DNA Research **12**, 27–38 (2005)

Transcription Profiling of Signal Transduction-Related Genes in Sugarcane Tissues

Flávia Stal Papini-Terzi,[1,†] Flávia Riso Rocha,[1,†] Ricardo Zorzetto Nicoliello Vêncio,[2] Kátia Cristina Oliveira,[1] Juliana de Maria Felix,[3,4] Renato Vicentini,[4] Cristiane de Souza Rocha,[4] Ana Carolina Quirino Simões,[1] Eugênio César Ulian,[5] Sônia Marli Zingaretti di Mauro,[6] Aline Maria Da Silva,[1] Carlos Alberto de Bragança Pereira,[2] Marcelo Menossi,[3,4] and Gláucia Mendes Souza[1,*]

Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748 05508-900, São Paulo, Brasil,[1] *Departamento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010 05508-090, São Paulo, Brasil,*[2] *Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, CP 6010, 13083-970, Campinas, Brasil,*[3] *Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, CP 6010, 13083-970, Campinas, Brasil,*[4] *Centro de Tecnologia Copersucar, CP 162, 13400-970, Piracicaba, São Paulo, Brasil,*[5] *and Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, Universidade Estadual Paulista, Centro de Estocagem Clones (BCCCenter) Via de Acesso Professor Paulo Donato Castellane, S/N 14870-000, Jaboticabal, São Paulo, Brasil*[6]

(Received 28 September 2004; revised 28 December 2004)

Abstract

A collection of 237,954 sugarcane ESTs was examined in search of signal transduction genes. Over 3,500 components involved in several aspects of signal transduction, transcription, development, cell cycle, stress responses and pathogen interaction were compiled into the Sugarcane Signal Transduction (SUCAST) Catalogue. Sequence comparisons and protein domain analysis revealed 477 receptors, 510 protein kinases, 107 protein phosphatases, 75 small GTPases, 17 G-proteins, 114 calcium and inositol metabolism proteins, and over 600 transcription factors. The elements were distributed into 29 main categories subdivided into 409 sub-categories. Genes with no matches in the public databases and of unknown function were also catalogued. A cDNA microarray was constructed to profile individual variation of plants cultivated in the field and transcript abundance in six plant organs (flowers, roots, leaves, lateral buds, and 1st and 4th internodes). From 1280 distinct elements analyzed, 217 (17%) presented differential expression in two biological samples of at least one of the tissues tested. A total of 153 genes (12%) presented highly similar expression levels in all tissues. A virtual profile matrix was constructed and the expression profiles were validated by real-time PCR. The expression data presented can aid in assigning function for the sugarcane genes and be useful for promoter characterization of this and other economically important grasses.

Key words: *Saccharum*; sugarcane; microarray; signal transduction; transcriptome

1. Introduction

The unraveling of signal transduction pathways is of strategic importance to the understanding of fundamental processes such as growth and development as well as cellular responses triggered by biotic and abiotic stresses. In recent years, the wealth of information related to signal transduction generated by several genome sequencing projects, coupled with the global transcription profiling of a diversity of organisms, has brought many aspects of signaling under scrutiny. Protein superfamilies, such as protein kinases and transcription factors, have been systematically classified and analyzed following their identification by the sequencing projects[1–4] and comparative studies of complete genomes are defining the conserved signaling modules and revealing their inherent differences.[5–7]

Communicated by Satoshi Tabata

* To whom correspondence should be addressed. Tel. +55-11-3091-3815 ext. 216, Fax. +55-11-3091-2186, E-mail: glmsouza@iq.usp.br

† The first two authors contributed equally to this work.

‡ Accession numbers: The gene expression data was deposited in The Gene Expression Omnibus (GEO) Database under access numbers GSE1702 (series), GPL1375 and GPL1376 (platforms) and GSM29453 to GSM29506 (samples).

§ Supplementary Materials can be found at http://www.sucest-fun.org/pub/SUCAST.

The tropical crop sugarcane (*Saccharum* sp.) is of great economical interest, contributing to about two-thirds of the world's raw sugar production. In some countries, part of the crop is also destined to the production of ethanol, a less polluting fuel alternative. Traditional breeding programs that select for varieties showing high productivity and resistance to stresses and diseases are slow. Therefore, it could be greatly advantageous to have genes associated with desirable traits targeted for directed improvement of sugarcane varieties. With the aim of expediting sugarcane genomics, the SUCEST consortium (http://sucest.lad.ic.unicamp.br/public) sequenced and annotated 237,000 expressed sequence tags (ESTs) derived from 26 cDNA libraries.[8] The sequences were assembled into 43,141 contigs or sugarcane assembled sequences (SASs) covering an estimated 90% of the expressed genome. As for all other cDNA and genomic sequences released, the challenge now is to attribute relevant biological information to the extracted data. Several studies have described particular features of sugarcane's general metabolism, growth and development based on the analysis of the data from the SUCEST project.[9] Notwithstanding, given the enormous amount of data generated by a project of this magnitude, many topics remain to be investigated. The SUCAST Project (Sugarcane Signal Transduction)[10] is an ongoing effort that aims to identify the sugarcane signaling components and define their role in grasses. In this study, we present the SUCAST Catalogue and its categories, and investigate gene expression patterns using cDNA microarrays.

2. Materials and Methods

2.1. Annotation

The Sugarcane cDNA sequences can be found at the SUCEST database (http://sucest.lbi.ic.unicamp.br/public/) and GenBank under Accession Numbers CA064599-CA301538. Members of the SUCAST catalogue were identified using the BLAST algorithm[11] with conserved protein sequences as drivers. Conserved protein family domains were identified by searches at the Pfam[12] and SMART[13] databases using default parameters.

2.2. PCR amplification and array printing

Sugarcane cDNA plasmid clones of 1,632 ESTs obtained from the SUCEST collection were re-arranged and amplified in 100-μl PCR reactions (40 cycles, annealing at 51°C), directly from bacterial clones in culture, using T7 and SP6 primers. Ninety percent of the clones had their identity validated by re-sequencing. PCR products were purified by filtration using 96-well filter plates (Millipore Multiscreen® MAFBN0B50). Samples were visualized on 1% agarose gels to inspect PCR amplification quality and quantity. Purified PCR products (in 10 mM Tris-HCl solution at pH 8.0) were mixed with an equal volume of DMSO in 384 well V-bottom plates. Microarrays were constructed by arraying cDNA fragments on DMSO optimized, metal-coated glass slides (type 7, Amersham Biosciences) using the Generation III Microarray Spotter (Molecular Dynamics/Amersham Pharmacia Biotech). Each cDNA fragment was spotted on the slides at least four times (i.e., technical replicates). Following printing, the slides were allowed to dry and the spotted DNA was bound to the slides by UV cross-linking (50 mJ).

2.3. Sugarcane tissue samples

Two different samples (i.e., biological samples) were collected for cDNA microarray tissue profiling from leaves (LV), flowers (FL), lateral buds (LB), roots (RT), first internode (IN1), and fourth internode (IN4) of distinct plants. Five leaf samples, each from a single field grown plant, were collected and tested for field variability (LV-1, LV-2, LV-a, LV-b, LV-c). Culms of the commercial variety SP80-3280 were planted in May 2001 and May 2002 at the Copersucar Experimental Station. The first leaf with a visible dewlap (leaf+1) was collected from a 12-month-old plant for the LV-1 sample, from a 14-month-old plant for the LV-2 sample (both planted in 2001) and from 12-month-old plants for LV-a, LV-b and LV-c (planted in 2002). Two flower samples were collected from immature inflorescences (variety SP87-342) with 5 to 30 cm (FL-1) or ∼50 cm (FL-2) in length. Lateral bud and root samples derived from single-eyed seed setts were collected from 12- to 14-month-old field grown plants (variety SP80-3280). For the LB-1 and RT-1 samples, seed setts were treated with Benlate (Benomyl) 0.6 g/l and Decis (Deltamethrin) 5 ml/l, and germinated in the dark on wet paper towels for 10 days at 25°C. For the LB-2 and RT-2 samples, seed setts were planted in 200-ml plastic cups containing moist white sand and tissues and were collected after 12 days. The internode samples were collected from field grown plants of the commercial variety SP80-3280. For the IN1-1 and IN1-2 samples, the leaves were removed and the first and second internodes visible below the apical meristem were used. For the IN4-1 and IN4-2 samples, the fourth internode was collected. Also, an independent collection of leaves, flowers, lateral buds, roots, first internode and fourth internode was performed, which were used in real-time PCR assays. Tissues were sectioned, frozen in liquid nitrogen, and stored at −80°C.

