Genetics and Molecular Biology, 24 (1-4), 25-34 (2001)

The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane

Glaucia Mendes Souza, Ana Carolina Quirino Simoes, Katia Cristina Oliveira, Humberto Miguel Garay, Leonardo Costa Fiorini, Felipe dos Santos Gomes, Milton Yutaka Nishiyama-Junior and Aline Maria da Silva*

Abstract

EST sequencing has enabled the discovery of many new genes in a vast array of organisms, and the utility of this approach to the scientific community is greatly increased by the establishment of fully annotated databases. The present study aimed to identify sugarcane ESTs sequenced in the sugarcane expressed sequence tag (SUCEST) project (http://sucest.lad.ic.unicamp.br) that corresponded to signal transduction components. We also produced a sugarcane signal transduction (SUCAST) catalogue (http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm) that covered the main categories and pathways. Expressed sequence tags (ESTs) encoding enzymes for hormone (gibberellins, ethylene, auxins, abscisic acid and jasmonic acid) biosynthetic pathways were found and tissue specificity was inferred from their relative frequency of occurrence in the different libraries. Whenever possible, transducers of hormones and plant peptide signaling were catalogued to the respective pathway. Over 100 receptors were found in sugarcane, which contains a large family of Ser/Thr kinase receptors and also photoreceptors, histidine kinase receptors and their response regulators. G-protein and small GTPases were analyzed and compared to known members of these families found in mammalian and plant systems. Major kinase and phosphatase pathways were mapped, with special attention being given to the MAP kinase and the inositol pathway, both of which are well known in plants.

INTRODUCTION

The analysis of the complete *Arabidopsis thaliana* genome sequence (The *Arabidopsis* Genome Initiative, 2000) has revealed the striking conservation of genetic mechanisms required for developmental and physiological processes while pointing to the unique properties of individual plant systems. Comparison of the *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Arabidopsis thaliana* genomes indicates that many signal transduction modules are conserved and even though the signals and end results of the pathways may be different many components are shared by these very diverse organisms (McCarty and Chory, 2000; The *Arabidopsis* Genome Initiative, 2000). The differences in the structure of plant hormones in comparison with animal hormones has led to the assumption that plants probably evolved an array of regulatory molecules that were totally different from those of animals, although the identification of signal transduction components in many plant systems has contradicted this view. In the work presented in this paper we describe putative signal transduction components of sugarcane revealed by a systematic mining of the whole EST data set generated in the SUCEST project (http://sucest.lad.ic.unicamp.br). The main categories were catalogued as well as the number of their related ESTs clusters, paving the way for the functional analysis of signal transduction pathways in sugarcane.

METHODS

The SUCEST database stores over 250,000 quality controlled 5' and 3' ESTs reads from 37 cDNA libraries prepared from several sugarcane tissues (calli, root, stalk, leaves, flowers, developing seed, etc.) grown under different environmental conditions (Vettore *et al.*, 2001). The ESTs were organized by sequence similarity into clusters and singletons and automatically annotated according to their similarities to sequences in the National Center for Biotechnological Information (NCBI) non redundant protein database (Telles *et al*., 2001). The basic local alignment search tool (BLAST) was used to perform a bi-directional search against the SUCEST clusters consensi generated using the CAP3 algorithm (Telles and da Silva, 2001) using query sequences from known signal transduction components or domains and the TBLASTN algorithm (Altschul *et al*., 1997). Hits with E-values in most cases lower than 10^{-10}, were manually inspected after previous automatic BLASTX annotation which had been performed by the SUCEST bioinformatics team. Positive matches were aligned using the multiple sequence alignment CLUSTAL method (Jeanmougin *et al*., 1998) and

Instituto de Química, Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900 São Paulo, SP, Brazil. Send correspondence to Glaucia Mendes Souza. E-mail: glmsouza@iq.usp.br.

their amino acid translated sequences were compared with protein families and domains in the PROSITE (Hofmann *et al*., 1999) and Protein Family (PFAM) (Bateman *et al*., 2000) databases. If 100% of the composition of their ESTs reads were derived from the same set of cDNA libraries the clusters were considered tissue-specific, whereas if only 80% of the reads belonged to a specific library they were considered tissue-enriched. The sequences of clusters and their identities are available at the SUCEST web site at http://sucest.lad.ic.unicamp.brand and at the SUCAST web site at http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm). The main protein categories so far catalogued and the number of related clusters found are shown in Table I, supplementary Information is available at http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm.

RESULTS AND DISCUSSION

Plant hormones and signaling peptides

Plant hormones and peptides transduce signals such as temperature, light, water, nutrient and microbe-plant interactions which induce cellular responses locally and/or throughout the plant. We have investigated the existence of the major sugarcane routes for the synthesis of ethylene, abscisic acid, auxins, gibberellins and jasmonates. The enzymes for which clusters have been found are shown in Figure 1.

Structurally, the simplest plant hormone is the gas ethylene, which has numerous roles including plant development, sex determination, fruit ripening, flower and leaf senescence and defense (Johnson and Ecker, 1998). Ethylene is synthesized from S-adenosyl methionine (AdoMet) to make 1-aminocyclopropane-1-carboxylic acid (ACC) which is converted to ethylene, CO_2, and HCN by ACC oxidase. Four putative ACC synthases were found in sugarcane and 6 ESTs clusters were identified as being similar to ACC oxidases. This was expected since these enzymes are encoded by multigene families in several plant species. ACC synthase genes are regulated by developmental signals, hormones and environmental stimuli. Control of ethylene synthesis is largely attributed to ACC synthase but altered expression patterns of ACC oxidase (ACO) genes in senescence, fruit ripening and wounding suggests that the latter contribute to regulation of ethylene production as well (Johnson and Ecker, 1998). The 4 clusters corresponding to ACC synthase ESTs are made up of 14 reads of which 9 come from root or root zone transition libraries, indicating that there are higher levels of this enzyme in roots. Ethylene has been implicated in the production of root hairs but to our knowledge increased levels of ACC synthase in roots has not yet been detected.

