Plant Molecular Biology **56:** 527–539, 2004.

An EST resource for cassava and other species of Euphorbiaceae

James V. Anderson[1,*], Michel Delseny[2], Martin A. Fregene[3], Veronique Jorge[2,4], Chikelu Mba[3], Camilo Lopez[2], Silvia Restrepo[3], Mauricio Soto[3], Benoit Piegu[2], Valerie Verdier[2], Richard Cooke[2], Joe Tohme[3] and David P. Horvath[1]

[1]*USDA/ARS, Biosciences Research Laboratory, 1605 Albrecht Blvd., P.O. Box 5674, State University Station, Fargo, ND 58105, USA;* [2]*University of Perpignan, Genome and Plant Development Laboratory, UMR 5096 CNRS-IRD-UP, 52, Avenue Paul Alduy, 66860 Perpignan, France;* [3]*International Center for Tropical Agriculture (CIAT), AA6713, Cali, Colombia; (*author for correspondence; e-mail andersjv @fargo.ars.usda.gov);* [4]*present address: INRA, Unite Amelioration Genetique et Physiologie Forestieres, Av de la pomme de pin, BP 20619 ARDON, 45166 OLIVET Cedex, France*

Received 23 July 2003; accepted in revised form 2 April 2004

Key words: cassava, genomics, leafy spurge, stress

Abstract

Cassava (*Manihot esculenta*) is a major food staple for nearly 600 million people in Africa, Asia, and Latin America. Major losses in yield result from biotic and abiotic stresses that include diseases such as Cassava Mosaic Disease (CMD) and Cassava Bacterial Blight (CBB), drought, and acid soils. Additional losses also occur from deterioration during the post-harvest storage of roots. To help cassava breeders overcome these obstacles, the scientific community has turned to modern genomics approaches to identify key genetic characteristics associated with resistance to these yield-limiting factors. One approach for developing a genomics program requires the development of ESTs (expressed sequence tags). To date, nearly 23000 ESTs have been developed from various cassava tissues, and genotypes. Preliminary analysis indicates existing EST resources contain at least 6000–7000 unigenes. Data presented in this report indicate that the cassava ESTs will be a valuable resource for the study of genetic diversity, stress resistance, and growth and development, not only in cassava, but also other members of the Euphorbiaceae family.

Abbreviations: BAC, bacterial amplified chromosome; bp, base pair; CAPS, cleaved amplicon polymorphisms; CBB, cassava bacterial blight; CMD, cassava mosaic disease; EST, expressed sequence tag; PCR, polymerase chain reaction; QTLs, quantitative trait loci; RFLP, restriction fragment length polymorphism; SAGE, serial analysis of gene expression; SNP, single nucleotide polymorphisms

Introduction

Cassava is one of the most important crops in the tropical, inter-tropical, and sub-Saharan regions of the world for human food, because nearly 600 million people eat cassava every day. Worldwide, cassava acreage is more than 16 million hectares and annually produces root yields of more than 170 million tons. The increasing importance of cassava production as a staple to the world food supply is evidenced by an increased production of more than 75% during the last 30 years.

Losses in yields to cassava farmers are usually attributed to biotic and abiotic stresses such as diseases, drought, and acid soils, or from deterioration during the post-harvest storage of roots. To improve yields, cassava breeders are developing programs that target improved resistance to these factors. One approach is to improve yield by

limiting the impact of diseases. A major disease of cassava is Cassava Bacterial Blight (CBB) caused by *Xanthomonas axonopodis* pv. *manihotis* (Lozano, 1986). Other important diseases result from a variety of viruses including Cassava African Mosaic Virus, which causes Cassava Mosaic Disease (CMD) (Akano *et al*., 2002). A second target concerns starch production in the root, which is a major source of calories for human food and is a valuable source for the starch industry and its derivatives (Munyikwa *et al*., 1997). Other target goals include reducing losses caused by post-harvest processing and by environmental factors such as dehydration- and cold-stress, and acid soils.

Unfortunately, conventional cassava improvement is fraught with problems of breeding a long growth cycle, highly heterozygous crop. To increase the cost-effectiveness of achieving desired goals, a number of resources and molecular tools have been developed during the recent years to enhance breeding. They include construction of genetic maps using RFLP, isoenzymes, microsatellite markers (Fregene *et al*., 1997; Mba *et al*., 2001) that have already allowed the identification of a variety of QTLs and a major gene (*CMD2*) for CMD resistance (Jorge *et al*., 2000, 2001; Akano *et al*., 2002; Okogbenin and Fregene, 2002). Such markers are limited in their application to breeding, and a more precise approach to gene mapping using candidate genes is required. Unfortunately, relatively few genes have been identified so far in cassava. In order to isolate additional genes, BAC libraries have been constructed, covering most of the genome (Fregene *et al*., 2003). However, these BAC libraries have neither been ordered nor anchored on the genetic map. As a consequence, there is an important need for additional sequenced markers that would facilitate more detailed genetic and physical mapping.

The era of genomics and bioinformatics has increased our ability to identify markers, and increased our knowledge of plant genome structure, organization, and gene function. The recent explosion of genetic and genomic data for a wide range of plant and animal species has led to a proliferation of publicly available information databases throughout the internet (http://www.ncbi.nlm.nih.gov/dbEST, http://arabidopsis.org, http://rgp.dna.affrc.go.jp). Two complete plant genomes are available for *Arabidopsis* and rice (The Arabidopsis Genome Initiative, 2000; Goff *et al*., 2002; Delseny, 2003). The expressed sequence tags (ESTs; which are partial sequences [200–800 bp] of expressed genes randomly picked from a cDNA library) databases are currently the fastest growing and largest portion of these publicly available DNA sequence databases (Cooke *et al*., 1996; Ohlrogge and Benning, 2000). To date, at least 20 gene indices for plants (both dicot and monocot) that integrate data from international EST sequencing, genome sequencing, and gene research projects are publicly available (Quackenbush *et al*., 2001; www.tigr.org/tdb/plant.shtml). These databases are important for identifying expressed genes that are further used for developing DNA microarrays (Richmond and Somerville, 2000). The cDNA microarray technology, first developed by Schena *et al*., (1995), depends on the availability of ESTs. This technology has been widely received (Duggan *et al*., 1999) and used in plants to identify specific gene functions (Aharoni *et al*., 2000; Gutierrez *et al*., 2002), evaluate transcript profiles induced by various physiological or environmental conditions (Reymond *et al*., 2000; Van Hal *et al*., 2000; Lee *et al*., 2002; Oztur *et al*., 2002; Potokina *et al*., 2002; Zhu *et al*., 2003), and evaluate transcript profiles between genetically modified and control species (Van Hal *et al*., 2000). Although these are just a few of the excellent examples that have materialized from genomics initiatives, continued genome sequencing projects for many important crops are still underway and are expected to provide future benefits. However, the traditional funding communities have overlooked other important plant families, having impacts on world economies. In particular are members of the genetically diverse plant family Euphorbiaceae that includes, apart from cassava, other globally important agricultural species such as: castor bean (*Ricinus communis*), an important oil crop; Rubber tree (*Hevea brasiliensis*), an important source of rubber; Poinsettia (*Poinsettia pulcherrima*), an important horticultural crop; leafy spurge (*Euphorbia esula*) an important perennial pest weed that affects range, recreational, and right of way lands in North American plains and prairies; and annual weeds such as hophornbeam copperleaf (*Acalypha ostryifolia*), and endangered species such as *Akoka* and telephus spurge.

