulated expression of a barley *Hva1* cDNA. *Plant Sci* 163:525–532
- Rudd S (2005) openSputnik—a database to ESTablish comparative plant genomics using unsaturated sequence collections. *Nucleic Acids Res* 33:D622–D627
- Rudd S, Schoof H, Mayer K (2005) PlantMarkers—a database of predicted molecular markers from plants. *Nucleic Acids Res* 33:D628–D632
- Sakai T, Nakagawa Y, Uritani I, Data ES (1994) Occurrence and characterization of stress metabolites in cassava roots. In: Uritani I, Garcia VV, Mendoza EMT (eds) *Postharvest biochemistry of plant food-materials in the tropics*. Japan Scientific Society Press, Tokyo, pp 95–110
- Satou Y, Kawashima T, Kohara Y, Satoh N (2003) Large scale EST analyses in *Ciona intestinalis*: its application as Northern blot analyses. *Dev Genes Evol* 213:314–318
- Scott KD, Egger P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theor Appl Genet* 100:723–726
- Sharp RE, Wu Y, Voetberg GS, Saab IN, LeNoble ME (1994) Confirmation that abscisic acid accumulation is required for maize primary root elongation at low water potentials. *J Exp Bot* 45:1743–1751
- Sivamani E, Bahieldin A, Wraith JM, Al-Niemi T, Dyer WE, Ho THD, Qu R (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci* 155:1–9
- Taylor N, Chavarriaga P, Raemakers K, Siritunga D, Zhang P (2004) Development and application of transgenic technologies in cassava. *Plant Mol Biol* 56:671–688
- Tobias CM, Twigg P, Hayden DM, Vogel KP, Mitchell RM, Lazo GR, Chow EK, Sarath G (2005) Analysis of expressed sequence tags and the identification of associated short tandem repeats in switchgrass. *Theor Appl Genet* 111:956–964
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Treviño MB, O'Connell MA (1998) Three drought-responsive members of the nonspecific lipid-transfer protein gene family in *Lycopersicon pennellii* show different developmental patterns of expression. *Plant Physiol* 116:1461–1468
- Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Curr Opin Biot* 17:113–122
- Vierling E (1991) The role of heat-shock proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* 42:579–620
- Wood AJ, Duff RJ, Oliver MJ (1999) Expressed sequence tags (ESTs) from desiccated *Tortula ruralis* identify a large number of novel plant genes. *Plant Cell Physiol* 40:361–368
- Yamaguchi-Shinozaki K, Shinozaki K (1993) The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of *rd22*, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Mol Gen Genet* 238:17–25
- Zhang JZ, Creelman RA, Zhu J-K (2004) From laboratory to field. Using information from *Arabidopsis* to engineer salt, cold and drought tolerance in crops. *Plant Physiol* 135:615–621
- Zimmerman G, Bäumlein H, Mock H-P, Himmelbach A, Schweizer P (2006) The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. *Plant Physiol* 142:181–192