Indole-3-acetic acid (IAA) is the major naturally occurring auxin and has been implicated in the regulation of growth and development of many plant species. Two major routes have been described for IAA biosynthesis, the first being a tryptophan-dependent pathway where Trp is converted to indole-3-acetaldoxime (IAOx) and then via indole-3-acetonitrile (IAN) to IAA, and a second Trp-independent pathway which has not, as yet, been very well characterized (Hull *et al*., 1999). In the tryptophan-dependent pathway the first step is the conversion of Trp to IAOx catalyzed by cytochrome P450 (CYP79B2 and CYP79B3) leading to the conversion of IAN to IAA by nitrilases (NIT1, NIT2, NIT3 and NIT4). The nitrilase related clusters (a total of seven) were also present at higher levels in root tissue, and one of them was specific to libraries from sugarcane infected with *Glucoacetobacter diazotrophicus*. A well established effect of auxin stimulation is the induction of ACC synthase (an early-auxin-response gene) leading to ethylene biosynthesis. The presence of many reads of ACC synthase in roots together with the prevalence of ESTs related to auxin synthesis in this organ leads us to think that the pathway leading from auxin to ethylene production may occur predominantly in sugarcane roots. The auxin receptor is unknown but auxin binding to the plasma membrane elicits the activation of selective protein degradation by the ubiquitin-proteosome pathway. Ubiquitin conjugation requires the sequential activity of three protein complexes, ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2) and ubiquitin-protein ligase (E3) (del Pozo and Estelle, 2000). Putative candidates for transducers of the auxin response have been found, including E1, E2 and E3 complexes and their candidate targets, the IAA domain proteins. IAA homology domains are found in a large family of auxin induced proteins, with possibly over 30 members in sugarcane.

Gibberellins (GA) play key roles in plant growth and development, mediating light-stimulated seed germination through phytochromes (Kamiya and Garcia-Martinez, 1999), mobilization of reserves by aleurone cells (Lovegrove and Hooley, 2000), leaf expansion, stem elongation, flower initiation, and flower and fruit development (Sun, 2000). In the major pathway of gibberellin synthesis trans-geranyl diphosphate (GGGP) is converted to ent-copalyl diphosphate (CPP) by ent-copalyl diphosphate synthase (CPS) and then to ent-kaurene by ent-kaurene synthase (KS) (Hedden and Phillips, 2000), these two enzymes have few clusters (CPS has one cluster and KS two) which occur most often in flowers. Sequential oxidations catalyzed by the enzymes ent-kaurene 19-oxidase, ent-kaurenoic acid 7-B-hydroxilase and GA12-aldehyde synthase produce GA12-aldehyde. Reads with similarity to the ent-kaurene 19-oxidase GA3 from *Arabidopsis* were found in several sugarcane tissues (flowers, internode, apical meristem, roots and stem bark) indicating widespread distribution in sugarcane, much as was seen in *Arabidopsis* (Helliwell *et al.*, 1998). Mono-oxygenases catalyze the

conversion of GA12-aldehyde to GA12 and GA53, which are substrates for the biosynthesis of gibberellins. The bio-active gibberellins GA1, GA4, GA3 and GA7 are formed by the 2-oxoglutarate-dependent dioxygenases, GA 20-oxidase and GA 3-β-hydroxylase. The dioxygenase genes are major targets for light regulation of GA metabo-

Table I - Catalogued sugarcane signal transduction components.

Category	Gene or gene family similarities	Number of clusters*
Ethylene synthesis	ACC synthase	4
	ACC oxidase	6
ABA synthesis	Zeaxanthin epoxidase	2
	9-cis-Epoxycarotenoid dioxygenase	3
Auxin synthesis	Nitrilases	7
Giberellin synthesis	CPS - copalyl diphosphate synthase	1
	KS (ent-kaurene synthase)	2
	GA3 (ent-kaurene oxidase)	2
	GA20ox	1
	GA3h	1
	GA2ox	2
Jasmonate synthesis	Linoleic acid desaturase	8
	Lipoxygenase	10
	Allene oxide synthase	3
	Allene oxide cyclase	2
	12-oxo-phytodienoate reductase	6
Receptors	Serine/threonine receptor kinases	93
	G-protein coupled receptors	1
	Photoreceptors	
	Phytochromes	4
	Blue Light receptors	6
	Histidine kinase-like receptors	
	Ethylene receptor	6
	Others	2
	Cytokinin receptor	5
Related to the two-component system	Phosphorelay intermediates ATHP1, 2 and 3	3
	Response regulators	
	ARR1, 2, 11	4
	ARR3, 4, 5, 6, 7	9
	Pseudo response regulator	1
Ubiquitination	E1	8
	E2	43
	E3	7
	SKP1	4
	F-box protein	11
	Poly-ubiquitin	29
	IAA proteins	30
G-proteins	α-subunit	3
	β-subunit	12
	γ-subunit	1
Small-GTPases	Ras	none
	Rab	28
	Rop/Rac	4
	Ran	7
	Arf	10
GTPases regulators	Rac GAP	3
	Ran GAP	5
	Rab GAP	none
	Rho GAP	1
	Rho GDI	4
Cyclases	Adenylyl cyclase	none
	Guanylyl cyclase	none
Non receptor protein kinases	MAPK	7
	MAPKK	5
	MAPKKK	17
	cAMP dependent protein kinase - regulatory	none
	cAMP dependent protein kinase - catalytic	1
	cGMP-dependent protein kinase	none
	Calcium dependente protein kinase	25
	Glycogen synthase kinase (MSK)	13
	Casein kinase	16
	CLB interacting kinase (CIPK)	2
Protein phosphatases	PPP family	
	Ser/thr protein phosphatase 1/catalytic (PP1c)	7
	Ser/thr protein phosphatase 2A/catalytic (PP2Ac)	11
	Ser/thr protein phosphatase 4/catalytic (PP4c)	2
	Ser/thr protein phosphatase 5/catalytic (PP5c)	1
	Ser/thr protein phosphatase 6/catalytic (PP6c)	1
	Ser/thr protein phosphatase 7/catalytic (PP7c)	2
	Undefined PPP catalytic subunit	6
	PP2A regulatory subunits	21
	PPM family	
	PP2C	11
	Kinase associated protein phosphatase	1
	Tyrosine-specific protein phosphatase	3
	Dual-specificity protein phosphatase	4
Inositol	PtdIns-3 kinase	2
	PtdIns-4 kinase	3
	PtdIns4P-kinase	13
	Ins-polyP-5-phosphatase	12
	Ins-1(or 4)-monophosphatase	2
	Phospholipase C	9
	myo-Inositol 4-O-methyltransferase	7
	myo-Inositol 2-dehydrogenase	1
	myo-Inositol-1-phosphate synthase	1
Plant peptides	Enod40	1
Calcium	Calmodulin	7
	Calreticulin	9
	Calnexin	2
	Calcium channel	1
	Calmodulin and cyclic nucleotide regulated cation channel	11
	Calcineurin B-like	6

*Number of clusters found with high similarity. For clusters sequences and identification visit the SUCEST web site at http://sucest.lad.ic.unicamp.br and the SUCAST web site at http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm.
Key: ACC (1-aminocyclopropane-1-carboxylic acid); GA (giberellin); ABA (abscisic acid); PtdIns (phosphatidyl inositol).