Although extensive studies have been carried out on a few model species, little is still known about basic physiological processes controlling crop plant development and resistance to stress and diseases. A significant understanding of the conservation and diversity of genes between members of the Euphorbiaceae family is also lacking. Consequently, it is currently difficult to design treatments or breeding programs to improve genetic stocks of desirable species or develop methods to control the growth of undesirable species. Many of these problems could be solved by the development of sequence databases and genomic-based research strategies for members of this family. Several groups have realized the success of large-scale genome sequencing for developing genomics resources and are starting to address the task of generating ESTs from cassava and related species, such as leafy spurge, in order to prepare gene catalogues that will be important for future development of DNA arrays and virtual Northerns. Since transformation systems already exist for cassava (Schöpke *et al*., 1996), the development of an EST resource will eventually allow one to isolate and manipulate key genes and metabolic pathways when they are established, and to introduce new genes when necessary. This review presents resources that are being established and will be available in the near future to the scientific community. We also provide evidence to show the importance of using these resources to understand genetic diversity and conservation within gene sequences and demonstrate potential advantages to a family-based rather than species-based genomics approach. We also provide data showing the effectiveness of several high throughput approaches to identify key genes involved in stress responses using data generated from EST sequencing. Finally, we will make the argument that these successes should be built upon and could be enhanced with additional sequencing efforts. Most of these data are not yet published but the aim of this article is to inform the scientific community of the progress being made, preliminary results, and developing/planned resources.

Materials and methods

Several materials and strategies have been used independently, and at different international locations to generate the present day resources. Libraries and ESTs from *Euphorbia esula* (leafy spurge) were generated at the Biosciences Research Laboratory in Fargo, North Dakota, USA. Libraries targeted to CMD were made under collaborations between CIAT in Cali, Colombia, and the Iwate Biotech Research Center (IBRC) in Kitakami, Japan. Libraries and ESTs for analyzing CBB and starch metabolism resulted from collaborations between labs at CIAT and at the University of Perpignan (France).

Plant material

Cassava plantlets, derived from meristem cultures (genotypes TME 3, TME 117, and TMS 30572), and for the development of future normalized cDNA libraries, were initially obtained from the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. TME 3 and TMS 30572 comprise different sources of genetic tolerance to CMD. TME 117, a drought-tolerant variety, is desired for its' sweet taste and texture throughout Africa, even after boiling, and comprises a relatively low cyanide content. Individual plantlets were transferred to 4 inch, square plastic pots containing Sunshine Mix #1. Each transplanted cassava plantlet was put inside a commercially available zip lock baggy and sealed. The plantlets were allowed to acclimate to growth chamber conditions for 3–4 weeks prior to opening the zip lock bags and allowing acclimation to continue an additional 2–3 weeks. After acclimating to growth chamber conditions, the plants were transferred to larger plastic pots containing 1 part Sunshine Mix #1 and 2 parts sandy loam. At this stage, the plants were transferred to greenhouse conditions. Plants grown in the greenhouse were fertilized once weekly using Prolific 20-20-20 (N-P-K). Temperatures were maintained at approximately 25 °C, 16/8 h day/night cycles with daylight supplemented by 400 W high-pressure sodium lamps in the greenhouse or with 60 W cool white high output fluorescent lamps supplemented with 60 W incandescent bulbs in the growth chambers. Light fluencies were approximately 350 μmoles m^{-2} s^{-1} in the greenhouse and approximately 80 μmoles m^{-2} s^{-1} in the growth chambers (LiCor-185 photometer, LiCor, Lincoln, NE). Cassava plants used for the CBB and starch cDNA libraries were derived from the

CIAT cassava germplasm. They were grown in greenhouse at 28/19 °C (day/night temperatures), under a 12-h day light photoperiod and 80% relative humidity. The cassava plants were grown from mature stem cuttings in sterile soil. Leafy spurge plant material used for growth induced, adventitious root bud cDNA libraries was grown as previously described (Anderson and Horvath, 2001).

Tissue treatment

Cassava plants (2 reps) from each variety were incubated at 25 °C (control) or 42 °C (heat shock) in an environmental growth chamber for a period of 4 h. Plants used for the heat shock study were obtained from the greenhouse 6–8 weeks after being transferred from the growth chamber to the greenhouse as previously described. The upper 4–6 inches of the plant, including the stem and meristem, were collected and immediately frozen in liquid N_2 and pulverized prior to storage at −80 °C.

To study the effects of dehydration-stress on cassava plant tissue, water was withheld from mature plants (1 year after transferring to greenhouse conditions) for a period of 6 days. Young leaf and petiole, and mature leaf, were collected from all three varieties of cassava, and individually ground in liquid N_2 prior to storage at −80 °C. Cold-treated plant tissue was obtained by placing the plants in an incubator set at 4–6 °C. After 30 h of cold-treatment, young leaf and petiole, and mature leaf were collected and processed as previously described. Control tissue was collected from untreated greenhouse plants.