2.4. RNA extraction

Frozen tissues were ground using a homogenizer. Tissue samples of 2–2.5 g were weighed and ground to a fine powder in liquid nitrogen using a pre-cooled mortar and pestle. The pulverized tissue was transferred to a 50-ml tube and homogenized with 5 ml of Trizol® (Life

Technologies) per gram of tissue, according to the manufacturer's instructions. RNA pellets were resuspended in 20 μl of warm diethyl pyrocarbonate-treated water, vortexing gently for about 15 min. RNA samples were quantified in a spectrophotometer and loaded on 1.0% agarose/formaldehyde gels for quality inspection. An equimolar pool of RNA samples of five sugarcane tissues (flower, leaf, stem, root, bud) was prepared for use as a common reference in all hybridizations.

2.5. Probe preparation and hybridization

Ten micrograms of total RNA were reverse-transcribed, labeled, and hybridized using the reagents provided with the *CyScribe Post-Labeling kit* (Amersham Biosciences), according to the manufacturer's instructions. The products of the labeling reactions were purified in Millipore Multiscreen® filtering plates to remove unincorporated labeled nucleotides. Microarrays were co-hybridized with the fluorescently labeled probes. Hybridizations were performed overnight at 42°C in humid chambers. The slides were then washed in 1×SSC and 0.2% SDS (10 min, 55°C), twice in 0.1×SSC and 0.2% SDS (10 min, 55°C), and in 0.1×SSC (1 min, at room temperature). Slides were rinsed briefly in filtered milli-Q water and dried under a nitrogen stream. Each experimental step was carefully monitored to ensure high quality of the slides and extracted data. The hybridizations were performed as displayed in Table 1.

2.6. Data acquisition, processing and statistical analysis

Slides were scanned using the *Generation III Scanner*™ (Molecular Dynamics) adjusting the photomultiplier tube (PMT) to 700 for both channels. Images were processed and data collected using the *ArrayVision* (Imaging Research Inc.) software. Local median background was subtracted from the median-based trimmed mean (MTM) density for each spot. Data from clones that generated poor quality PCR fragments (no amplification or unspecific bands) or poor quality spots (visually inspected) were excluded. The data were stored and managed by the BioArray Software environment[14] free web-based database.

A set of custom programs based on R language[15] were developed for data processing based on methods described previously[16] (available at http://verjo19.iq.usp.br/xylella/microarray/). Pearson correlation values among the leaf samples were calculated using normalized expression ratios obtained from leaf versus pool hybridizations for 1,280 genes (Table S-1, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t1.html). We used homotypic or 'self-self' hybridizations of the reference pool sample to define intensity-dependent cutoff levels that would indicate differentially expressed genes. Based on these results, eight intervals were set integrating the probability density function to 99.5% for different signal intensity levels, which were used to define differentially expressed genes in the inspected tissues. Figure S1 (http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_f1.html) shows the data from four 'self-self' hybridizations of the reference pool sample computed to establish the limits of the random variations in the SUCAST microarray experiments. The fluorescence ratios were normalized to account for systematic errors using the LOWESS fitting[17] and used to calculate the expression ratios for all genes between the tissue sample and the reference sample (Tables S-1, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t1.html and S-2, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t2.html). For every gene, the percentage of replicates within or outside the cutoff limits was calculated in each tissue sample (Table S-3, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t3.html). Genes with at least 70% of the replicate points above or below the cutoff limits were considered differentially expressed in that particular sample, while genes with 55% of the points within the cutoff were considered ubiquitous among samples.

For the clustering analysis and the visualization of a profile matrix, a single intensity value for each gene was obtained by calculating the median of all replicate points representing the same clone. Data were clustered hierarchically using the unweighted pair-group method average (UPGMA) algorithm with the euclidian distance as a measure. Further details are available at the supplementary web site (http://www.sucest-fun.org/pub/SUCAST).

2.7. Validation of microarray results by real-time PCR (RT-PCR)

Five micrograms of total RNA were treated with DNAse (Promega) according to the manufacturer's instructions and an aliquot of 7.5 μl of the treated RNA was reverse-transcribed using the *SuperScript First-Strand Synthesis System for RT-PCR* (Invitrogen). The 20-μl reverse transcription reactions contained the RNA template; 2 μl 10X RT buffer; 0.5 mM each dATP, dGTP, dCTP and dTTP; 50 ng random hexamers; 0.25 μg oligo(dT); 5 mM $MgCl_2$; 10 mM DTT (dithiothreitol); 40 U Rnase OUT; and 50 U *SuperScript II Reverse Transcriptase*. RNA, random hexamers, dNTPs, and oligo(dT) were mixed first, incubated at 70°C for 5 min and placed on ice. Subsequently, the remaining components, except the *SuperScript II Reverse Transcriptase*, were added to the reaction and the mixture was heated to 25°C for 10 min and then incubated at 42°C for 2 min. The *SuperScript II Reverse Transcriptase* was added to each tube and the reaction was incubated at 42°C for 1.5 hr, 72°C for 10 min, and chilled on ice. An identical reaction without the reverse transcriptase was performed

Table 1. cDNA microarray hybridizations performed with sugarcane tissue samples. All hybridizations were performed against a reference sample (pool of tissues composed of an equimolar mixture of flower, leaf, stem, root and bud RNA). The table indicates which CyDye was used to label each sample in each different hybridization.

flower			lateral bud			leaf			root			1st internode			4th internode		
Cy3		**Cy5**	Cy3		**Cy5**	**Cy3**		**Cy5**	**Cy3**		**Cy5**	**Cy3**		**Cy5**	**Cy3**		**Cy5**
Pool	vs.	FL-1	Pool	vs.	LB-1	Pool	vs.	LV-1	Pool	vs.	RT-1	Pool	vs.	IN1-1	Pool	vs.	IN4-1
Pool	vs.	FL-1	Pool	vs.	LB-1	Pool	vs.	LV-1	Pool	vs.	RT-1	Pool	vs.	IN1-1	Pool	vs.	IN4-2
FL-1	vs.	Pool	LB-1	vs.	Pool	Pool	vs.	LV-2	RT-1	vs.	Pool	IN1-1	vs.	Pool	IN4-1	vs.	Pool
FL-2	vs.	Pool	LB-2	vs.	Pool	LV-1	vs.	Pool	RT-2	vs.	Pool	IN1-2	vs.	Pool			
						LV-2	vs.	Pool									
						LV-a	vs.	Pool									
						LV-b	vs.	Pool									
						LV-c	vs.	Pool									

as a control to confirm the absence of genomic DNA. The cDNA product was treated with 2 U of RNAseH (Invitrogen) for 30 min at 37°C and for 10 min at 72°C. Real-time PCR reactions were performed using *SYBR Green PCR Master Mix* (Applied Biosystems) in a *GeneAmp 5,700 Sequence Detection System* (Applied Biosystems). Primers were designed using the *Primer Express 2.0* Software (Applied Biosystems). BLAST searches against the SUCEST database were conducted to ensure the specificity of the selected primers. The primer sequences designed are listed in Table S-4 (http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t4.html). Each reaction was performed in duplicate and contained 2 μl of a 1:10 dilution of the synthesized cDNA, primers to a final concentration of 600 nM each, 12.5 μl of the *SYBR Green PCR Master Mix* and PCR-grade water to a total volume of 25 μl. The parameters for the PCR reaction were 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec and 60°C for 1 min. The specificity of the amplified products was evaluated by the analysis of the dissociation curves generated by the equipment. Negative controls were also prepared in order to confirm the absence of any contamination. The ratio between the relative amounts of the target gene and the endogenous control gene in the RT-PCR reactions was determined based on the $2^{-\Delta\Delta Ct}$ method[18] with modifications. The normalized expression level was calculated as $L = 2^{-\Delta Ct}$ and $\Delta C_T = C_{T,\text{target}} - C_{T,\text{reference}}$, for each tissue. To classify a gene's distribution of expression levels among the different tissues, ranging from ubiquitous to tissue-specific, we used the entropy measure[19]:

$$H_{gene} = \sum p_t \cdot \log_6(1/p_t)$$

where $p_t = L_t / \sum L_t, L_t$ is the expression level of the gene in the t-th tissue, and the sums are taken over the six tissues.

3. Results and Discussion

3.1. The SUCAST catalogue

Plant responses to developmental and environmental signals rely on the activity of different cellular components, which detect these signals and transduce them through the cytoplasm and nucleus to trigger the appropriate metabolic answer. These signaling pathways coordinate growth and development, as well as responses to stress and pathogens. With the aim of creating a signal transduction catalogue for sugarcane we undertook a detailed survey of 43 thousand transcripts identified by the SUCEST project.[8] This EST project sequenced the 5′ and 3′ end of clones from 26 libraries prepared from 11 different sugarcane tissues and plants submitted to three stress treatments. The large sampling of many tissues allowed possibly 90% of the sugarcane expressed genes to be tagged.