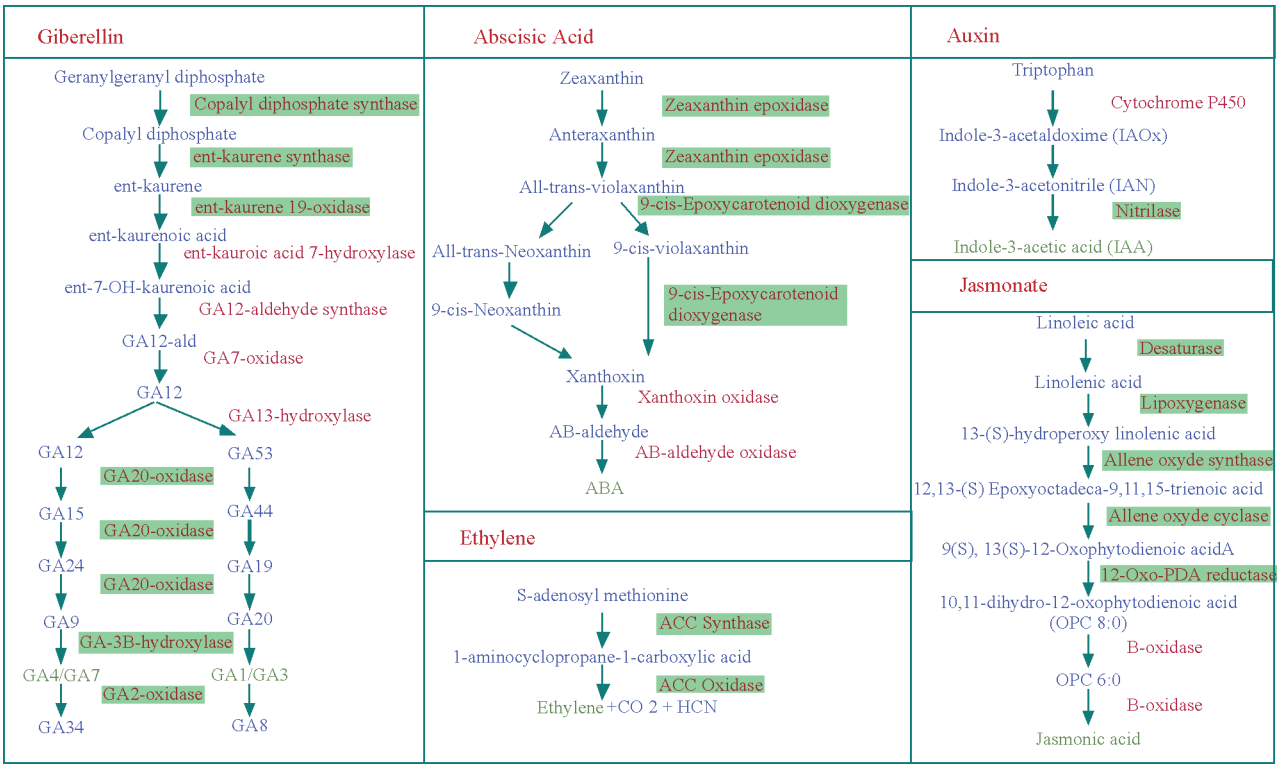

Figure 1 - Major routes of hormone biosynthesis. The enzymes for which corresponding clusters have been found in sugarcane are highlighted in green.

lism (Kamyia and Garcia-Martinez, 1999). One putative cluster with similarity to both GA 20-oxidase and GA 3-β-hydroxylase was found in sugarcane, with no apparent tissue specificity observed. The dioxygenase GA 2-oxidase, which deactivates gibberellins also was detected in sugarcane (2 clusters). Gibberellins are detected at the plasma membrane but no receptors have been cloned. There is evidence of increased cytosolic calcium and induction of calmodulin levels following stimulation of aleurone cells by gibberellins and that G proteins (see below) transduce gibberellin signals (Lovegrove and Hooley, 2000).

Plant peptides are involved in various signaling pathways, including cellular communication in meristems mediated by CLAVATA3 and phytosulfokines (PSKs), *Rhizobium* nodule formation (Enod40) and defense responses (systemin) (Bisseling, 1999). The CLAVATA3 peptide, together with its receptor CLAVATA1, coordinates the growth and differentiation of meristem cells (Fletcher *et al*., 1999). The CLAVATA1 gene encodes a putative receptor kinase that controls the size of shoots and floral meristems in *Arabidopsis* (Clark *et al*., 1997). Although we did not find any CLAVATA3-related sugarcane ESTs, we did find a large number of clusters for receptor serine/threonine kinases similar to CLAVATA1 (see below). Phytosulfokines are peptides, made up of 4-5 amino acids, which stimulate cell proliferation (Matsubayashi *et al*., 1999). A cDNA clone encoding a PSK precursor has been isolated from rice, but no PSK-related ESTs clusters were detected in the SUCEST database. Likewise no significantly similar clusters were found for systemin, a defense polypeptide released by some plants upon damage. Enod40 is one of the first plant genes activated during *Rhizobium* induced nodule formation. It is induced in the root pericycle, and the only sugarcane ESTs cluster found that had any similarity to this gene was derived from the leaf/root-zone transition library. Receptors for PSK and Enod40 have not yet been identified in any system.

Abscisic acid (ABA) is a cleavage product of carotenoids which is produced by plants in response to stress (Cutler and Krochko, 1999). Water stress induces accumulation of ABA, presumably from the cleavage of 9-cis-epoxycarotenoids (9-cis-violaxanthin and 9-cis-neoxanthin) to xanthoxin by the enzyme 9-cis-epoxycarotenoid dioxygenase. The 9-cis-epoxycarotenoids are produced from zeaxanthin by the enzyme zeaxanthin epoxidase. ABA seems to counteract the increase of calmodulin and calcium levels in aleurone signaling by gibberellins (Lovegrove and Hooley, 2000) but ABA receptors have not been defined yet. No significant enrichment of ABA biosynthesis clusters were found in the sugarcane cDNA libraries.