Extraction of RNA for Northern and macroarray blotting

RNA was extracted from cassava plant material using the pine tree extraction method of Chang *et al.* (1993) using modifications described by Anderson and Horvath (2001). Total RNA was separated on a 1% denaturing agarose gel and blotted onto a positively charged nylon membrane (Hybond-N, Amersham Pharmacia Biotech) using standard protocols (Sambrook *et al.*, 1989). Northern blot hybridizations were accomplished using ^{32}P radiolabeled DNA probes (Rediprime II random prime labeling system; Amersham Pharmacia Biotech) incubated in Rapid-hyb buffer (Amersham Pharmacia Biotech) at 65 °C. RNA blots were washed under high stringency (0.1 × SSC, 0.1% SDS at 65 °C) and visualized on a Packard Instant Imager with approximately 1 h of exposure and by autoradiography. All experiments were replicated with separate sets of treated plants. Following visualization, filters were washed once with boiling 0.1% (v/v) SDS solution and allowed to cool to room temperature. Filters were rinsed with 0.1% (v/v) SDS solution at room temperature and removal of all radioactivity was assured by visualization of the clean filter for 1 h on the imager prior to re-probing with a new cDNA probe.

For macroarray probing, RNA was extracted as previously described above. Total RNA (30 μg) was labeled with [^{32}P]-dCTP using reverse transcriptase (SuperScript II, Invitrogene). With the exception of using radiolabeled dCTP vs. dATP, all labeling, hybridizations, and washes were done using the protocol described by Uhde-Stone *et al.* (2003). All arrays were visualized by autoradiography. Spot intensities (radioactivity) were determined using a Packard Instant Imager.

cDNA library construction used for developing Euphorbiaceae EST resources

cDNA libraries for cassava bacterial blight and starch content

The cDNA libraries have been made using a Stratagene kit. For starch biosynthesis studies, two cDNA libraries were made, one from genotype CM523-7 (a high dry matter content variety) and another from genotype Mper183 (a low dry matter content variety). Both libraries were made from root material collected 6 months after planting. For analyzing response to *Xanthomonas axonopodis* (*Xam*), several cDNA libraries were made from stems using genotypes MCol1522 (susceptible to *Xam* strain CI0151) and genotype MBra685 (resistant to *Xam* strain CI0151). Leaf and stem inoculations were done as previously described (Restrepo *et al.*, 2000). Additionally, several subtracted cDNA libraries have also been made with genotypes SG107-35 (highly resistant to *Xam* strain CI0-46), MCol1522 and MBra685. Tissue was collected 6, 12, 24, 48, 72 h and 7 days after inoculation in order to enrich the EST database with genes over-expressed in inoculated

material. cDNA synthesized from RNA obtained from either healthy or wounded stems were pooled and used as ''drivers''. The cDNA from the inoculated tissues was used as ''tester''. An additional subtractive library was made from leaf material collected from genotype MBra 685. In addition, a limited number of cDNA-AFLP fragments were sequenced and included in the data set. All these libraries are listed in Table 1.

Bacterial clones generated from these cDNA libraries were randomly distributed in 96 well microtiter plates prior to processing for plasmid isolation and sequencing using a 5′-specific primer. Sequencing was accomplished using an ABI 3100 capillary sequencing machine. Raw data were collected and processed to evaluate the quality of the sequence (using Phred) and to eliminate vector sequence. They were organized as multiFasta files and were used to construct contigs of overlapping sequences and singletons. These data will be transferred to GenBank immediately after submission of this manuscript.

cDNA libraries constructed for cassava mosaic disease resistance

A cDNA library was constructed in pYES (Invitrogen Inc.) according to the manufacturer's instructions using mRNA from CMD resistant progeny from the cassava genotype TME 3, the source of *CMD2*. Two microliters of the cDNA library were electroporated into 40 μl of *E. coli* HB101 cells (Gibco BRL) and plated on LB agar plates + ampicillin (100 μg ml^{-1}). A total of 5000 colonies were picked and placed in 70 μl of LB media + ampicillin (100 μg ml^{-1}) in 384 well plates. Plasmid isolation was done using the MONTAGE 96-well plate system (Millipore Inc). Primer designed from the 3′ end of the multiple cloning site of pYES (Invitrogen Inc.) and 5 μl of plasmid miniprep were used for sequencing each clone. Sequencing reactions were accomplished using the ABI Prism® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) on a 9600 Perkin Elmer Machine or an MJ Research DNA engine. The sequence reactions were cleaned using the multi screen 96-well plate format (Millipore Inc.) and analyzed on a Shimadzu RISA 384 capillary sequencing machine. Sequences obtained were manually cleaned from vector sequences and combined into one single text file using a program written in Perl, running on a SunSparc Station (Sun Microsystems Inc.). A program was written in Perl to perform batch BLAST (Altschul *et al*., 1997) similarity searches for sequence identification using the CIAT local BLAST site (http://gene2/BLAST/inicio.htm). Sequencing was done from the 3′ end in order to compare the EST sequences with a collection of SAGE tags derived from the same material (Fregene, unpublished data).

cDNA libraries constructed for leafy spurge (Euphorbia esula)

Construction of a cDNA library using mRNA isolated from the adventitious root buds (shoot buds below the crown) of 3-day excised plants, plasmid isolation, sequencing, and analysis of ESTs was done as previously described (Anderson and Horvath, 2001). EST sequences were submitted to the GenBank EST database (dbEST; Boguski

Table 1. Characteristics of the libraries constructed from cassava for CBB, CMD, and starch characteristics, and from leafy spurge for growth-induction in underground adventitious root buds.

Cultivar	Phenotype/Condition	Organ	Designation
CM523-7	High matter dry content	Roots	Starch-CM
MPer183	Low matter dry content	”	Starch-Mper
MCol1522	Sensible/Inoculated	Stem	MCol-48h
	Sensible/Subtracted	”	mc_ssh
MBra685	Tolerant/Non inoculated	Stem	MBra
	Tolerant/Not subtracted	”	mb_nosub
	Tolerant/Subtracted	Leaf	mb_dsc
SG107-35	Resistant/Not subtracted	Stem	sg_nosub
	Resistant/Subtracted	”	sg_ssh
	Resistant/Subtracted	”	sg_dsc
TME 3	Resistance to CMD/Field exposure	Leaf/stem	YEST
LS001	Growth-induced	Root buds	LS3-dgi

et al., 1993) where they were given GenBank accession numbers and kept in a publicly available archive. Contig sets and singletons in the leafy spurge EST-database are available to the public at the University of Minnesota, Center for Computational Genomics and Bioinformatics (http://web.ahc.umn.edu/biodata/euphorbia).

