UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



**Sandra Rodrigues de Camargo**

# "Identificação de genes e uso de promotores modulados por etanol em cana-de-açúcar"



Tese apresentada ao Instituto de Biologia para obtenção do Título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento.

Orientador: Prof. Dr. Marcelo Menossi Teixeira

Co-Orientador: Dr. Eugênio César Ulian

**Campinas, 2007**

**FICHA CATALOGRÁFICA ELABORADA PELA**
**BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

**C14e**

Camargo, Sandra Rodrigues de
Identificação de genes e uso de promotores modulados por etanol em cana-de-açúcar / Sandra Rodrigues de Camargo. – Campinas, SP: [s.n.], 2007.

Orientador: Marcelo Menossi Teixeira.
Co-orientador: Eugênio César Ulian.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Cana-de-açúcar. 2. Etanol. 3. Expressão gênica. 4. Regiões promotoras (Genética). 5. Arranjos de DNA. I. Teixeira, Marcelo Menossi. II. Ulian, Eugênio César. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

**Título em inglês:** Identification of genes and use of promoters regulated by ethanol in sugarcane.
**Palavras-chave em inglês:** Sugarcane; Ethanol; Gene expression; Promoter regions (Genetics); DNA arrays.
**Área de concentração:** Genética Vegetal e Melhoramento.
**Titulação:** Doutora em Genética e Biologia Molecular.
**Banca examinadora**: Marcelo Menossi Teixeira, Sabrina Moutinho Chabregas, Adriana Natalício Capella, Vicente Eugênio de Rosa Junior, Michel George Albert Vincentz.
**Data da defesa**: 23/08/2007.
**Programa de Pós-Graduação:** Genética e Biologia Molecular.

Campinas, 23 de agosto de 2007

**BANCA EXAMINADORA**

Prof. Dr. Marcelo Menossi Teixeira (Orientador)

Assinatura

Profa. Dra. Sabrina Moutinho Chabregas

Assinatura

Profa. Dra. Adriana Natalício Capella

Assinatura

Prof. Dr. Michel George Albert Vincentz

Assinatura

Dr. Vicente Eugênio de Rosa Junior

Assinatura

Dr. Ricardo Augusto Dante

Assinatura

Profa. Dra. Gláucia Mendes Souza

Assinatura

Profa. Dra. Anete Pereira de Souza

Assinatura

**Dedico**

Aos meus pais, Benedito Luiz e Terezinha, e ao meu irmão Marcelo – as pessoas mais importantes em minha vida.

*Agradeço ao Prof. Dr. Marcelo Menossi pela orientação durante o desenvolvimento deste trabalho.*

*Aos membros da banca examinadora Dra. Adriana Capella, Dr. Ricardo Dante, Dr. Vicente De Rosa Jr. e Dr. Michel Vincentz pelas críticas e sugestões.*

*Aos pesquisadores do Centro de Tecnologia Canavieira (CTC) pela amizade e pelo fornecimento do material vegetal utilizado neste trabalho. Dr. Eugênio Ulian pela co-orientação; a Dra. Sabrina Chabregas pela contribuição durante o exame de qualificação e pré-banca e a Dra. Maria Cristina Falco pelo auxílio nas análises das plantas.*

*À toda equipe do Laboratório Genoma Funcional, meus companheiros de bancada, Juliana, Kiara, Renata, Agustina, Ligia, Vanúsia, Graziela, Lucia, Ariane, Rafael, Marcelo, Fábio, Edna, Eduardo Mariano, Aline, Renato, Rodrigo, Paulino, Jorge, Melina, Carolina, Michele.*

*Aos pesquisadores e amigos do CBMEG que muito me ajudaram e contribuíram para a minha formação Fábio Nogueira, Vicente, Eduardo Kiyota, Silvia, Mario, Fernanda Soardi.*

*Ao Geraldo Cançado, meu ex-companheiro de bancada, porém companheiro para a vida toda. Obrigada por toda a ajuda, pelo encorajamento na busca novos desafios, pelo carinho e compreensão.*