Jasmonate is a 12-carbon fatty acid-derivative (synthesized from linoleic acid via the octadecanoid pathway) which has a role in the defense against different types of microbial pathogens (Reymond and Farmer, 1998). Most of the enzymes involved in the synthesis of jasmonate appear to be expressed in sugarcane and some ESTs clusters for the 12-oxo-phytodienoate (PDA) reductase were found in libraries derived from plantlets infected with

Glucoacetobacter or *Herbaspirillum*. This agrees with the proposed role for jasmonate in the plant response to biotic and abiotic stresses as suggested by Reymond *et al*. (2000).

Receptors

Analysis of the *Arabidopsis* genome sequence has indicated that the largest and most diverse family of receptors in plants is the receptor serine/threonine kinase family, which has over 300 genes (The *Arabidopsis* Genome Initiative, 2000). We found 93 clusters similar to receptor kinases in sugarcane, of which 6 contain leucine-rich-repeat (LRR) domains, including one cluster for a receptor serine/threonine kinase containing a lectin domain and another cluster for a receptor serine/threonine kinase containing a tetratricopeptide repeat (TPR) domain.

It appears that plants have evolved different signaling pathways compared to mammals and other metazoans since no receptor tyrosine kinase or evidence of the Ras pathway has been found in plants (McCarty and Chory, 2000). It is not surprising then that we did not find any EST clusters related to tyrosine kinases or Ras (see below), and only one cluster related to the G-protein coupled receptor family.

The two-component histidine kinase pathway transduces ethylene and cytokinin signaling (Urao *et al*., 2000). Our analysis revealed 6 histidine kinase clusters similar to the ethylene receptor and 5 similar to the cytokinin receptor. Moreover, we catalogued 3 clusters corresponding to phosphotransfer intermediate proteins and 13 clusters related to response regulators.

In *Arabidopsis* there are two cryptochromes and five phytochrome photoreceptors. They overlap in function and transduce the blue and far-red light which regulates gibberellin synthesis (Kamiya and Garcia-Martinez, 1999), inhibition of hypocotyl elongation, anthocyanin production and the sensitivity of flowering to the photoperiod (Cashmore *et al*., 1999). The phytochromes are an interesting family of proteins with light-dependent serine/threonine-specific kinase activity. It has been proposed that these photoreceptors have evolved from ancestral histidine kinases (McCarty and Chory, 2000). We have found 4 ESTs clusters similar to phyA-D and 6 clusters similar to the blue light receptors.

G-protein and small GTPases

The current sugarcane EST data set appears to contain clusters similar to the α, β and γ subunits of the G-protein, three clusters for the α-subunit, twelve clusters for the β-subunit and one cluster for the γ-subunit. The sequencing of the *Arabidopsis* genome has confirmed the existence of a single gene for each of the G-protein α and β-subunits and recent studies have identified the only γ-subunit (The *Arabidopsis* Genome Initiative, 2000). Our findings contrast with previous observations that in plants G-protein subunits are not members of large gene families, indicating an increased number of these transducers in sugarcane.

In animal cells the Ras superfamily of small guanosine triphosphatases (GTPases) is categorized into the Ras, Rab, Arf, Ran and Rho families according to their guanosine triphosphate (GTP) binding domain, effector and insertion sequences. Plants do not appear to contain members of the Ras family (McCarty and Chory, 2000). Accordingly we did not find any cluster related to this group in sugarcane. The most predominant small GTPase family found in the SUCEST database was the Rab family (Table II) with 28 EST clusters mapping to this family. The most predominant member of the Rho family in plants are the Rac (or Rops) GTPases (Valster *et al*., 2000) but only four clusters, for RacA, RacB and RacC from *Zea mays*, and no *bona fide* Rho (*i.e.* one with a characteristic LKCD GTP-binding domain) were found in the SUCEST database. The second largest group of small GTPases which we found in sugarcane belonged to the Arf group with 13 EST clusters occurring in the SUCEST database. We also found 7 Ran family clusters. In general, the clusters were most similar to their *Oriza* and *Zea* counterparts, with the Rab family being the most diverse family with additional members similar to those found in *Lycopersicum* and *Arabidopsis*. The GTP-binding proteins are regulated by GTPase activating proteins (GAPs), GDP dissociation inhibitors (GDIs) and GTP-exchange factors (GEFs). We found several GAPs and GDIs in sugarcane but no GEFs of the Dbl-type, indicating that only regulation by inhibition can be inferred for these proteins so far.

Second messengers

Inositol signaling in plants has been shown to play a role in cell growth and elongation, mediating membrane trafficking and calcium levels (Stevenson *et al*., 2000). Production of inositol triphosphate (Ins(1,4,5)P3)) is a common response to salt and hyper-osmotic stress in plants as well as to the effects of gravity (gravi-stimulation). A search for the enzymes involved in inositol metabolism in sugarcane indicated the pathways shown in Figure 2. Several clusters related to these enzymes showed tissue specificity or were enriched in the root, root-zone transition, flower or infected plant libraries. No inositol triphosphate receptor was found in sugarcane, nor, to our knowledge, in any other plants, suggesting that the plant and animal receptors might not share much sequence similarity. Cyclic ADP ribose (cADPR) has also been shown to trigger the release of calcium from the intracellular compartments of plants, but we did not find any EST clusters similar to ADP-ribosyl cyclases or cyclic ADP-ribose hydrolases in the SUCEST database.

In our analyses, we found no EST clusters related to proteins containing a guanylate cyclase domain or to cyclic

Table II - Catalogue of sugarcane small GTP-binding proteins[1].