Results

Preliminary characterization of the EST resources

Characterization of ESTs obtained from cDNA libraries targeted for CBB and Starch

Good quality ESTs (11954) were obtained with an average length of 467 bp (Table 2). All were derived from the 5′ end of mRNA. They could be grouped into 1875 contigs containing 9218 ESTs, and singletons composed of 3825 ESTs. Therefore, at this stage, a unigene set can be built with 5700 sequences that include the 848 cassava sequences already present in GenBank (NCBI). A first observation is the relatively high redundancy within unsubtracted libraries, indicating that some of the analyzed tissues are highly specialized. For example, in root libraries, we have respectively an average of 2.9 ESTs/contig for high starch compared to 1.9 ESTs/contig for low starch. There are also more singletons in the low starch library.

A second characterization consists in comparing cassava ESTs with other sequences in publicly available databases. So far, about 50% of the singletons and contigs have significant homology with *Arabidopsis* of which half correspond to *Arabidopsis* hypothetic proteins. This data suggests that these hypothetic proteins are likely to be real proteins. About 18% of the ESTs have no homology to any known genes in available databases and may represent new plant genes specific of the Euphorbiaceae family. Classification of these ESTs into functional categories is underway.

Characterization of ESTs obtained from cassava cDNA libraries targeted for CMD

The 3′ end sequencing of about 5000 cDNA clones generated a total of 4000 ESTs (average length of 481 bp). The ESTs could be organized into 1505 unigenes (500 contigs, containing 2995 ESTs, and 1005 singletons). Homology with known genes and proteins deposited in public databases were obtained using the local BLAST (Altschul *et al.*, 1997) at CIAT. The identity of about 800 unigene sequences could be ascertained with a good confidence level. Redundancy found in sequences of known functions was about 30%. The ESTs were used for SAGE tag annotation. The annotation of the ESTs have been described elsewhere (Fregene *et al.*, 2003). The most abundant tags were easily annotated, for example, identity of the 10 genes that make up 5% of all expressed transcripts were

Table 2. Statistics of the EST collection for CBB and starch characterization.

Library	No. of sequences generated	No. of sequences analyzed	Number of TC[a]	Number of singleton[b]
Starch CM	5376	3608	642	555
Starch Mper	4992	3391	607	1127
MCol-48h	2304	1721	184	1178
mc_ssh	384	258	45	142
Mbra	2688	1560	158	1049
mb_nosub	384	258	22	124
mb_dsc	768	438	54	41
sg_nosub	288	128	6	95
sg_ssh	384	210	24	128
sg_dsc	768	382	29	17
Aflp		241	27	179
Genbank		848	98	534
Total[c]			1875	3825

[a]Number of contig was calculated as the contigs present in each library independent of other libraries.
[b]Number of singleton was calculated as the singleton present in each library independent of other libraries.
[c]The total was calculated as the number of singletons or contigs from all the libraries.

found by ESTs, but annotation of less abundant tags is not as efficient. This suggests that the PCR method of tag annotation is a more powerful route to annotating SAGE tags compared to ESTs from regular cDNA libraries (Fregene *et al*., 2003). On the other hand, ESTs from a normalized cDNA library may be a more efficient means of tag annotation compared to non-normalized libraries. The EST data will be submitted to GenBank.

Characterization of ESTs obtained from a leafy spurge cDNA library targeted for dormancy in adventitious root buds

The average length generated from sequencing runs on 1983 isolated plasmids was 469 bases. From the 1983 sequencing reactions analyzed, quality reads were obtained for 1814 ESTs. Analysis of the ESTs indicated the presence of 246 contigs composed of 2 or more overlapping sequences. In all, 642 ESTs (35%) were present in the 246 contigs with the remaining 1172 representing singletons. This data suggests a redundancy ratio of ~25%. Approximately 28% of the ESTs are classified as either unknowns or hypothetical proteins. Other classifications representing significant numbers within the EST-database include ribosomal proteins, elongation factors, protein kinases, cell cycle proteins, heat shock proteins, aquaporins, and DNA-binding proteins (Anderson and Horvath, 2001).

Comparison of selected orthologous genes within cassava and leafy spurge databases

From the small number of accessions in GenBank for cassava and leafy spurge, a select number of orthologous genes (Table 3) were analyzed for similarity and the data suggested a good probability for cross-hybridization on heterologous systems. Recent experiments using *Arabidopsis* cDNA microarrays have indicated that substantial cross-hybridization between species could provide relevant expression data (Horvath *et al*., 2003). Consequently, efforts were undertaken to use leafy spurge ESTs to produce both macro- and microarrays, and as direct probes for Northern blot analysis of cassava RNA. Preliminary analysis from experiments done using a first round of leafy spurge microarrays, developed at the Fargo Lab, indicated that approximately 15–25% of the leafy spurge clones hybridise well to labeled target DNA from cassava leaf tissue (Anderson and Horvath, personal communication). Substantially higher levels of hybridization were detected when cassava RNA from dehydration-stressed leaf tissue was

Table 3. Comparison of sequence identities between homologues of cassava and leafy spurge.

Gene Product	Spurge accession	Cassava accession	(%) identity	Consensus length
Dna J	AF239932	BI325097	78.9	152
Dna J	AF239932	BI325100	69.9	216
Histone H1	AF222804	BI325218	55.4	725
Histone H3	AF239930	BI325152	53.4	696
Histone H3	AF239930	BI325185	81.9	155
Histone H3	AF239930	BI325226	79.5	215
Histone H3	AF239930	BI239930	82.5	388
Histone H4	BI946403	BI325254	68.2	320
Histone H4	BI946403	BI325167	62.4	225
Histone H4	BI946403	BI325241	65.2	157
Lhcb	AF220527	BI325250	77.8	234
Lhcb	AF220527	BI325141	77.2	433
Lhcb	AF220527	BI325235	79.9	402
Lhcb	AF220527	BI325188	56.7	236
Polyubiquitin	AW944681	BI325109	60.5	208
Polyubiquitin	BG345192	BI325109	68.9	148
Polyubiquitin	BI975267	BI325109	71.4	139
RuBisCo	BI993507	BI325191	69.3	386
RuBisCo	BI975169	BI325191	67.6	283
14-3-3	AF222805	BI325181	69.1	137
		Average	70	293