We used BLAST[11] searches, Pfam[12] and SMART[13] domain analyses to identify conserved signal transduction components such as receptors, adapters, G-proteins, small GTPases, members of the two-component relay system, nucleotide cyclases, protein kinases, protein phosphatases and elements of the ubiquitination machinery and infer their putative functions. Around 2000 SASs encoding signal transduction related proteins and also 611 transcription factors were indexed in the SUCAST catalogue, which is organized into 29 categories and 409 subcategories (Table S-5, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t5.html). These elements represent 5% of the total SASs from the current SUCEST dataset. In addition, 717 SASs that might be involved in processes triggered by stress and pathogens or that may play a role in growth and development were also catalogued. Table 2 summarizes the SUCAST categories. The combined analysis of the sugarcane EST data bank, by means of an in depth annotation and gene architecture analysis, generated a catalogue with 3,563 members, which covers several aspects of signaling and transcription. It includes around 100 SASs

Table 2. The SUCAST Catalogue. The number of SASs in the catalogue and the number selected for the cDNA microarray analysis for each category are shown. For a list of all SASs refer to the Supplementary Material, Table S-5 (http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t5.html).

SUCAST classification	# SAS	# SAS in the array
Protein categories:		
Receptors	477	181
Adapters	12	9
G proteins	17	10
Small GTPases	75	36
Two component relay	19	8
Cyclase	1	0
Calcium metabolism	68	31
Inositol metabolism	46	21
Protein Phosphatases	107	36
Protein kinases	510	65
Ubiquitination	106	41
Transcription factors	611	175
Hormone biosynthesis	75	30
Hormone related	22	13
Functional categories:		
Development	30	13
Cell cycle	34	10
Stress	305	119
Pathogenicity	382	104
'No matches' and unknown proteins	548	294
Others	118	84
TOTAL	3563	1280

for hormone biosynthesis[10] and also 548 SASs with no similarities to known proteins, which were selected due to our interest in associating function to new genes.

On the basis of sequence analysis, it has been inferred that 13% of the Arabidopsis genes are involved in transcription or signal transduction.[20] The automated categorization of the SUCEST data indicated that 13.6% of the tagged genes belong to these categories.[8] With a genome size expected to be similar to the rice genome, the sugarcane genome might have around 3 thousand genes encoding putative signal transduction components.

3.2. The SUCAST cDNA microarray

To evaluate the expression profile of the SUCAST components in different sugarcane tissues we constructed glass slide cDNA microarrays with PCR products derived from 1632 cDNA clones. For 21% of the clones we could not obtain satisfactory PCR fragments and the corresponding data were removed from the analysis. Therefore, data of transcript abundance for 1,280 SASs are presented as indicated in the categories of Table 2. As a reference sample in all microarray hybridizations we used an equimolar pool of total RNA extracted from flowers, lateral buds, leaves, stems and roots.

3.3. Assessing individual variability in the field

Since sugarcane is propagated vegetatively, the genetic variability among the individuals should be low. Expression patterns obtained assaying few individuals of the same variety should be representative of a population in the field, provided that growth conditions are similar. To minimize individual differences and differences attributable to local field variations, RNA samples were typically obtained from more than one plant in our experiments. Even so, we reasoned that it would be important to evaluate whether the individual variability was as low as expected. With that purpose, we collected leaves from five different sugarcane individual stools and extracted the RNA separately. Leaves were collected in May 2002 (LV-1), July 2002 (LV-2) and May 2003 (LV-a, LV-b, LV-c). The RNA samples were labeled and hybridized to the microarrays against the reference sample (Table S-1, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t1.html). Pair-wise Pearson correlation calculations show a high correlation between leaves of the three individuals collected at the same time or within a short interval of time (p=0.84 to 0.88), and a lower correlation between individuals collected in different years (p=0.61 to 0.64) (Fig. S2, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_f2.html). The results imply sufficiently low individual variation within each sampling event, and even between close events, and indicate that pooling a large number of plants to represent a subpopulation is not necessary.

3.4. Differentially and evenly distributed genes

Total RNA samples extracted from six different sugarcane tissues were labeled and hybridized to the microarrays against the reference sample. Two different biological samples of each tissue were analyzed, and the results of at least two technical replicates were computed. Median ratio values for each gene in each sample can be found in Table S-2 (http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t2.html). Cutoff limits for differential expression were calculated based on 'self-self' hybridizations (see Methods). To estimate replicate data consistency, the expression ratio versus signal intensity data of the replicates of a given gene in different 'tissue vs. pool' hybridizations (Table 1) were studied (Fig. 1). Genes with at least 70% of the replicate points outside the cutoff limits (above or below) in both biological samples of one or more tissues were considered differentially expressed, whereas genes with more than 55% of the replicates within the cutoff limits were considered ubiquitous. Figure 1 shows four

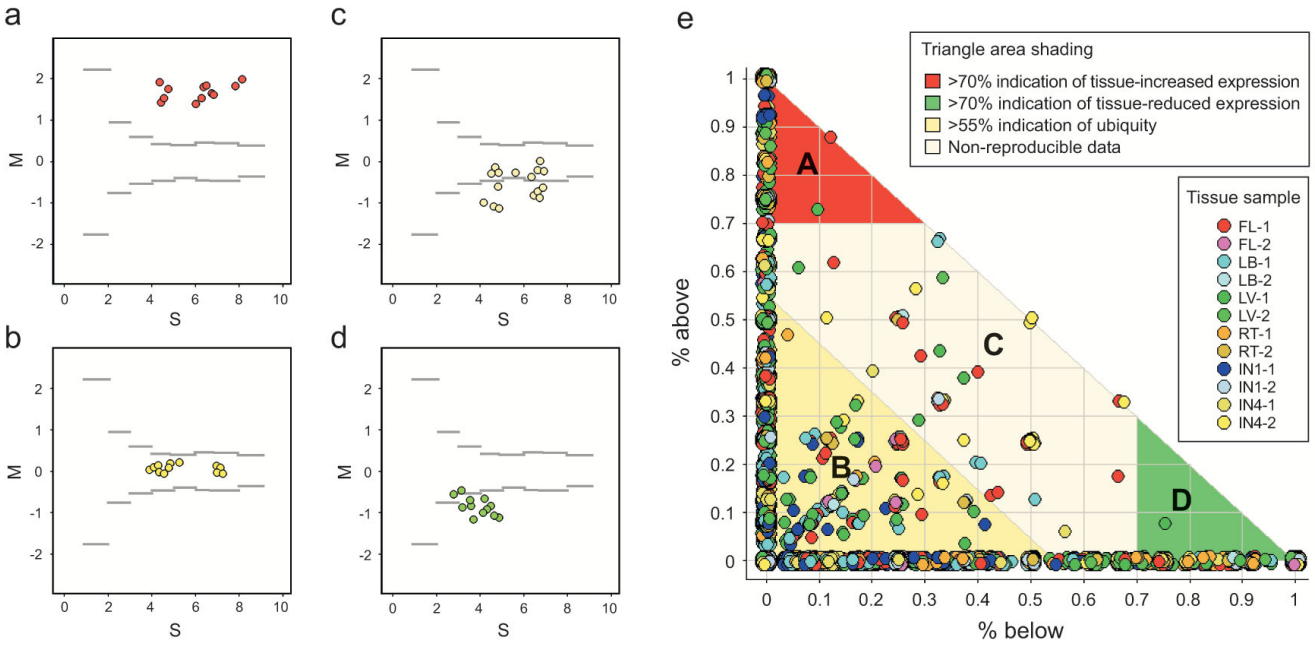

Figure 1. Differential expression consistency. a through d - Normalized log ratios against log average intensity plots, where M=$\log_2$(Cy5 / Cy3), and S=$\log_2$(0.5*(Cy5+Cy3)). The graphics correspond to examples of the data distribution of the technical replicates for a gene we considered to have tissue-increased expression (a), ubiquity or no differential expression (b), low reproducibility (c), or tissue-reduced expression (d). The bars indicate the intensity-dependent cutoffs. E - Global distribution of SUCAST microarray data. The position of each dot in the triangle relates to the percentage of reproducible replicates in each hybridization. The areas A, B, C and D contain data as exemplified in the corresponding graphics a, b, c and d.