Similar to rab				28 clusters
Effector loop	GTP binding domain	Similar to	Organism[3]	# clusters [2]
TIGIDF	NKAD	Ethylene-responsive	Le	3
TIGIDF	NKVD	Ethylene-responsive	Le	1
TIGVEF	NKAD	GTP binding ras-like	At	4
TIGVEF	NKCD	Rab11e	Ct	2
TIGVEF	NKCD	Rab2	Ss	1
TIGVEF	NKID	Rab11c	At	2
TIGVDF	NKCD	YPTM1	Zm	1
TIGVDF	NKSD	YPTM2	Zm	3
TIGVEF	NKSD	Ras related Ric2	Os	2
TIGVEF	NKSD	Ras related RGP1	Os	1
TIGVEF	NKSD	Ras related RGP2	Os	2
TIGVEF	NKSD	Rab11d	Ct	2
TIGVDF	NKVD	GTP-binding protein	At	1
TVGASF	NKAD	Rab5B	Os	1
TIGVDF	NKVD	Rab18	At	1
TIGVDF	NKCD	ORAB	Do	1
Similar to ran				7 clusters
GTP-binding domain I	GTP binding domain II	GTP-binding domain III	Organism[3]	# clusters[2]
GDGGTGKTTFV	DTAGQ	NKVD	Zm	1
GDGGTGKTTFV	DTAGQ	NKVD	Os	3
GDGGTGKTTFV	DTAGQ	NKVD	Ca	1
Undefined GTP-binding domains			At	2
Similar to rac (rop)				4 clusters
Effector loop	GTP domain	Insert region	Similar to	Organism[3]	# clusters[2]
NKFPTDYIPTVFDNFSANVVV	TKLD	HYLMDHPGLVPV	RacA	Zm	1
NKFPTDYIPTVFDNFSANVSV	TKLD	QFFVDHPGAVPI	RacB	Os, Zm	1
NKFPTDYIPTVFDNFSANVSV	TKLD	SYLADHSAASI	RacC	Zm	1
NKFPTDYIPTVFDNFSANVSV	TKLD	AYLADHPGASTI	RacC	Zm	1
Similar to arf					10 clusters
GTP-binding domain I	GTP-binding domain II	GTP-binding domain III	Similar to	Organism[3]	# clusters[2]
GLDAAGKT	DVGGQ	NKQD	Arf1	At	1
GLDAAGKT	DVGGQ	NKQD	Arf1	Os	1
GLDAAGKT	DVGGQ	NKQD	Arf	Os	1
GLDAAGKT	DVGGQ	incomplete	Arf ?	Os	1
GLDAAGKT	DVGGQ	NKQD	ESTArf	Os	2
GLDAAGKT	DVGGQ	NKQD	Arf like	At	1
GLDAAGKT	DVGGQ	NKQD	Arf3	At	1
GLDNAGKT	DLGGQ	NKQD	Arf3/Arf1	At/Os	1
incomplete	DVGGQ	incomplete	Arf	Ca	1

[1]The domains described characterize each small GTP-binding protein family.
[2]Numbers of clusters for each class.
[3]Also shown are the most similar protein hit and corresponding organism (At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Zm, *Zea mays*; Le, *Lycopersicon esculentum*; Do, *Dyscopige ommata*; Ss, *Sporobolus stapfianus*; Ct, Common tabaco; Ca, *Capsicum annum*).
For clusters sequences and identification visit the SUCEST web site at http://sucest.lad.ic.unicamp.brand and the SUCAST web site at http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm.

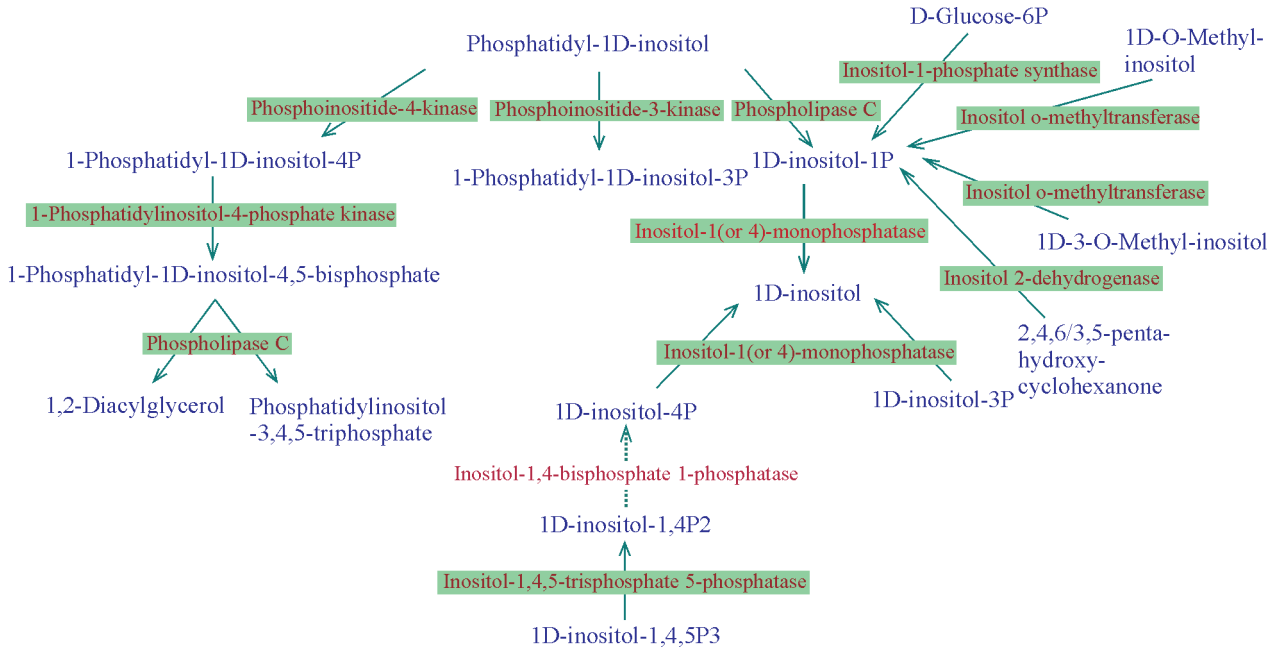

Figure 2 - Inositol metabolism in sugarcane. The enzymes for which clusters have been found in sugarcane are highlighted in green. Also shown are the inositol intermediate compounds predicted to exist.

guanidine monophosphate (cGMP) dependent protein kinases. We did find one putative cluster with sequence similarity to the catalytic subunit of the cyclic adenosine monophosphate (cAMP) dependent protein kinase (PKA) but no clusters for the regulatory subunit (R subunit) or adenylyl cyclase were found. These cAMP and cGMP regulatory components seem to be absent in sugarcane but cyclases and R subunits typical of animal systems do not seem to be present in other plant species as well.