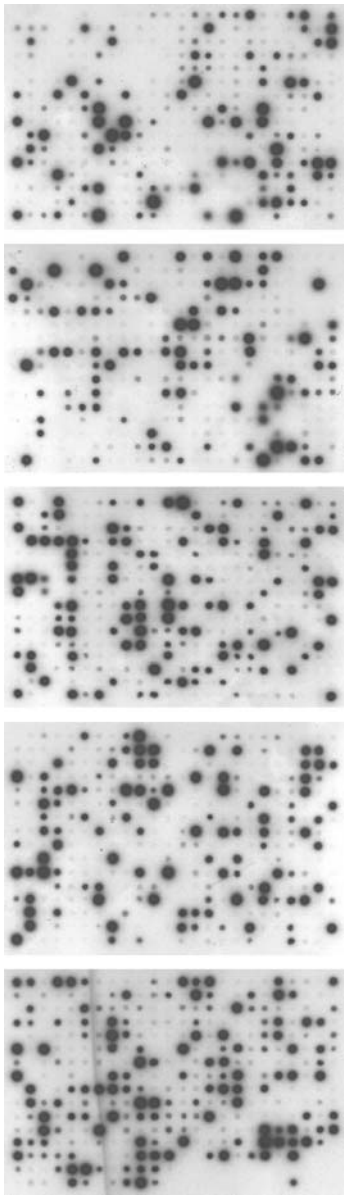

Figure 1. Leafy spurge DNA macroarray developed from an EST-database and probed with target cDNA developed from total RNA isolated from young leaf tissue of cassava (genotype TME 117). PCR-amplified cDNA corresponding to each EST in the leafy spurge database were spotted onto Hybond-N nylon filters from 384 well plates using a 384 well replicator. Hybridizations and washes were done as described by Uhde Stone *et al*. (2003). Figure shown represents replicate 1 data obtained during the probing of two replicate blots. Both replicate blots showed identical patterns of hybridization.

used to probe macroarrays developed from the leafy spurge EST database (Figure 1). At least 35% of the leafy spurge clones showed greater than 2 × above background hybridization with the

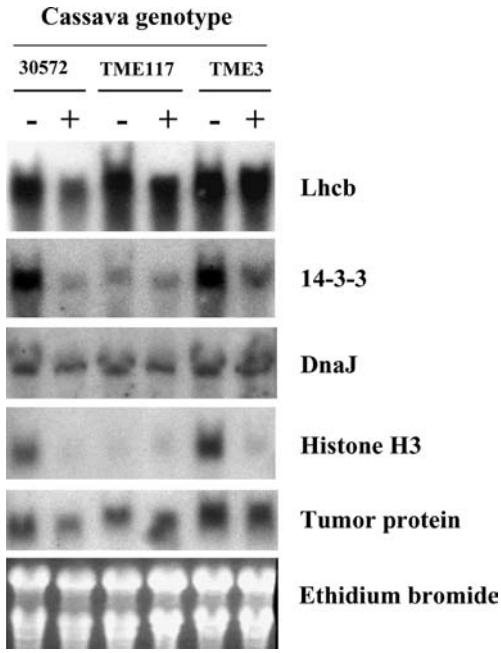

Figure 2. RNA blot of cassava leaf tissue probed with labeled cDNAs from leafy spurge. Each lane contains 20 μg of total RNA extracted from cassava leaf tissue either incubated at 25 °C (−) or 42 °C (+) for 4 h. Leafy spurge clones Lhcb (accession #AF220527), 14-3-3 (accession #BE095293), DnaJ (accession #AW840603), Histone H3 (accession #AF239930), and Tumor protein (accession #BI993560) were used to develop radioactive probes used for each hybridization. Replicate blots showed similar patterns of hybridization.

dehydration-stressed cassava target sample (Figure 1).

To determine if cross species hybridization could provide meaningful expression data, several genes identified as showing greater than 55% similarity (Table 3), or showing hybridization to micro- or macro-arrays during preliminary screenings, were used directly to probe Northern blots of corresponding cassava RNA. The results confirmed the differential expression of several genes in cassava leaf tissue exposed either to heat-, dehydration-, or cold-stress (see Figures 2 and 3). TME 117 (a drought-tolerant cassava variety) showed the least differential gene expression during heat-shock treatments compared to TMS 30572 and TME 3 (Figure 2). Interestingly, down regulation of Histone H3 (a marker for the S-phase of cell division) and 14-3-3 (involved in stress-related signal-transduction) during heat shock was observed in TMS 30572 and TME 3. The apparent stability of gene expression in TME

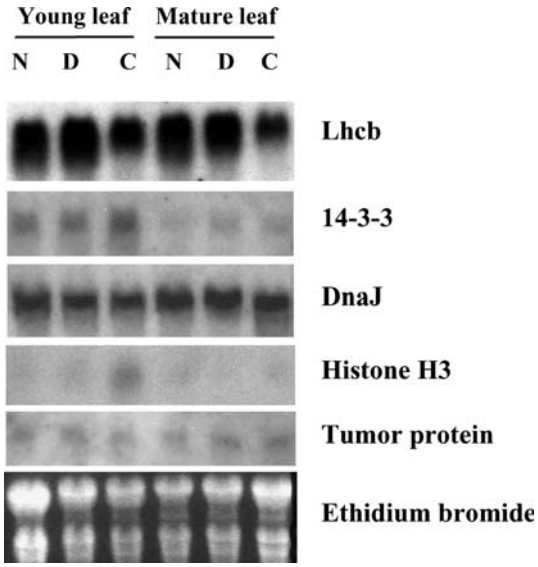

Figure 3. RNA blot of cassava leaf tissue probed with labeled cDNAs from leafy spurge. Young and mature leaf tissue was obtained from cassava genotype TME 117 and treatments are indicated as: N, normal leaf tissue; D, 7-day dehydration stressed; C, 30 h cold-stressed. Leafy spurge clones Lhcb (accession #AF220527), 14-3-3 (accession #BE095293), DnaJ (accession #AW840603), Histone H3 (accession #AF239930), and Tumor protein (accession #BI993560) were used to develop radioactive probes used for each hybridization. Replicate blots showed similar patterns of hybridization.