Table 3. Distribution of differentially expressed and ubiquitous genes among sugarcane tissue samples. Genes with at least 70% of replicate points outside the cutoff limits in both biological samples of one or more tissues were considered differentially expressed, whereas genes with more than 55% of replicates within the cutoff limits in all samples were considered ubiquitous. The percentage of genes with expression above, within or below the ratio cutoff limits are indicated for each sample.

tissue	sample	% of genes above	% of genes within	% of genes below	% of genes variable	Total # of SAS analyzed
flowers	FL-1	3.19%	78.59%	8.33%	9.89%	1224
	FL-2	5.00%	80.15%	6.87%	7.97%	1179
lateral buds	LB-1	1.18%	90.10%	3.47%	5.25%	1182
	LB-2	4.61%	77.30%	5.36%	12.73%	1194
leaves	LV-1	8.84%	72.50%	9.98%	8.67%	1222
	LV-2	7.36%	76.05%	9.13%	7.46%	1073
roots	RT-1	5.52%	80.46%	7.31%	6.71%	1177
	RT-2	11.78%	61.33%	10.46%	16.43%	1205
internodes	IN1-1	3.91%	86.34%	4.58%	5.16%	1201
	IN1-2	6.20%	69.31%	5.80%	18.69%	1225
	IN4-1	4.40%	88.33%	1.91%	5.36%	1045
	IN4-2	8.26%	54.57%	10.77%	26.40%	1235
Average		5.87%	76.02%	7.05%	11.05%	1180

cases of gene expression distribution: (a) over 70% of the data points are above the cutoff limits, indicating that this gene was more expressed in the tissue being tested than in the reference; (b) all the replicates are within the cutoff limits, indicating that there was no differential expression; (c) a variable pattern among the replicates was observed, showing low reproducibility for the expression levels of the gene; (d) over 70% of the data points are below the cutoff limits, indicating that this gene is less expressed in the tissue being tested than in the reference. A graphical representation of the global distribution of the data in the M-S space, taking into account the reproducibility of the technical replicates for each gene in each hybridization is seen in Fig. 1e and summarized in Table 3. The number of genes analyzed for each tissue varied from 1045 to 1235, due to differ-

ences in the quality of some spots in different slides. The majority of the genes analyzed (avg. 76%) showed expression levels in each particular tissue similar to those of the pool of tissues. This result is in agreement with the observations of Obayashi and colleagues,[21] who have identified a large cluster of ubiquitously expressed genes after global macroarray analysis of the *Arabidopsis* transcriptome. It is important to note, however, that only part of the sugarcane transcriptome is represented in our array. Nevertheless, this could be an indication that the majority of the signaling elements in sugarcane are not differentially expressed in the different tissues analyzed. The highest percentages of preferential expression in one tissue were found in leaves (8.84% and 7.36%), in the RT-2 root sample (11.78%), and in the internode sample IN4-2 (8.26%). Likewise, a high proportion of the genes in these samples showed reduced expression in one particular tissue. The flower samples exhibited a high number of underrepresented or non-expressed genes (8.33% and 6.87%, respectively, for each of the two different biological samples). An average of 11% of the genes showed a variable pattern of expression, with high variation among the technical replicates.

For the majority of the genes present in our microarray there is little information in the literature regarding tissue distribution of transcripts. Cho and colleagues[22] performed microarray hybridizations using samples from seven different organs of maize. This approach allowed the elucidation of organ relationships and the detection of organ-specific gene expression. Recently, a comprehensive study of organ-specific gene expression has been reported for *Arabidopsis*.[21] Other reports focus on a few genes or specific metabolic routes, involving families of closely related genes such as the MADS transcription factors,[23] genes involved in particular pathways such as the acyl lipid metabolism,[24] and often rely solely on EST *in silico* data. Watson and collaborators[25] described the mapping of the proteome of the model legume barrel medic (*Medicago truncatula*). Spatial mapping of the transcriptome and proteome of diverse plant species can shed light on the regulation of many developmental pathways.

Our results indicated 217 genes that presented differential expression in both biological samples in at least one of the tissues analyzed. These elements were clustered according to their expression patterns, evidencing groups of genes with marked expression in leaves, roots, or internodes (Fig. S3, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_f3.html). Smaller groups of genes with prominent expression in flowers and lateral buds were also uncovered. Table 4 lists the SASs with enriched or reduced expression in the various tissues.

Several differentially expressed genes encode transcription factors. Among these, we detected ten genes highly expressed in flowers: a GARP transfactor, an AP2, four zinc-finger and four MADS-box domain-containing proteins. MADS and the zinc finger YABBY transcription factor play important roles during organ development and together with AP2 and zinc finger C3H proteins were shown to have enhanced expression levels in flowers.[26]

A discrete group of eight receptor genes was found to be preferentially expressed in leaves, most of which are members of the receptor-like kinase (RLK) family. We also found a sugarcane receptor possibly involved in signaling pathways in the sugarcane reproductive tissues. This SASs is very similar to MSP1 from rice and EXS/EMS1 from *Arabidopsis*. MSP1 and EXS/EMS1 are genes expressed in reproductive tissues and, among other functions, control the fate of germinative cells.[27–29] These observations indicate that this SAS possibly represents an ortholog of the MSP1 and EXS/EMS1 genes. Three SASs were found that code for putative receptors containing a protein kinase domain and a Ubox as predicted by the SMART database.[13] All of them showed a homogeneous transcript distribution among the tissues analyzed. One of them, SAS SCQSRT2031C08.g, possesses a complex structure, comprised by TPRs (tetratricopeptide repeats), low complexity regions, a pkinase domain and a Ubox domain followed by a ZnF_NFX domain (a presumed zinc binding domain) near the C-terminus. The U-box is believed to have a role in ubiquitination.[30] Protein kinases containing the U-box domain have already been reported for *Arabidopsis*.[31] However, the function of these plant proteins remains to be determined. Moreover, 13 receptors of unknown function were found to have a differential expression pattern. Six of these presented predominant expression and seven showed weak expression in at least one of the tissues. The elucidation of the expression profiles of new receptors is of great interest, since it can help in assigning putative functions to these proteins.

It is remarkable that several genes related to the ubiquitination system have been found to be more expressed in the internodes than in the other tissues examined. The ubiquitin/26S proteasome pathway[32] is implicated in selected protein breakdown, used to control the level and activity of proteins in a diverse range of metabolic routes. In sugarcane, an intense protein degradation activity in the internodes could be related to their specialization in sucrose storage.

A group of hormone-related elements, including four nitrilases and three lipoxygenases, showed prevalent expression in root tissues. The nitrilases are homologous to the *Arabidopsis nit*4 gene, which was characterized as being predominantly expressed in roots.[33] The lipoxygenases (LOX) are a functionally diverse class of dioxygenases implicated in physiological processes such as growth, senescence, and stress responses in plants, that show different organ-specific expression in different plants.[34,35] Another group of hormone-related genes composed by members of biosynthetic pathways of salicylic

Table 4. SUCAST SASs showing differential expression among different sugarcane tissues. The table lists all SASs that are found enriched or lacking in both biological samples of the various tissues. The data corresponds to the results of at least eight replicate points for each SAS. The plus sign indicates that the SAS expression is higher in the tissue than observed in the pool, while the minus sign indicates that the SAS expression is lower in the tissue than in the pool.