An early response to many physiological stimuli (light, cold, heat, movement, hypoxia, drought, phytohormones and pathogens) include an elevation in the level of free calcium in the cytosol. Plasma membrane and vacuolar Ca^{2+} channels have been implicated in multiple signaling pathways involving calcium-binding proteins and calcium dependent protein kinases (CDPKs). We identified a few sugarcane clusters corresponding to calcium channels and calmodulin regulated channels, some of which were enriched in or specific to flower libraries. Clusters for the calcium binding proteins calmodulin, calreticulin and calnexin were also found. Sugarcane also express genes for the new family of calcium sensors called calcineurin B-like proteins (CLB) and we detected 6 EST clusters related to these proteins. It has been shown that *Arabidopsis* express distinct members of a CBL gene family and one of them (AtCBL-1) is strongly induced by stress signals (Kudla *et al*., 1999). In *Arabidopsis*, some members of the CBL family appear to interact with a subset of a family of protein serine/threonine kinases called CLB interacting protein kinases (CIPKs) while other CBLs are thought to target distinct proteins (Kim *et al*., 2000). Two CIPK-related clusters were found in the SUCEST database.

Kinases and phosphatases

A strong and common theme of signaling networks in eukaryotic organisms is regulation by reversible protein phosphorylation. The importance of protein kinases and protein phosphatases in the control of most of the pathways involved in growth and development as well as in responses to changes in the environment is reflected in the large number of genes which encode these enzymes in eukaryotes (Hunter, 1995). It has been reported that the *Arabidopsis thaliana* genome encodes nearly 1000 genes belonging to the protein kinase superfamily and almost 300 genes encoding protein phosphatases (The *Arabidopsis* Genome Initiative, 2000). As expected, a large number (so far 80 clusters have been annotated) of sugarcane EST clusters related to protein kinases were found in the SUCEST database, including receptor-like protein kinases and two-component histidine kinases (Table I). However, in accordance with previous observations for other plant species (McCarty and Chory, 2000), sugarcane appears to lack typical receptor tyrosine kinases.

The MAP kinase (MAPK) cascade is the most studied phosphorylation pathway in plants and appears to transduce a vast array of signals (Figure 3). The input receptors include the histidine kinases and receptor-like serine/threonine kinases. The MAP kinases are targets of ethylene through the two-component system leading to changes in gene expression mediated by EIN3 (Johnson and Ecker, 1998), and a MAPKKK has been shown to be associated with a Ras-like small GTPase Rop in the signal transduction of the CLAVATA1 receptor-like kinase in meristem signaling (Valster *et al*., 2000). As shown in Figure 3, other

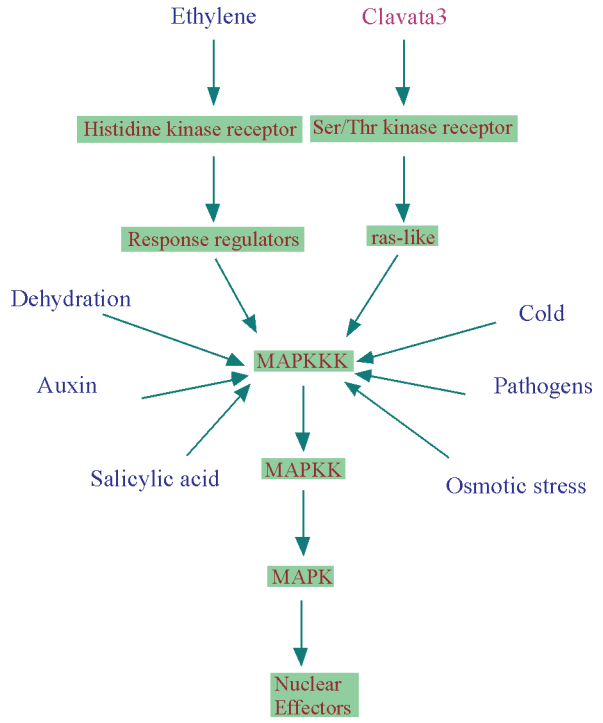

Figure 3 - Signals mediated by the MAP kinase module. The dashed lines represent undefined pathways. The components for which clusters have been found in sugarcane are highlighted in green.

signals transduced by the MAP kinase cascade include plant hormones and many environmental signals, however their complete pathways are still undefined.

Protein phosphatases that de-phosphorylate phosphoserine/threonine residues are encoded by the PPP and PPM gene families, which have distinct amino acid sequences and crystal structures. Members of the PPP family usually exist, *in vivo,* as multimeric holoenzymes where a limited number of catalytic subunits are largely controlled by the nature of the associated regulatory subunit (Barford, 1996; Cohen, 1997). We were able to assign 30 clusters to catalytic subunits of the PPP family, covering its major members with the exception of calcineurin (PP2B). To our knowledge the PP2B catalytic subunit has not yet been detected in other plants, despite the existence of the calcineurin regulatory subunit-like proteins known as CLBs (see above).

It may be possible that the number of genes encoding PPP catalytic subunits in sugarcane may actually be larger, and that their ESTs are clustered together due to the striking sequence conservation among members of this family. In addition, we found at least 19 EST clusters encoding regulatory subunits for members of the PPP family (including B-type PP2A regulatory proteins) which serves to remind us of the highly conserved regulation strategy of this family of enzymes.

PP2C are monomeric magnesium-dependent enzymes classified in the PPM family, several members of the PP2C group being related to the ABI1/ABI2 *Arabidopsis* enzymes implicated in the negative regulation of the abscisic acid pathway (Merlot *et al*., 2000). In the SUCEST database, we found 11 sugarcane EST clusters with sequences very similar to PP2C. One of them (from the apical meristem and flower libraries) was significantly similar to a kinase associated protein phosphatase (KAPP), an enzyme related to the PP2C family that has been found to be part of a signaling complex involving the CLAVATA1 receptor and Rho GTPase-related protein (Trotochaud *et al*., 1999). It is tempting to speculate that a similar complex might exist in sugarcane.

No tyrosine specific kinase has been found in higher plants or yeast. However it has been demonstrated (Zhang and Klessig, 1997) that plant MAPK activation follows its phosphorylation by MAPK kinases, a major group of dual-specificity kinases that phosphorylate MAPKs at both threonine and tyrosine, similar to what is seen in mammalian and yeast cells. Moreover it has been demonstrated that plants express tyrosine phosphatases (PTP) and that *Arabidopsis* PTP1 is encoded by a stress-responsive gene (Fordham-Skelton *et al*., 1999; Xu *et al*., 1998). We found three sugarcane EST clusters that appear to encode PTP-like enzymes and nine clusters for a subfamily of PTPs known as dual specificity phosphatases (DSPP) which have also been implicated in the negative regulation of an *Arabidopsis* MAPK (Gupta *et al*., 1998).