*À minha amada família por estar sempre ao meu lado. Obrigada pela confiança, pelo incentivo, pelo apoio e, principalmente, por todo o amor e carinho. Pai, Mãe e Marcelo, vocês são maravilhosos!*

*À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida.*

A cana-de-açúcar é uma cultura de grande importância social e econômica para o Brasil. Além da produção de açúcar, que coloca o país como o maior produtor mundial, a cana-de-açúcar tem alcançado grande destaque na produção de energia renovável e pouco poluente, o etanol. Devido ao rápido crescimento das áreas cultivadas e do aumento no número de indústrias processadoras da cana-de-açúcar, milhares de novos empregos têm sido gerados no Brasil. Por se tratar de uma cultura tão importante para a sociedade e economia brasileira, a cana-de-açúcar vem ganhando cada vez mais destaque e atenção das instituições de pesquisa públicas e privadas. Grande parte da pesquisa e experimentação desenvolvida atualmente para esta cultura visa o desenvolvimento de variedades mais adaptadas às condições edafoclimáticas do Brasil e mais resistentes e tolerantes contra pragas e doenças. Outro importante campo de estudo que tem sido bastante focado no momento é a compreensão dos mecanismos bioquímicos da síntese de sacarose com a finalidade de aumentar a produção deste açúcar e conseqüentemente de etanol.

O desenvolvimento de um sistema genético capaz de modificar o metabolismo da planta, através de um estímulo artificial, pode ser muito útil no aumento da produtividade e qualidade ou na redução dos custos envolvidos com a produção de cana-de-açúcar. Esta tecnologia poderá contribuir significativamente para um avanço ainda maior da cana-de-açúcar no país. A partir desta premissa, o presente estudo buscou identificar e caracterizar um sistema de indução de expressão gênica disparado pela aplicação externa de etanol em folhas de cana-de-açúcar.

Para isso, fazendo uso da técnica de macroarranjos de cDNA, analisamos a expressão gênica de 3.575 ESTs, isolados de folhas de cana-de-açúcar, em resposta a aplicação de etanol. Setenta ESTs apresentaram padrão de expressão alterado pelo etanol. Dentre estes, foram identificados ESTs que codificavam para proteínas relacionadas ao processo de transcrição e tradução e estresse abiótico. Muitos genes cuja função ainda permanece desconhecida também foram identificados nesta análise. Dos 70 ESTs identificados, 48 tiveram expressão induzida pela aplicação de etanol.

A partir dos ESTs selecionados como induzidos pelo tratamento com etanol, foi possível selecionar o gene *ERD* (early responsive dehydration) como candidato para identificação e caracterização da sua seqüência promotora. O EST do gene *ERD* apresentou

um perfil de expressão interessante, com expressão basal baixa durante os tratamentos controle e indução da expressão após a primeira hora de tratamento com etanol.

Estudos realizados com várias espécies de plantas demonstram que o sistema *alc* é um sistema eficiente de expressão gênica baseado na indução por etanol. A partir desta informação, desenvolvemos um vetor de expressão contendo o promotor *alcA* regulando a expressão do gene repórter da β-glucuronidase e em seqüência, o promotor constitutivo *Ubi*-1 de milho controlando a expressão do gene *alcR*, um fator de transcrição indispensável para a ativação do sistema. Posteriormente, esta construção foi utilizada para transformação genética de plantas de cana-de-açúcar que no momento estão sendo selecionadas e analisadas quanto a introgressão estável do cassete gênico.

Outra etapa do presente estudo foi realizada com o gene *SsNAC23*, já descrito como induzido por estresse de frio, estresse hídrico e herbivoria. Neste trabalho foi demonstrado que *SsNAC23* também é induzido pela aplicação de etanol nas folhas de cana-de-açúcar logo após uma hora de exposição ao agente indutor.

Concluindo, os resultados obtidos neste trabalho contribuem para elucidar questões importantes do processo de regulação da expressão gênica induzida por etanol, além da validação de um processo útil para aplicação na agricultura. A partir dos dados gerados nas análises de macroarranjos de cDNA foi possível ainda, validar um novo método de quantificação do erro estatístico resultante das análises realizadas no programa SOM (Self Organizing Maps).