117 during heat-shock may be a reflection of its resistance to dehydration-stress. However, the down-regulation of Lhcb in cold-treated TME 117 does indicate that this cassava variety is sensitive to other environmental stressors (Figure 3).

Future prospects

Sequencing of approximately 25000 Euphorbiaceae ESTs has been accomplished from various tissues, genotypes, and species. A thorough comparison of the EST resources developed at the different international laboratories, indicated in this report, has not yet been accomplished and will need to be done before a true set of ''Euphorbiaceae-specific'' unigenes can be confirmed. However, estimates based on our preliminary analysis indicate that ~6000–7000 unigenes have been identified for cassava and leafy spurge.

Consequently, sufficient resources are available to begin large-scale expression profiling for these genes that should identify genes responsive to specific stresses or developmental processes important to controlling the growth and productivity of these agronomically important species. The availability of an EST resource for Euphorbiaceae opens up a number of research avenues concerning not only cassava but also several plants from the same family. About 750 ESTs are already available for castor bean seeds (Van de Loo *et al*., 1995) and 910 for *Hevea* (Ko and Han, unpublished, dbEST). The ability to detect single-copy orthologous genes within families, and the recent concept that conserved orthologous sets (COS) of genes can be used as COS markers for comparative mapping studies between highly divergent plant genomes, has advanced the study of comparative genomics in higher plants (Paterson *et al*., 1996; Fulton *et al*., 2002; Salse *et al*., 2002; Dominguez *et al*., 2003; Gebhardt *et al*., 2003). Since genomics projects have been initiated for several of the Euphorbiaceae species, such as cassava and leafy spurge (discussed in this report), and rubber tree (Ko and Han, unpublished), the potential for unlocking genetic diversity within important Euphorbiaceae family members exists.

Gene identification and discovery

A first perspective is to analyze, in more detail, the available sequences currently in our databases that have been raised in order to identify genes involved in response to pathogen-associated diseases (CBB, CMD), starch metabolism, dehydration stress, and dormancy. Therefore, in the present data set, we have the potential to identify a large proportion of the known plant genes related in response to pathogens, genes for enzymes involved in starch biosynthesis, modification and degradation, as well as dehydration tolerance, and genes regulating dormancy status. Eventually, we expect to find new variants for these genes, with perhaps new specificities. Although the genetic diversity among genotypes may make identifying key trait specific gene expression difficult, the developing tools and databases described here will serve as critical resources for more detailed studies with inbreed or transgenic lines.

A second effort is to identify the sequences that show no homology or similarity with other plant or animal genes. These ''unknowns'' might represent new genes, although many of them might simply correspond to divergent coding sequences of already known genes or to their relatively spe-

cific 5′ or 3′ untranslated sequences. Because Cassava belongs to a poorly studied yet important family, the chance of identifying genes specific for this crop or this family is reasonably high. So far, the resources are based on the limited number of ESTs that have been sequenced from a small number of tissues and biological situations. Obviously, resources in the order of 6–7000 unique genes is small with respect to the expected number of genes which has been estimated at ~26000, or fewer, in *Arabidopsis* (The Arabidopsis Genome Initiative, 2000) and ~32000–50000 in rice (Goff *et al*., 2002; Delseny, 2003). More ESTs need to be prepared from various developmental stages in order to successfully identify a full unigene set that can be used for additional studies. This will require using different strategies: sequencing the 3′ ends of the available clones, subtraction of the present libraries to isolate rare clones, and preparing new libraries from different organs and tissues in different biological conditions. Currently, the main efforts are being directed at increasing the number of ESTs, and this should be considered as one of the goals of the Cassava Global Genome Project. As part of this goal, efforts are currently underway in collaborations between IITA and the USDA Agricultural Research Service to develop two normalized cDNA libraries for cassava genotype TME 117 (a drought-tolerant variety) that will be used to produce an additional 5000–8000 unigene set. When all these data are available, bioinformatics resources will be needed for processing them to identify as many functions as possible and to contig the different sequences from all of the existing EST resources. In particular, sequencing the 3′ ends of available clones should improve the analysis of these contigs and discriminate sequences that have been artificially contiged with each other.

Finally, another resource that has still to be developed is the creation of a full-length cDNA collection. So far no real effort has been made to raise such high quality libraries. Experience in *Arabidopsis* and rice demonstrated these resources are strategic both for genomic annotation and functional characterization of the genes by ectopic expression (Seki *et al*., 2002).

Gene expression and profiling

An important application of EST programmes is global gene expression analysis using DNA chips (microarrays). The first generation of DNA chips for cassava and leafy spurge (Euphorbiaceae) are representing only part of the genome and they are made directly by spotting amplified clones corresponding to each EST. Such a DNA chip is presently being made at CIAT from one of the unigene sets described in this paper and will be further complemented with additional resources that have to be merged into a single unigene set when new resources are available. Nevertheless, this first generation of ''Euphorbiaceae-specific'' chips will be extremely useful for classifying unknown genes into functional categories. The first set of experiments will analyze transcript responses to pathogens, dissect starch metabolism pathways, and pathways acting in drought-tolerance, and other important biotic- and abiotic-stresses. Data presented in this report (Figures 1–3) already shows the potential for using these Euphorbiaceae-specific micro-arrays/macro-arrays to screen for differentially expressed genes within varieties of cassava.

Development of oligonucleotide chips will be initiated when more information about gene sequence is available in order to decrease cross hybridization interferences due to members of multigene families. This problem can be anticipated from the allopolyploid nature of Cassava (Olsen and Schaal, 1999).

Mapping and comparative genomics

As mentioned in the introduction, and this issue, a genetic map of Cassava has been developed as well as several BAC libraries. The EST resource should be invaluable to increase the density of gene markers on the genetic map and to contribute to the ordering of the BAC clones into a physical map and anchoring it on the genetic map.

ESTs can be used either as RFLP or CAPS markers on the mapping populations. It can be expected that several thousand ESTs can be mapped in the coming 2 or 3 years. Meanwhile they can be located on BAC clones even more rapidly because there is no need for polymorphism and because this can be done using high throughput PCR strategies. We can anticipate that a number of known genes of interest will be mapped first but it is also important to randomly map a large number of sequences. Particularly important is the mapping of members of multigene families because cassava is an allotetraploid genome.