Signal transduction of plant-microbe interactions

Plants are constantly attacked by a wide variety of microorganisms and have developed an array of responses to either survive pathogen attacks or, in the case of endophytes, to profit from these interactions. An effective response depends on sensing and transducing a particular microorganism's presence, leading to a specialized gene expression response that, for example, confers disease resistance on the plant. A number of resistance genes are induced by salicylic acid, ethylene and jasmonic acid when plants are exposed to pathogens. Jasmonate, for instance, has been shown to be essential for the defense of tomato against hornworm larvae and *Arabidopsis* against flies and fungal attacks (Reymond and Farmer, 1998) inducing expression of defensins.

To begin to access the signaling mechanisms that may be involved in sugarcane-microbial interactions, we performed a search for signal transduction components specifically expressed when sugarcane plantlets were infected with *Herbaspirillum rubrisubalbicans* (the HR cDNA library) and *Glucoacetobacter diazotroficans* (the AD cDNA library). Both bacteria are diazotrophic endophytes that present a unique type of association with sugarcane, *H. rubrisubalbicans* appearing to cause mottled stripe disease in susceptible sugarcane cultivars (Reinhold-Hurek and Hurek, 1998). In our libraries, though, the *H. rubrisubalbicans* strain used for infection was non-pathogenic for the host sugarcane variety used (Vettore *et al.*, 2001).

From the 650 signal transduction-related ESTs clusters we have so far catalogued, 23 were specifically found

only in AD and HR libraries. The analysis of these clusters revealed that an enzyme involved in the jasmonate synthesis pathway, 12-oxo-phytodienoate reductase, was specific for these libraries, indicating that jasmonate synthesis is probably induced under these conditions. Ethylene signaling components were also represented in the AD and HR specific clusters. Two histidine kinase receptor clusters similar to the ethylene receptor and a response regulator were found specifically in the infected libraries. Six receptor serine/threonine kinases were specific to the AD or HR libraries, one cluster being similar to a wall-associated kinase from *Arabidopsis* induced when these plants are exposed to pathogens and postulated to protect *Arabidopsis* against the attacks by microbial pathogens (He *et al*., 1998). We also detected one cluster encoding an authentic type 1 serine/threonine phosphatase catalytic subunit (PP1c) which appeared to be AD and HR specific.

Inositol signaling is prominent in clusters specific to the AD and HR libraries, where we found a 1-phosphatidylinositol-4-phosphate kinase, a inositol(myo)-1(or 4)-monophosphatase and two phospholipases C, but calcium signaling is probably also involved, as indicated by the presence of one calnexin and one calreticulin.

It appears that G-protein coupled sensing is also involved in the transduction of plant-microbe interaction signaling because a G-protein β-subunit was specifically induced in the sugarcane tissues experimentally infected with *H. rubrisubalbicans* and *A. diazotroficans*.

This preliminary attempt to define the signal transduction components induced when diazotrophic endophytes associate with sugarcane suggests that the plant actively participates in this process instead of behaving as a silent host for the growth of these bacteria. Even so, the inferences we have drawn are based on comparing several different non-infected tissues with infected plantlets, and a wider gene expression analysis of sugarcane infected with these endophytes will be necessary to prove these assumptions.

Future perspectives

So far the current sugarcane EST collection (SUCEST) has enabled the identification of over 650 clusters for signal transduction components, the analysis of which indicates that most of the signaling modules typical of plants are conserved. The mining of the SUCEST database for signal transduction components is an on-going effort, and if the same number of components is found in sugarcane as has been found in *Arabidopsis,* we may expect at least 5 thousand genes to be included in this category. Many multi-gene families have also been detected, the receptor serine/threonine kinases being the most striking example with almost 100 members. The work reported in this paper summarizes the catalogued clusters and main signal transduction pathways that can be found at the SUCAST Web site (http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm). It is hoped that this resource will aid future functional analysis of the sugarcane genome.

RESUMO

O sequenciamento de ESTs (etiquetas de sequencias transcritas) tem possibilitado a descoberta de muitos novos genes em uma ampla variedade de organismos. Um aumento do aproveitamento desta informação pela comunidade científica tem sido possível graças ao desenvolvimento de base de dados contendo seqüências completamente anotadas. O trabalho aqui relatado teve como objetivo a identificação de ESTs de cana de açúcar seqüenciadas através do projeto SUCEST (http://sucest.lad.ic.unicamp.br) que codificam para proteínas envolvidas em mecanismos de transdução de sinal. Nós também preparamos um catálogo dos componentes de transdução de sinal da cana de açúcar (SUCAST) englobando as principais categorias e vias conhecidas (http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm). ESTs codificadoras de enzimas envolvidas nas rotas de biossíntese de hormônios (giberelinas, etileno, auxinas, ácido abscísico, ácido jasmônico) foram encontradas e sua expressão específica nos tecidos foi inferida a partir de seu enriquecimento nas diferentes bibliotecas. Quando possível, transmissores do sinal hormonal e da resposta a peptídeos produzidos pela planta foram associados a suas respectivas vias. Mais de 100 receptores foram encontrados na cana de açúcar, entre os quais uma grande família de receptores Ser/Thr quinase e também de fotoreceptores, receptores do tipo histidina quinase e seus respectivos reguladores da resposta. Proteínas G e GTPases pequenas foram também analisadas e comparadas com membros destas famílias já conhecidos em mamíferos e plantas. As vias principais que envolvem a participação de proteínas quinases e fosfatases foram mapeadas, em especial as vias da quinase MAP quinase e do inositol que são bem estudadas em plantas.

ACKNOWLEDGMENTS

This work is supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). A.C.Q.S and L.C.F. are FAPESP fellows. H.M.G. was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). G.M.S and A.M.S are partially supported by CNPq.

REFERENCES

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

Altschul, S.F., Madden, T., Schaffer, 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.

Barford, D. (1996). Molecular mechanisms of protein serine/threonine phosphatases. *Trends Biochem. Sci. 21*: 407-412.

Bateman, A., Birney E., Durbin, R., Eddy, S.R., Howe, K.L. and **Sonnhammer, E.L.L.** (2000). The Pfam Protein Families Database. *Nucleic Acids Res. 28*: 263-266.

Bisseling, T. (1999). The role of plant peptides in intercellular signaling. *Curr. Opin. Plant Biol. 2*: 365-368.

Casmore, A.R., Jarillo, J.A., Wu, Y-J. and **Liu, D.** (1999). Cryptochromes: Blue light receptors for plant and animals. *Science 284*: 760-765.

Clark ,S.E., Williams, R.W. and **Meyerowitz, E.M.** (1997). The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell 89*: 575-585.