SAS	Putative cellular function	flower	lateral bud	leaf	root	internode1	internode4
SCCCRZ2001F07.g	Calmodulin-binding protein			+			
SCCCCL2001B01.b	Apyrase		+	-			
SCCCRZ1C01H06.g	Apyrase		+	-			
SCEZLB1012F10.g	GNGC calmodulin-binding protein	+			-		
SCCCRZ1003A03.g	HSP70s (heat shock protein)			-			
SCEPLR1051D05.g	Calreticulin			-		-	-
SCRFLR2037F09.g	Calreticulin			-		-	-
SCEPSD1069D02.g	Calreticulin			-		-	
SCEQRT2090E04.g	Calreticulin			-			
SCRULR1020D11.g	Fructose-bisphosphate aldolase			+	-		
SCCCRZ2001F03.g	Fructose-bisphosphate aldolase			-			
SCUTLR1058G02.g	Cyclin III	-	-		+	-	
SCRLLR1038H01.g	Zeaxanthin epoxidase (ABA2)		-	+	-		
SCJLLR1011C06.g	Zeaxanthin epoxidase (ABA2)					-	
SCRFLR1012D12.g	Nitrilase - Auxin biosynthesis	-	-	-	+	-	
SCCCCL6002B05.g	Nitrilase - Auxin biosynthesis	-			+		
SCEZHR1047G06.g	Nitrilase - Auxin biosynthesis				+		
SCEQRT1028H06.g	Nitrilase - Auxin biosynthesis	-		-		-	
SCVPLR2012A10.g	ACC oxidase - Ethylene biosynthesis			+			
SCJFRT1005C11.g	ACC oxidase - Ethylene biosynthesis	-					
SCSGCL6070B03.g	ACC synthase - Ethylene biosynthesis			+			
SCSBAM1085B06.g	Linoleic acid desaturase - Jasmonic acid biosynthesis	-				-	
SCUTLR2030B03.g	Lipoxygenase - Jasmonic acid biosynthesis				+	-	
SCCCRT1001E01.g	Lipoxygenase - Jasmonic acid biosynthesis	-			+		
SCJFRT1007H07.g	Lipoxygenase - Jasmonic acid biosynthesis	-			+		
SCEQRT1024E12.g	Phenylalanine ammonia-lyase (SA biosynthesis)			+	-	-	
SCSGAM1094D05.g	Phenylalanine ammonia-lyase (SA biosynthesis)				+		
SCCCLR2002F08.g	Auxin repressed/dormancy-associated			+			+
SCRLLR1038D05.g	Jasmonate responsive protein	-		+	+		
SCCCLR1001E04.g	Rubisco - small subunit			+	-		
SCRFLR1012F12.g	Caffeic acid 3-O-methyltransferase				-		+
SCCCLR1C02F07.g	Myo-Inositol-1-phosphate synthase		-		-		
SCSBFL1109C05.g	Cyclopropane fatty acid synthase				-		
SCACLR1126E09.g	No match		-	+		-	
SCBFSD2038H10.g	No match			+			
SCBFST3136A06.g	No match			+			
SCCCLR1001D09.g	No match			+			
SCRLFL3009C03.b	No match			+			
SCSBSB1051A09.g	No match			+			
SCCCCL1001H07.b	No match			+			
SCAGLR2026G12.g	No match				+		
SCRFLR2034D09.g	No match			-	-	+	
SCRFLB2056A06.g	No match				-	+	
SCCCLR1066G12.g	No match					+	
SCBGLR1023G07.g	No match				-	+	+
SCCCLR2001E10.g	No match					+	+
SCCCCL3004C02.b	No match						+
SCRUAD1063D03.g	No match						+
SCJFST1012G07.g	No match	-					
SCRUFL1022F09.g	No match (non-coding)		-	-	-	-	-
SCRLLR1111D02.g	No match (non-coding)			-		-	-
SCRUFL1017D09.g*	No match				-	-	-
SCMCAM1101G01.g	No match						-
SCUTAM2114F04.g	No match				-		
SCJFST1009H11.g	No match			-			
SCCCCL4007H07.g	No match		-				
SCEQLR1091A10.g	60S Ribosomal protein L23			-			
SCUTLR1037F12.g	60S Ribosomal protein L5			-			
SCACLR2022H05.g	Acyl carrier protein-like			-			
SCCCRZ2002C09.g	Alpha tubulin			-			
SCSBAD1084C01.g	Alpha-2 tubulin			-	-		
SCBFAD1045D12.g	Beta-glucosidase			-	-		
SCCCLR1022F10.g	Glycine hydroxymethyltransferase	-		-			
SCCCLR2002D04.g	Histone H4			-		-	-
SCCCLR2002G09.g	Histone H4				-		
SCCCRZ2001F06.g	60S ribosomal protein			-			
SCEPAM2015H01.g*	ATP-dependent DNA helicase						+
SCCCLR2002E04.g	Bet v I pollen allergen	+			-		
SCCCRT2001A03.g	Homeodomain protein			+			
SCCCRT2002B03.g	Ribosomal protein			-			
SCCCLR1067A05.g	Sugar transporter			+			
SCRUSB1062E12.g	Triacylglycerol lipase			-	-		

SAS	Putative cellular function	flower	lateral bud	leaf	root	internode1	internode4
SCCCRZ1002H08.g	Saposin B domain-containing protein				+		
SCEPLB1043F03.g	WD40 protein			-			
SCCCSD2001E05.g	Thaumatin		-			-	-
SCCCSD1003E02.g	Thaumatin		-			-	
SCVPRZ2035H07.g	R-gene Mlo			+	-		
SCCCCL3080A08.g	R-gene - NBS/LRR			+			
SCRLAD1098H03. g	R-gene - NBS/LRR				+		
SCBFAD1089A09 .g	R-gene - NBS/LRR				+		
SCQSRZ3038H06.g	R-gene - NBS/LRR					+	
SCCCLR1065F03.g	R-gene - NBS/LRR	-					
SCBGLR1023D05.g	R-genes transduction - Isd1					+	
SCEQRT1028C03.g	Pathogenicity - PR protein	-	-	-		-	-
SCQSRT1036D03.g	Pathogenicity - PR protein	-	-	-		-	-
SCACLR1036B06.g	Protein kinase/CBL-interacting	-		+			
SCEQRT2099H01.g	Protein kinase - CDPK			+			
SCQSLR1018F05.g	Protein kinase - CDPK			+			
SCCCCL4001E05.g	Protein kinase - GSK3-like				-		
SCCCRZ1003E02.g	Protein kinase - MAPK			+			
SCSGRT2064G11.g	Protein kinase - MAPK	-				+	
SCJLST1025H09.g	Protein kinase - cAMP dependent				-		
SCCCLR1076F09.g	Protein kinase - S6 ribosomal-like						+
SCEPRZ1010E06.g	Protein Phosphatases - 2C		-				
SCMCCL6048A12 .g	Protein Phosphatase 2A				+		
SCEZRZ3052A09.g	Protein Phosphatase 2A			-			
SCMCST1051F08.g	Dual Specificity Protein Phosphatases (DSPP)			+			
SCSBHR1056H08.g	Ethylene receptor EIN2		+	-			+
SCJFST1017H03.g	Receptor - NPH1-like				-		
SCRUAD1133D10.b	Photoreceptor - blue light - cry1			+			
SCAGST3138B05.g	Photoreceptor - blue light - cry1			+			
SCJFRZ027C05.g	Receptor - Clavata1			-	-		-
SCBGFL5080E07.g	EXS receptor kinase	+		-	-		
SCEZRZ1012E03.g	Leucine-rich transmembrane kinase (LTK1)			-	-		
SCQSAM2100F06.g	Leucine-rich transmembrane kinase (LTK1)			-			
SCAGLR1021B11.g	Receptor Ser/Thr kinase			+			
SCQSLR1061E06.g	Receptor Ser/Thr kinase			+			
SCSGAD1009D03.g	Receptor Ser/Thr kinase			+			
SCSGRT2065D10.g	Receptor Ser/Thr kinase			-	-	+	
SCEZAM2058E08. g	Receptor Ser/Thr kinase					+	
SCJLST1020G08.g	Receptor Ser/Thr kinase				-		
SCSBAM1086H09.g	Receptor Ser/Thr kinase				-		
SCEPLR1030G10.g	Receptor Ser/Thr kinase			-			
SCJFRT1009B11.b	Receptor Ser/Thr kinase				+		
SCVPCL6041G02.g	Receptor Ser/Thr kinase			-	-		-
SCCCLB1023E12.g	Receptor Ser/Thr kinase			-	-		
SCVPLR2005F07.g	Receptor Ser/Thr kinase			-	-		
SCJLRZ1021D12.g	Receptor Ser/Thr kinase			-			
SCEQRT1029G10.g	Receptor Ser/Thr kinase with lectin domain			+			
SCCCCL3120C09.g	Receptor Ser/Thr kinase with LysM domain						+
SCCCRZ2003C02.g	Receptor - Somatic embryogenesis			+			
SCMCST1053A06.g*	Receptor - Somatic Embryogenesis					-	
SCQGST1031D02.g	Receptor - Somatic embryogenesis				-		
SCBFSB1048C08.g	S-receptor	-		+			
SCEPAM1017E02.g	Wall-associated Ser/Thr kinase receptor	-		-			
SCCCRT2001H11.g	Small GTPase - Arf						+
SCCCLR1066D10.g	Small GTPase - Arf						+
SCJLRT1019H05.g*	Small GTPases - Arf			-			
SCCCRT1001E12.g	Small GTPase - Rab		+				
SCBGLR1117A05.g	Small GTPase -Ran			-			
SCCCAD1004H02.g	Catalase	-				-	-
SCSGRT2066D05.g	Cytochrome P450			+			
SCJLRZ1024A03.g	Cytochrome P450				+		
SCJFRT2055E06.b	Cytochrome P450 - CYP71C	-	-	-		-	-
SCCCRZ2C04A07.g	Cytochrome P450 - CYP71E				+		
SCCCAM2001B04.g	Cytochrome P450 - CYP71E					+	
SCAGLR1064G02.g	Cytochrome P450 - CYP73				+		
SCSGFL4193B05.g	Cytochrome P450 - CYP73				+		
SCAGLR1043E04.g	Cytochrome P450 - CYP74A				+		
SCEQRT1026H08.g	Cytochrome P450 - CYP75			+			
SCMCAM2084B09.g	Cytochrome P450 - CYP75			+			
SCEZRZ1012A02 .g	Cytochrome P450 - CYP90				-	+	
SCJLLB2077E09.g	Dioxygenase	-	-		+	-	
SCJFLR1074E10.g	Cysteine proteinase precursor			+	+		