Another important question is when and how many rounds of polyploidization have occurred (Wolfe, 2001; Simillion *et al*., 2002; Blanc *et al*., 2003; Bowers *et al*., 2003). It will be interesting to compare the available map with that of the presumed cassava ancestors. The EST resource should be a fantastic tool to examine evolutionary relationships within the Euphorbiaceae family. For example, preliminary experiments have shown that DNA arrays constructed from the small set of leafy spurge ESTs, described in this report, have consistently shown between 15 and 35% hybridization 2 × above background with target DNA from cassava leaf tissue (Figure 1, and Anderson *et al*., 2001). Additionally, as already mentioned, some ESTs already exist for *Ricinus* and *Hevea*, and they can also be compared with those of Cassava. Similarly, there is already a detailed genetic map of rubber tree (Lespinasse *et al*., 2002) and many ESTs from cassava cross-hybridize with *Hevea* DNA and can be used for further mapping purposes. Such an approach should help in establishing syntenic regions between the different genomes and in facilitating positional cloning of genes of interest in these species. Thus, additional high throughput gene discovery and sequencing projects directed toward specific problems in Euphorbiaceae members should provide the groundwork for unlocking genetic diversity within this family. Of equal importance, it should enhance our ability to control the growth and productivity of various members of this plant family and increase the possibility for map-based cloning.

Genetic and allelic diversity

Contigs available for existing ESTs already revealed a high degree of genetic diversity between cassava genotypes. This is not unexpected because of the allotetraploid nature of Cassava and because the domestication of this crop has not been as intense as that of maize or rice. Generating additional cassava ESTs from the same cultivars already analyzed, and from others, should give some indication about the distribution and divergence of orthologous genes and their allelic diversity. Such knowledge should provide new tools for mapping and breeding based on the SNP diversity that already exists.

As a conclusion, we hope that developing these EST resources and its derivatives will contribute to a more rapid improvement of cassava breeding, increase cassava production, unlock the genetic diversity within the Euphorbiaceae family, and assist in our quest to regulate growth and development in undesirable species. Most of the ESTs presented in this review are already in the process of being transferred to public databases so that they are available for the public community. Obviously, this resource still needs to be amplified, but it already makes a significant contribution to basic knowledge in Euphorbiaceae family members such as cassava and leafy spurge. Developing it is a strategic point of the Cassava Global Genome Project.

Acknowledgements

The CBB and starch EST project was funded by CGIAR through the Agropolis (Montpellier, France) platform for developing countries and benefited from the support of the Montpellier Languedoc-Roussillon Génopole; C Lopez was supported by a PhD fellowship from IRD (Institut pour la Recherche et le Developpement). Funding for the initiation and development of normalized cDNA libraries and EST-databases for cassava variety TME 117 was obtained from USAID-linkage funds through the International Institute for Tropical Agriculture, (IITA), Ibadan, Nigeria, in collaboration with Alfred Dixon, Ivan Ingelbrecht, and Francis Moonan. Special thanks to Dr Earnest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, for providing bioinformatics and hosting of the leafy spurge EST database. Thanks also to Ryohei Terauchi for help with the cDNA library for CMD resistance, and to the IBRC and JSPS for funding.

References

Aharoni, A., Keizer, L.C.P., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M.M.L., De Vos, R.C.H., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J. and O'Connell, A.P. 2000. Identification of the *SAAT* gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12: 647–661.

Akano, O., Dixon, A., Mba, C., Barrera, E. and Fregene, M. 2002. Genetic mapping of a dominant gene conferring resistance to Cassava mosaic disease. Theor. Appl. Genet. 105: 521–525.

Altschul, S.F., Madden, T.L., Shaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs'. Nucleic Acids Res. 25: 3389–3402.

Anderson, J.V., Gedil, M., Horvath, D.P. and Dixon, A. 2001. Preliminary studies directed towards the development of Euphorbiaceae-specific microarrays. In: N.J. Taylor, F. Ogbe and C.M. Fauquet (Eds.), 5th International Scientific Meeting of the Cassava Biotechnology Network: Abstract Book Donald Danforth Plant Science Center, St. Louis, Missouri, pp. S5-02.

Anderson, J.V. and Horvath, D.P. 2001. Random sequencing of cDNAs and identification of mRNAs. Weed Sci. 49: 581–589.

Blanc, G., Hokamp, K. and Wolfe, K.H. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res. 13: 137–144.

Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. 1993. DbEST-database for "expressed sequence tags". Nat. Genet. 4(4): 332–333.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosome duplication events. Nature 422: 433–438.

Chang, S., Puryer, J. and Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Rep. 11: 113–116.

Cooke, R., Raynal, M., Laudier, M., Grellet, F., Delseny, M., Morris, P.C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y.F., Mache, R., Krivitzky, M., Gy, I.J.J., Kreis, M., Lecharny, A., Parmentier, Y., Marbach, J., Fleck, J., Clement, B., Phillips, G., Herve, C., Bardet, C., Tremousaygue, D., Lescure, B., Lacomme, C., Roby, D., Jourjon, M.F., Chabrier, P., Charpenteau, J.L., Desprez, T., Amselem, J., Chiapello, H. and Hofte, H. 1996. Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. Plant J. 9: 101–124.

Delseny, M. 2003. Towards an accurate sequence of the rice genome. Curr. Opin. Plant Biol. 6: 101–105.

Dominguez, I., Graziano, E., Gebhardt, C., Barakat, A., Berry, S., Arus, P., Delseny, M. and Barnes, S. 2003. Plant genome archeology: evidence for conserved ancestral chromosome segments in dicotyledonous species. Plant Biotechnol. J. 1: 91–99.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J.M. 1999. Expression profiling using cDNA microarrays. Nat. Genet. Supplement 21: 10–14.

Fregene, M.A., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J. and Bonierbale, M.W. 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 95: 431–441.

Fregene, M.A., Matsumura, H., Akano, A., and Dixon, A. and Terauchi, R. 2003. Serial analysis of gene expression (SAGE) of host plant resistance to the cassava mosaic disease (CMD). Plant Mol. Biol. (in press).

Fulton, T.M., Van der Hoeven, R., Eannetta, N.T. and Tanksley, S.D. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457–1467.