Cohen, P.T.W. (1997). Novel protein serine/threonine phosphatases: variety is the spice of life. *Trends Biochem. Sci. 22*: 245-251.

Cutler, A.J. and **Krochko, J.E.** (1999). Formation and breakdown of ABA. *Trends Plant Sci. 4*: 472-478.

del Pozo, J.C. and **Estelle, M.** (2000). F-box proteins and protein degradation: an emerging theme in cellular regulation. *Plant Mol. Biol. 44*: 123-128.

Fletcher, J.C., Brand, U., Running, M.P., Simon, R. and **Meyerowitz, E.M.** (1999). Signaling of cell fate decisions by CLAVATA3 in *Arabidopsis* shoot meristems. *Science 283*: 1911-1914.

Fordham-Skelton, A.P., Skipsey, M., Eveans, I.M., Edwards, R. and **Gatehouse, J.A.** (1999). Higher plant tyrosine-specific protein phosphatases (PTPs) contain novel amino-terminal domains: expression during embryogenesis. *Plant Mol. Biol. 39*: 593-605.

Gupta, R., Huang, Y., Kieber, J. and **Luan, S.** (1998). Identification of a dual-specificity protein phosphatase that inactivates a MAP kinase from *Arabidopsis*. *Plant J. 16*: 581-589.

He, Z.H., He, D. and **Kohorn, B.D.** (1998). Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. *Plant J. 14*: 55-63.

Hedden, P. and **Phillips, A.L. (**2000). Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci. 5*: 523-530.

Helliwell, C.A., Sheldon, C.C., Olive, M.R., Walker, A.R., Zeevaart, J.A., Peacock, W.J. and **Dennis, E. S.** (1998). Cloning of the *Arabidopsis* ent-kaurene oxidase gene GA3. *Proc. Natl. Acad. Sci. USA. 95*: 9019-9024.

Hofmann, K., Bucher, P., Falquet, L. and **Bairoch, A.** (1999). The PROSITE database, its status in 1999. *Nucleic Acids Res*. *27*: 215-219.

Hull, A.K., Vij, R. and **Celenza, J.L.** (1999). *Arabidopsis* cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proc. Natl. Acad. Sci. USA. 97*: 2379-2384.

Hunter, T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling. *Cell 80*: 225-236.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and **Gibson, T.J.** (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci. 23*: 403-405.

Johnson, P.R. and **Ecker, J.R.** (1998). The ethylene gas signal transduction pathway: a molecular perspective. *Annu. Rev. Genet. 32*: 227-254.

Kamiya, Y. and **Garcia-Martinez, J.L.** (1999). Regulation of gibberellin biosynthesis by light. *Curr. Opin. Plant Biol. 2*: 398-403.

Kim, K.N, Cheong, Y.H., Gupta, R. and **Luan, S.** (2000). Interaction specificity of *Arabidopsis* calcineurin B-like calcium sensors and their target kinases. *Plant Physiol. 124*: 1844-1853.

Kudla, J., Xu, Q., Harter, K., Gruissem, W. and **Luan, S.** (1999). Genes for calcineurin B-like proteins in *Arabidopsis* are differentially regulated by stress signals. *Proc. Natl. Acad. Sci. USA*. *96*: 4718-4723.

Lovegrove, A. and **Hooley, R.** (2000). Gibberellin and abscisic acid signaling in aleurone. *Trends Plant Sci. 5*: 102-110.

Matsubayashi, Y., Mrita, A., Matsunaga, E., Fruya, A., Hanai, N. and **Sakagami, Y.** (1999). Physiological relationships between auxin, cytokinin, and a peptide growth factor, phytosulfokine-a in stimulation of asparagus cell proliferation. *Planta 207*: 559-565.

McCarty, D.R. and **Chory, J.** (2000). Conservation and innovation in plant signaling pathways. *Cell 103*: 201-209.

Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A. and **Giraudat, J.** (2001). The ABI1 and ABI2 protein phosphatase 2C act in a negative feedback regulatory loop of the abscisic acid signaling pathway. *Plant J. 25*: 295-303.

Reinhold-Hurek, B. and **Hurek, T.** (1998). Life in grasses: diazotrophic endophytes. *Trends Microbiol. 6*: 139-144.

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*: 707-720.

Reymond, P. and **Farmer, E.E.** (1998). Jasmonate and salicylate as global signals for defense gene expression. *Curr. Opin. Plant Biol. 1*: 404-411.

Stevenson, J.M., Perera, I.Y., Heilmann, I., Persson, S. and **Boss, W.F.** (2000). Inositol signaling and plant growth. *Trends Plant Sci. 5*: 252-258.

Sun, T. (2000). Gibberellin signal transduction. *Curr. Opin. Plant Biol. 3*: 374-380.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. and **Martin, G.B.** (1999). Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell 11*: 15-29.

Telles, G. P., Braga, M. D. V., Dias, Z., Lin, T-L., Quitzau, J. A. A., da Silva, F. R. and **Meidanis, J.** (2001). Bioinformatics of the sugarcane EST project. *Genetics and Mol. Biol. 24* (1-4): 9-15.

Telles, G. P. and **da Silva, F. R.** (2001). Trimming and clustering sugarcane ESTs. *Genetics and Mol. Biol. 24* (1-4): 17-23.

Trotochaud, A.E., Hao, T., Wu, G., Yang, Z. and **Clark, S.E.** (1999). The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. *Plant Cell 11*: 393-405.

Urao, T., Yamaguchi-Shinozaki, K. and **Shinozaki, K.** (2000). Two-component systems in plant signal transduction. *Trends Plant Sci. 5*: 67-74.

Valster, A.H., Hepler, P.K. and **Chernoff, J.** (2000). Plant GTPases: the Rhos in bloom. *Trends Cell Biol. 10*: 141-146.

Vettore, A. L., da Silva, F. R., Kemper, E. L., and **Arruda, P.** (2001). The libraries that made SUCEST. *Genetics and Mol. Biol. 24* (1-4): 1-7.

Xu, Q., Fu, H-H., Gupta, R. and **Luan, S.** (1998). Molecular Characterization of a Tyrosine-Specific Protein Phosphatase Encoded by a Stress-Responsive Gene in *Arabidopsis*. *Plant Cell 10*: 849-857.

Zhang, S. and **Klessig, D.F.** (1997). Salicylic acid activates a 48-kD MAP kinase in tobacco. *Plant Cell 9*: 809-824.