SAS	Putative cellular function	flower	lateral bud	leaf	root	internode1	internode4
SCAGLR2011H07.g	Thioredoxin	-				+	
SCRFLB1056D04.g	Thioredoxin	-					
SCJFLR1013A09.g	Thiol protease					-	
SCCCAD1001C08.g	Peroxidase - P7X				+		
SCCCRZ1002F06.g	Enolase			-			
SCCCCL2001G03.b	Histone deacetylase (HOS1-like)			-			
SCUTST3084F06.g	Stress related/Low temperature induced (ßTI	-			+		
SCACCL6008H06.g	Stress related/Low temperature induced (ßTI	-					+
SCSBST3096E10.g	Aquaporin (tonoplast intrinsic protein)			-	+		
SCEQRT2100B02.g	Aquaporin (tonoplast intrinsic protein)			-	+		
SCCCLR1024C03.g	Aquaporin (tonoplast intrinsic protein)				+		
SCJLFL4186B12.g	Aquaporin (tonoplast intrinsic protein)	-					
SCCCST3001H12.g	Aquaporin (tonoplast intrinsic protein)			-			
SCQGLR1085F11.g	Dehydrin					+	
SCEZST3149E09.g	Glutathione S-transferase				+		
SCAGFL1089C03.g	Glutathione S-transferase	-		-			
SCCCCL3002C09.b	Glutathione S-transferase	-		-			
SCJLRT1020A09.g	Glutathione S-transferase	-	-			-	
SCJLHR1028C12.g	Histone H4			-	-	-	-
SCSFHR1043G09.g	S-adenosylmethionine-synthase			-			+
SCMCLR1122E10.g	NADPH-ferrihemoprotein reductase				+	-	
SCJLLR1011F11.g	Acyl-coA-oxidase				+		
SCJLRT1021D12.g	Chalconesynthase				+		
SCVPLR2027D02.g	Chalconesynthase	-	-			-	
SCUTRZ2022B12.g	Fatty acid desaturase			+			
SCCCLR1070D02.g	Photosystem II oxygen-evolving enhancer (wound-induced)			+			
SCCCCL3004C07.b	Transcription factor - ABL3/VP1					+	
SCSGFL5C05G12.g	Transcription factor - Alfin-like				-		
SCCCLR1079H06.g	Transcription factor - Alfin-like				-		
SCCCRZ1001G10.g	Transcription factor - Aux/IAA			-			
SCVPFL1132B06.g	Transcription factor - Aux/IAA			-			
SCRLRZ3041C03.g	Transcription factor - GARP			+			
SCRLAM1010D08.g	Transcription factor - knotted homeobox			-	-	+	
SCCCLR1066C06.g	Transcription factor - HSF			+			
SCCCRZ3002D03.g	Transcription factor - LIM			-			
SCEPRZ1008F02.g	Transcription factor - LIM			-			
SCCCFL6003A11.g	Transcription factor - MADS		-		-	-	
SCSGSB1007G01.g	Transcription factor - MADS	+	-		-		
SCCCLR2C03H07.g	Transcription factor - MADS				-		
SCSBSD1032G02.g	Transcription factor - MADS				-		
SCCCLR1C08G10.g	Transcription - Myb/LHY/CAA1			+			
SCCCLR2003E10.g	Transcription factor - NAC			+			
SCRLCL6032B05.g	Transcription factor - NAM			+			
SCEPLB1042D02.g	Transcription factor - NAC				+		
SCEZRT2019E05.g	Transcription factor - NAC				+		
SCCCLR1024F10.g	Transcription factor - Auxin responsive				-	-	
SCCCLR1001D10.g	Transcription factor - AP2-domain				-		
SCUTLR2023G02.g	Transcription factor - WRKY			+			
SCUTHR1065C05.g	Transcription factor - Zinc finger/ C2C2 / YABBY	+					
SCQSFL3031C01.g	Transcription factor - Zn finger C3H					-	
SCEZRZ1017C07.g	Pumilio - Translation			-			
SCCCLR1077F09.g	Response regulator (ARR-like)			+	-	+	
SCACLR1057G02.g	Response regulators (ARR-like)			+			
SCEPRT2048E07.g	Response regulators (ARR-like)					-	
SCVPCL6041F12.g	Ubiquitin-specific protease				-	-	
SCSBFL4011F05.g	Ubiquitin-conjugating enzyme E2				+		
SCACLR2014E12.g	Ubiquitin-conjugating enzyme E2						+
SCCCLR1C07A06.g	Ubiquitin-conjugating enzyme E2						+
SCCCRZ2001H09.g	Ubiquitin-conjugating enzyme E2						+
SCCCRZZC01F09.g	Ubiquitin-conjugating enzyme E2						+
SCQGLR1019D10.g	Ubiquitin-conjugating enzyme E2						+
SCCCLR1C01E03.g*	ubiquitination - F-box protein						+
SCSGLR1045F05.g	Unknown protein				+		
SCVPLR1049B12.g	Unknown protein					+	-
SCRFAM2129A08.g	Unknown protein					+	
SCBFRZ2045D11.g	Unknown protein				-	-	+
SCSBST3102C05.g	Unknown protein						+
SCSGSB1009D11.g	Unknown protein						+
SCRFLR2038D12.g	Unknown protein	-		-		-	
SCCCRZ2C03B08.g	Unknown protein	-					
SCSBSD2029F05.g	Unknown protein	-					
SCSGFL5C08F04.g	Unknown protein				-	-	
SCQGLR2032E12.g	Unknown protein				-		

* SAS identity not verified by re-sequencing

acid (phenylalanine ammonia-lyase), abscisic acid (zeaxanthin epoxidase), and ethylene (ACC oxidase, ACC synthase) biosynthetic pathways was mainly expressed in leaves.

We detected a caffeic acid 3-*O*-methyltransferase (COMT) gene expressed primarily in the fourth internode. This enzyme is involved in lignin biosynthesis and, in association with other enzymes like the CCOMT (caffeoyl CoA 3-*O*-methyltransferase), keeps in check the content and the composition of lignin in cells. A correlation between the lignin content of alfalfa internodes of progressive maturity and the activity of COMT and CCOMT has been demonstrated.[36] A sugarcane COMT has been cloned and exhibited a peak of expression in culms.[37] The SUCEST database indicates the presence of five complete sequences for this enzyme, that may represent a promising target for sugarcane genetic engineering with the aim of modifying the content and/or composition of sugarcane bagasse, allowing it to become a useful and low cost raw material for paper production and animal feed.

Among the 43,141 SASs in the SUCEST database, 35% did not show similarity to known proteins (no matches) and are therefore new genes of unknown function. We found nine no-matches to be predominantly expressed in internodes, seven in leaves, two in roots and two in flowers (Table 4). The latter are non-coding transcripts. The involvement of non-coding RNAs in floral development has been described[38–40] and thus it is possible that these sugarcane elements are involved in gene regulation during floral development. Among all the SUCEST no-matches, 2010 SASs correspond to sugarcane-specific non-coding sequences that could also contain regulatory elements.

In addition to the differentially expressed genes, we also investigated genes that showed similar expression levels in all tissues. The identification of "housekeeping" genes is of great interest in expression studies, since they are valuable experimental controls and indicate promoters useful for plant biotechnology. Among the analyzed genes, 153 presented over 55% of the replicate data points within the cutoff limits for all 12 samples analyzed (Table S-6, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t6.html). A total of 35 no-matches were found among them, an indication that these sugarcane-specific genes may have a central role in this plant's physiology.

3.5. Validation of microarray data by real-time PCR

To validate the present work, 25 SASs (Table S-4, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t4.html) were selected and analyzed by real-time PCR. To normalize the expression data, several ESTs were tested in search of a gene that showed strong and ubiquitous expression in the sugarcane tissues. An ideal reference gene has the same level of expression in all conditions under study. The commonly used tubulin gene did not show an adequate pattern, being expressed in varying levels in the tissues analyzed (Fig. 2a). The same was observed for an actin gene (not shown). Based on the number of ESTs sequenced in the SUCEST project and on the expression profile obtained from the microarray data we selected two SASs as references: SCCCLR1048F12.g (a 14-3-3 gene) and SCCCST2001G02.g (a polyubiquitin gene). These genes were found to be homogeneously expressed and adequate for normalization purposes, showing equivalent transcript levels in the tissue samples, except for a slight variation in flower tissues (for the polyubiquitin gene) and roots (for the 14-3-3 gene) (Fig. 2b,c). Therefore, all our real-time PCR data was normalized to both the 14-3-3 gene and the polyubiquitin gene. When the normalization was done using the 14-3-3 gene, expression data for the root tissue was disregarded. The same was done for the flower sample when normalizing data with the polyubiquitin gene. Although none of the reference genes tested presents absolutely the same expression levels in all tissues, the use of two different reference genes increases the reliability of the results. We considered that a gene had its expression profile validated when both results (using the 14-3-3 gene and the polyubiquitin gene as references) were consistent with the microarray data. Eighteen differentially expressed and seven ubiquitous SUCAST genes were assayed. To further confirm the expression profile obtained, the RNA samples used for the validation experiments were different from those used in the microarray hybridizations. Figure 2d-k shows the relative levels of transcripts for 8 SASs normalized to the polyubiquitin levels. The results using the 14-3-3 gene as a reference yielded essentially the same patterns (not shown).