Gebhardt, C., Walkemeier, B., Henselewski, H., Barakat, A., Delseny, M. and Stuber, K. 2003. Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. Plant J. 34: 529–541.

Goff, S.A., Rick,e D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quai,l 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, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92–100.

Gutierrez, R.A., Ewing, R.M., Cherry, J.M. and Green, P.J. 2002. Identification of unstable transcripts in *Arabidopsis* by cDNA microarray analysis: rapid decay is associated with a group of touch and specific clock-controlled genes. Proc. Natl. Acad. Sci. USA. 99(17): 11513–11518.

Horvath, D.P., Schaffer, R., West, M. and Wisman, E. 2003. *Arabidopsis* microarrays identify conserved and differentially-expressed genes involved in shoot growth and development from distantly related plant species. Plant J. 34: 125–134.

Jorge, V., Fregene, M.A., Duque, M.C., Bonierbale, M.W., Tohme, J. and Verdier, V. 2000. Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 101: 865–872.

Jorge, V., Fregene, M.A., Velez, C.M., Duque, M.C., Tohme, J. and Verdier, V. 2001. QTL analysis of field resistance to *Xanthomonas axonopodis* pv. *manihotis* in cassava. Theor. Appl. Genet. 102: 564–571.

Kohler, A., Delaruelle, C., Martin, D., Encelot, N. and Martin, F. 2003. The poplar root transcriptome: analysis of 7000 expressed sequence tags. FEBS Lett. 542(1–3): 37–41.

Lee, J.M., Williams, M.E., Tingey, S.V. and Rafalski, J.A. 2002. DNA array profiling of gene expression changes during maize embryo development. Funct. Integr. Genomics 2(1–2): 13–27.

Lespinasse, D., Rodier-Gout, M., Grivet, L., Leconte, A., Legnaté, H. and Seguin, M. 2002. A saturated genetic linkage map pf rubber tree (*Hevea* ssp.) based on RFLP, AFLP, microsatellite and isozyme markers. Theor. Appl. Genet. 100: 975–984.

Lozano, J.C. 1986. Cassava bacterial blight: a manageable disease. Plant. Dis. 70: 1089–1093.

Mba, R.E.C., Stephenson, P., Edwards, K., Melzer, S., Mkumbira, J., Gullberg, U., Appel, K., Gale, M., Tohme, J. and Fregene, M.A. 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor. Appl. Genet. 101: 21–31.

Munyikwa, T.R.I., Langeveld, S., Salehuzzaman, S.N.I.M., Jacobsen, E. and Visser, R.G.F. 1997. Cassava starch biosynthesis: new avenues for modifying starch quantity and quality. Euphytica 96: 65–75.

Ohlrogge, J. and Benning, C. 2000. Unravelling plant metabolism by EST analysis. Curr. Opin. Plant Biol. 3: 224–228.

Okogbenin, E. and Fregene, M. 2002. Genetic analysis and QTL mapping of early root bulking in an F1 population from non-inbred parents in cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 106: 58–66.

Olsen, K.M. and Schaal, B.A. 1999. Evidence on the origin of Cassava: phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. USA 96: 5586–5591.

Oztur, Z.N., Talame, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R. and Bohnert, H.J. 2002. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol. Biol. 48(5–6): 551–573.

Paterson, A.H., Lan, T.H. and Reischmann, K.P. 1996. Towards a unified genetic map of higher plants transcending the monocot-dicot divergence. Nat. Genet. 14: 380–382.

Potokina, E., Sreenivasulu, N., Altschmied, L., Michalek, W. and Graner, A. 2002. Differential gene expression during seed germination in barley (*Hordeum vulgare* L.). Funct. Integr. Genomics 2(1–2): 28–39.

Quackenbush, R.C., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvisi, B., Pertea, G., Sultana, R. and White, J. 2001. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res. 29: 159–164.

Restrepo, S., Duque, M.C. and Verdier, V. 2000. Characterization of pathotypes among isolates of *Xanthomonas axonopodis* pv. *manihotis* in Colombia. Plant Pathol. 49: 680–687.

Reymond, P., Weber, H., Damond, M. and Farmer, E.E. 2000. Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. Plant Cell 12: 707–719.

Richmond, T. and Somerville, S. 2000. Chasing the dream: plant EST microarrays. Curr. Opin. Plant Biol. 3: 108–116.

Salse, J., Piegu, B., Cooke, R. and Delseny, M. 2002. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequence project. Nucleic Acids Res. 30: 2316–2328.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning – A Laboratory Manual. 2nd edn. Cold Spring Harbor Laboratory Press, New York, USA.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467–470.

Schöpke, C., Taylor, N., Carcamo, R., N'Da, K.K., Marmey, P., Henshaw, G.G., Beachy, R.N. and Fauquet, C.M. 1996. Regeneration of transgenic cassava plants (*Manihot esculenta* Crantz) from microbombarded embryogenic suspension cultures. Nat. Biotechnol. 14: 731–735.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K. 2002. Functional annotation of a full-length *Arabidopsis* cDNA collection. Science 296: 141–145.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van de Peer, Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 15: 13627–13632.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Uhde-Stone, C., Zinn, K.E., Ramirez-Yanez, M., Li, A., Vance, C.P. and Allan, D.L. 2003. Nylon filter arrays reveal differential gene expression in proteoid roots of white lupin in response to phosphorus deficiency. Plant Physiol. 131: 1064–1079.

Van de Loo, F.J., Turner, S. and Somerville, C. 1995. Expressed sequence tags from developing castor seed. Plant Physiol. 108: 1141–1150.

Van Hal, N.L.W., Vorst, O., van Houwelingen, A.M.M.L., Kok, E.J., Peijnenburg, A., Aharoni, A., van Tunen, A.J. and Keijer, J. 2000. The application of microarrays in gene expression analysis. J. Biotech. 78: 271–280.

Wolfe, K.H. 2001. Yesterday's polyploids and the mystery of diploidization. Nature Rev. Genet. 2: 333–341.

Zhu, T., Budworth, P., Chen, W., Provart, N., Chang, H.S., Guimil, S., Su, W., Ester, B., Zou, G.Z. and Wang, X. 2003. Transcriptional control of nutrient partitioning during rice grain filling. Plant Biotechnol. J. 1: 59–70.