To rank the differential expression results obtained in real-time PCR analysis, we measured the entropy (H) of the distribution of expression levels among the tissues. The entropy is widely used in information theory to measure how distant the observed distribution is from a uniform distribution.[19] Ideally, the ubiquitous genes should have a uniform distribution of expression levels in all considered tissues. According to the entropy definition, this property is mathematically translated to H closer to 1. In contrast, tissue-specific genes should have relevant expression in just one of the considered tissues and the entropy of this expression level distribution is translated to H closer to zero (note, however, that this is not a linear scale). We observed that 6 out of 7 genes expected to present a ubiquitous expression profile in fact presented an H value equal to or higher than 0.9 (Table S-4, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t4.html), indicating that they can represent real "housekeeping" genes. Four differentially expressed genes seem to be highly specific to one tissue, with H values below 0.6. Eight differentially expressed genes showed enrich-

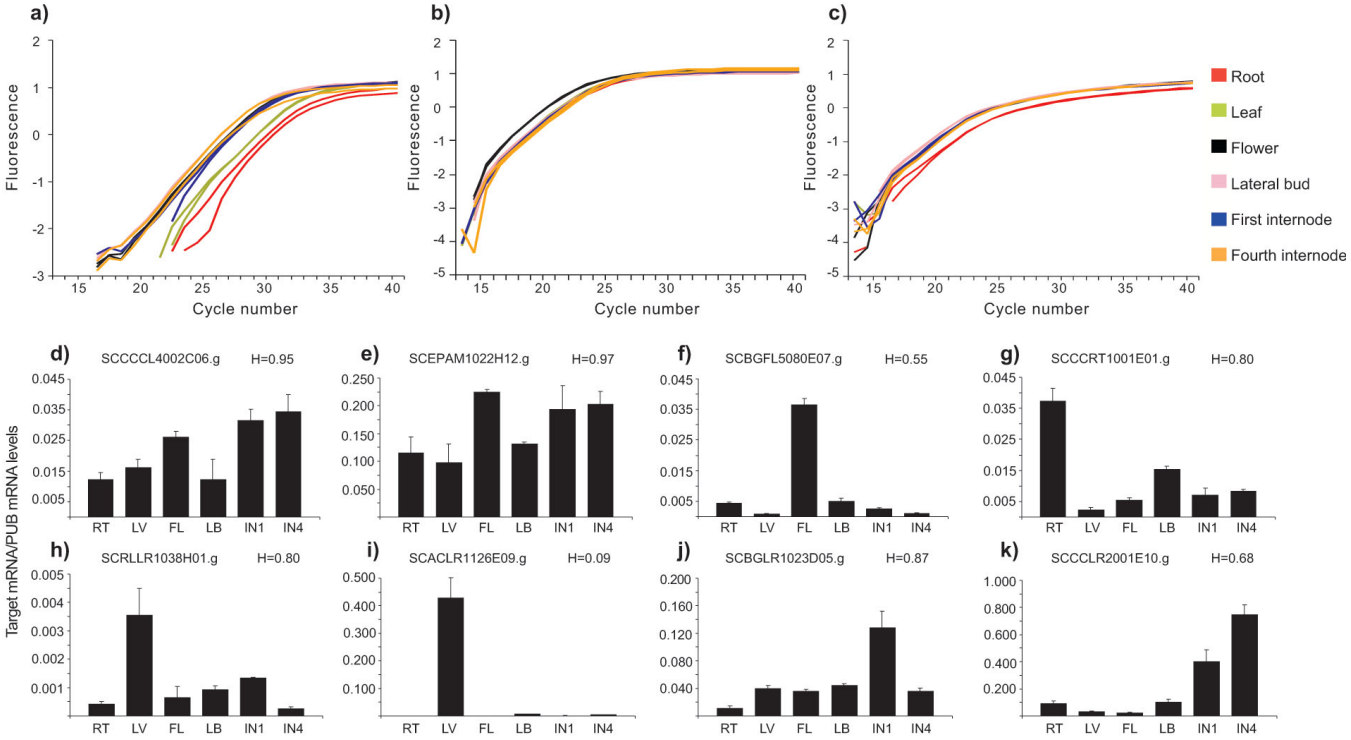

Figure 2. Validation of microarray results using real-time PCR. a through c - Raw values of the $\log_{10}$ fluorescence were plotted against the cycle number for a tubulin gene (a), a polyubiquitin gene (b), and a 14-3-3 gene (c). Each tissue analyzed is represented by a different color. All the reactions were carried out in parallel and each reaction was performed in duplicate. d through k - Real-time PCR results for a Phosphatase - PP6 / catalytic subunit (d), a Cytochrome P450 - CYP71A (e), an EXS receptor kinase (f), a lipoxygenase (g), a zeaxanthin epoxidase (h), SCACLR1126E09.g (no match) (i), LSD1 gene (j) and SCCCLR2001E10.g (no match) (k). The bars show target mRNA levels relative to the polyubiquitin mRNA. RT = root, LV = leaf, FL = flower, LB = lateral bud, IN1 = first internode, IN4 = fourth internode. The measured entropy (H) for each distribution obtained is indicated. Error bars were calculated as described by Livak and Schmittgen.[18]

ment in a particular tissue, as pointed out by the microarray data, but were also expressed at high levels in other tissues. In these cases, the H values were higher than the ones obtained for genes expressed in a single tissue, as expected, but were always lower than 0.9.

In summary, 18 out of the 25 genes tested (72%) had a profile in real-time PCR assays consistent with the differential or ubiquitous expression observed in the microarray experiments. It is important to stress that the RNA samples used in the real-time PCR experiments derived from a third biological sample, further suggesting that the data set generated in our microarray experiments is robust in indentifying ubiquitous and differentially expressed genes. The criteria used to select the differentially expressed and ubiquitous genes, although arbitrary, proved to be effective. The selection of data with at least 70% of the replicates in agreement with the cutoff for differential expression and 55% for ubiquitous expression was adequate, as shown by the high validation rate obtained. The less stringent value for ubiquity proved to be as effective probably because the genes selected had similar expression in all biological samples of all tissues.

3.6. The SUCAST expression matrix

As pointed out previously, all hybridizations were made against a common reference, consisting of a pool of tissues. When there are several samples to be compared, this strategy requires fewer hybridizations than a direct pair-wise setup, and is useful when there is no natural control (like a non-treated sample) as in treatment versus control studies. Additionally, the pool of transcripts theoretically represents the transcripts of all tissues, minimizing the occurrence of spots without a hybridization signal, for which it is not possible to calculate the expression ratio.

Although this approach allowed us to identify ubiquitous and differentially expressed genes among the sugarcane tissues, it generated relative — not absolute — information on the expression profiles. This means it evidences, for example, that a certain gene is more highly expressed in leaves than in the average of the tissues, but it does not tell us whether this same gene is more highly expressed in flowers than in roots. To get access to this type of information, we calculated "virtual ratios" between pairs of tissues using the reference values of the common sample (pool) as the common denominator.

This approach provided us with the expression pat-

terns of each individual SAS in the SUCAST microarray among all tissues analyzed (Table S-7, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t7.html and Fig. S4, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_f4.html). Additionally, the clustering of the patterns allows a spatial comparative picture of transcript abundance, which can complement the information provided by the ratio cut-off analysis (that uses 70% replication stringency levels). Using the matrix, we could note expression patterns not evident in the previous analysis. As an example, all four of the MADS transcription factors that were identified as differentially expressed due to lower expression in roots or other tissues than in the reference, are indicated by the matrix to be primarily expressed in flowers (Table S-7, http://www.dna-res.kazusa.or.jp/12/1/03/supplement/supplement_t7.html), although only SAS SCSGSB1007G01.g had been classified as such (Table 4).

3.7. Conclusions

The success of the sugarcane culture has relied for decades on traditional breeding of varieties resistant to plagues and diseases, with increased sucrose content, and more adaptable to different soils and environmental conditions, a slow and uncertain approach. Therefore, genomic data that could assist traditional breeding in the improvement of sugarcane varieties are awaited. There are very few molecular studies on sugarcane signaling response to environmental changes, and none on the distribution of these components in the different plant tissues. The comparison of the transcript complement found in six tissues using microarrays provided a spatial picture of the transcriptome of this grass, which can greatly contribute to the assignment of function to new genes. The present work focused on the identification of genes that may participate in tissue-specific activities and ubiquitous genes. The cloning of strong ubiquitous promoters or tissue-specific promoters can increase the availability of tools for sugarcane transformation and study. The identification of genes highly expressed in stems or leaves could also help in the understanding of metabolic pathways involved in sugar production and accumulation, and could constitute targets for crop improvement. The described signaling elements are currently being studied in search for candidates that might regulate hormone responses, the accumulation of sucrose in the stalk, and the response to several biotic and abiotic stresses allowing us to step forward in the efficient manipulation of sugarcane varieties. The knowledge accumulated on the role for signal transduction processes in the regulation of stress and pathogenesis responses brings the SUCAST components to center stage in the search for genes that might be modified to obtain plants with desired traits.

Acknowledgements: This work was funded by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo). We are indebted to Adriana Y. Matsukuma and Denise Yamamoto for technical assistance on microarray printing and scanning performed in the laboratory of the Cooperation for Analysis of Gene Expression (CAGE) inter-departmental Project, Dr. Sergio Verjovski-Almeida and Dr. Hugo Aguirre Armelin for their support throughout the development of this work, Apuã C. M. Paquola for valuable help providing bioinformatic tools and Dr. Jesus Ferro for coordinating the SUCEST cDNA clone collection in the beginning of this work. AMdS was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, (CNPq). FSPT, FRR, RZNV, KCO and ACQS were supported by FAPESP fellowships.

References

1. Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. 2002a, Evolution of protein kinase signaling from yeast to man, *Trends Biochem. Sci.*, **27**(10), 514–520.
2. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. 2002b, The protein kinase complement of the human genome, *Science*, **298**(5600), 1912–1934.
3. Chen, W., Provart, N. J., Glazebrook, J. et al. 2002, Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses, *Plant Cell*, **14**(3), 559–574.
4. Gribskov, M., Fana, F., Harper, J. et al. 2001, PlantsP: a functional genomics database for plant phosphorylation, *Nucleic Acids Res.*, **29**, 111–113.
5. Rives, A. W. and Galitski, T. 2003, Modular organization of cellular networks, *Proc. Natl. Acad. Sci. USA*, **100**(3), 1128–1133.
6. Lin, J., Qian, J., Greenbaum, D. et al. 2002, GeneCensus: genome comparisons in terms of metabolic pathway activity and protein family sharing, *Nucleic Acids Res.*, **30**, 4574–4582.
7. McCarty, D. R. and Chory, J. 2000, Conservation and innovation in plant signaling pathways, *Cell*, **103**(2), 201–209.
8. Vettore, A. L., da Silva, F. R., Kemper, E. L. et al. 2003, Analysis and Functional Annotation of an Expressed Sequence Tag Collection for the Tropical Crop Sugarcane, *Genome Res.*, **13**, 2725–2735.
9. Arruda, P. 2001, Sugarcane transcriptome. A landmark in plant genomics in the tropics, *Genetics and Mol. Biol.*, **24**, 1–296.
10. Souza, G. M., Simoes, A. C. Q., Oliveira, K. C. et al. 2001, The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane, *Genetics and Mol. Biol.*, **24**(1–4), 25–34.
11. Altschul, S. F., Madden, T., Schaffer, A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.

12. Sonnhammer, E. L. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. 1998, Pfam: multiple sequence alignments and HMM-profiles of protein domains, *Nucleic Acids Res.*, **26**, 320–322.
13. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. 1998, SMART, a simple modular architecture research tool: identification of signaling domains, *Proc. Natl. Acad. Sci. USA*, **95**(11), 5857–5864.
14. Saal, L. H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, Å., and Peterson C. 2002, Bioarray software environment: A platform for comprehensive management and analysis of microarray data, *Genome Biol.*, **3**, software0003.1–software0003.6.
15. Ihaka, R. and Gentleman, R. 1996, R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, **5**, 299–314.
16. Koide, T., Zaini, P. A., Moreira, L. M. et al. 2004, DNA microarray-based genome comparison of a pathogenic and a non-pathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence, *Journal of Bacteriology*, **186**, 5442–5449.
17. Yang, Y. H., Dudoit, S., Luu, P. et al. 2002, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Res.*, **30**(4), e15.
18. Livak, K. J. and Schmittgen, T. D. 2001, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method, *Methods*, **25**(4), 402–408.
19. MacKay, D. 2003, Probability, entropy, and inference. In: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, UK.
20. Arabidopsis Genome Initiative. 2000, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature*, **408**, 796–815.
21. Obayashi, T., Okegawa, T., Sasaki-Sekimoto, Y. et al. 2004, Distinctive features of plant organs characterized by global analysis of gene expression in Arabidopsis, *DNA Res.* **11**(1), 11–25.
22. Cho, Y., Fernandes, J., Kim, S. H., and Walbot, V. 2002, Gene-expression profile comparisons distinguish seven organs of maize, *Genome Biol.*, **3**(9), research0045.
23. Alvarez-Buylla, E. R., Liljegren, S. J., Pelaz, S. et al. 2000, MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes, *Plant J.*, **24**(4), 457–466.
24. Beisson, F., Koo, A. J. K., Ruuska, S. et al. 2003, Arabidopsis Genes Involved in Acyl Lipid Metabolism. A 2003 Census of the Candidates, a Study of the Distribution of Expressed Sequence Tags in Organs, and a Web-Based Database, *Plant Physiol.*, **132**, 681–697.
25. Watson, B. S., Asirvatham, V. S., Wang, L., and Sumner, L. W. 2003, Mapping the Proteome of Barrel Medic (*Medicago truncatula*), *Plant Physiol.*, **131**, 1104–1123.
26. Henning, L., Gruissem, W., Grossniklaus, U., and Kohler, C. 2004, Transcriptional programs of early reproductive stages in *Arabidopsis*, *Plant Physiol.*, **135**,1765–1775.
27. Zhao, D. Z., Wang, G. F., Speal, B., and Ma, H. 2002, The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther, *Genes Dev.*, **16**(15), 2021–2031.
28. Canales, C., Bhatt, A. M., Scott, R., and Dickinson, H. 2002, EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*, *Curr. Biol.*, **12**(20), 1718–1727.
29. Nonomura, K., Miyoshi, K., Eiguchi, M. et al. 2003, The MSP1 gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice, *Plant Cell*, **15**(8), 1728–1739.
30. Aravind, L. and Koonin, E. V. 2000, The U box is a modified RING finger — a common domain in ubiquitination, *Curr Biol.*, **10**, R132–134.
31. Azevedo, C., Santos-Rosa, M. J., and Shirasu, K. 2001, The U-box protein family in plants, *Trends Plant Sci.*, **6**(8), 354–358.
32. Vierstra, R. D. 2003, The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins, *Trends Plant Sci.*, **8**(3), 135–142.
33. Bartel, B. and Fink, G. R. 1994, Differential regulation of an auxin-producing nitrilase gene family in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA*, **91**(14), 6649–6653.
34. Kolomiets, M. V., Hannape, D. J., Chen, H., Tymeson, M., and Gladon, R. J. 2001, Lipoxygenase is involved in the control of potato tuber development, *Plant Cell*, **13**(3), 613–626.
35. Porta, H. and Rocha-Sosa, M. 2002, Plant lipoxygenases. Physiological and molecular features, *Plant Physiol.*, **130**(1), 15–21.
36. Inoue, K., Sewalt, V. J., Murray, G. B., Ni, W., Sturzer, C., and Dixon, R. A. 1998, Developmental expression and substrate specificities of alfalfa caffeic acid 3-*O*-methyltransferase and caffeoyl coenzyme A 3-*O*-methyltransferase in relation to lignification, *Plant Physiol.*, **117**(3), 761–770.
37. Selman-Housein, G., Lopez, M., Hernandez, D. et al. 1999, Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification, *Plant Sci.*, **143**, 163–171.
38. Schmid, M., Uhlenhaut, N. H., Godard, F. et al. 2003, Dissection of floral induction pathways using global expression analysis, *Development*, **130**, 6001–6012.
39. Aukerman, M. J. and Sakai, H. 2003, Regulation of Flowering Time and Floral Organ Identity by MicroRNA and Its APETALA2-Like Target Genes. *The Plant Cell*, **15**, 2730–2741.
40. Chen, X. 2004, A MicroRNA as a Translational Repressor of APETALA2 in *Arabidopsis* Flower Development, *Science*, **303**, 2022–2025.