## *ABSTRACT*

The sugarcane is a very important crop in Brazil, the biggest producer of the World. Besides sugar, the sugarcane is also used to produce ethanol, a very important renewable source of energy in Brazil. Nowadays this crop is responsible for generation of thousands of job positions directly and indirectly linked to the sugar and ethanol industry and market. Since the sugarcane is getting an important role in the Brazil economy, Brazilians researchers are focusing a lot of work in this crop to breed new varieties with improved traits and better adaptation to stressing environmental, such as cold, poor and acid soils, and attack of pests and diseases. Other important field of study is the improvement of sucrose content to increase the efficiency of ethanol production.

The development of a genetic system triggered by an external stimulus and able to modify the plant metabolism may be very useful to improve productivity and quality, besides the cost reduction of sugarcane production. This technology might permit Brazil to increase the expansion of this crop without degrading the environmental. Therefore, the aim of this work was the identification of an efficient system of gene expression induced by ethanol.

The technique of cDNA array was used to check the expression of 3,575 ESTs isolated from sugarcane leaves previously treated with ethanol. Seventy ESTs showed expression profile altered by ethanol. Among these ESTs with altered expression we have identified sequences encoding proteins involved with transcription and translation and genes previously reported as responsive to abiotic and biotic stresses. Forty eight of the 70 ESTs with altered expression were up-regulated by ethanol. The SsNAC23, a gene induced by cold, dry and insect attack, was also responsive to ethanol in this study. The expression of SsNAC23 in sugarcane leaves was increased after one hour of exposition to ethanol.

The expression analysis of genes induced by ethanol in arrays filters allowed us select a candidate gene whose promoter region will be identified and studied. The gene chosen was the ERD, Early Responsive to Dehydration, whose expression profile was the absence of expression under control treatment and high induction after one hour of ethanol application.

Previous works have showed that the genetic system based in ethanol induction, also known as *alc* system, can be very useful in crops. This system was tested in several

plant species and the results were very promising. Therefore, one of the goals of this work was the production of genetic altered sugarcane plants with the *alc* system. For that, we have developed an expression cassette using the *alcA* promoter regulating the expression of the *β-glucuronidase*, used as a reporter gene, and the constitutive promoter *Ubi-1* from maize controlling the expression of the gene *alcR*, a transcriptor factor essential to activation of the *alc* system. The transformed sugarcanes are being analysed and selected to construct introgression.

The results obtained in this work will contribute to understanding of gene expression regulation using an external and chemical inductor as well as provide new insights to the development of new approaches of gene expression control in sugarcane. Besides that, the array data was also used to validate a new statistical method of error estimation by using the SOM software.

**1. Origem e características da cana-de-açúcar**

A cana-de-açúcar (*Saccharum* spp) é uma gramínea semiperene cujo centro de origem não é bem definido. Acredita-se que essa espécie vegetal teve origem no Sudeste Asiático há cerca de 6.000 anos a.C. (Daniels e Roach, 1987). Desde então, a cana-de-açúcar vem se disseminando pelo mundo, principalmente em países tropicais e subtropicais.

O gênero *Saccharum,* do qual a cana-de-açúcar faz parte, compreende seis espécies descritas, *S. officinarum*, *S. barbieri*, *S. sinense*, *S. spontaneum*, *S. robustum* e *S. edule* (Mukherjee, 1957).

O melhoramento da cana-de-açúcar teve início concomitantemente com seu cultivo, tornando-a, juntamente com o milho, uma das espécies mais manipuladas pelos processos de melhoramento genético (Matsuoka, 1999a). As variedades modernas são resultado de hibridações intra e interespecíficas entre as espécies *S. spontaneum* e *S. officinarum* (Stevenson, 1965). A espécie *S. spontaneum* é tida como responsável pelo vigor vegetativo e resistência a estresses bióticos e abióticos enquanto que *S. officinarum* contribui com alto teor de açúcar (Ming et al, 1998).

Devido a sua origem multiespecífica, a cana-de-açúcar apresenta um genoma bastante complexo com nível de ploidia variável (2n = 40 até 128), além de aneuploidia (Heinz, 1987). Deste modo, a cana-de-açúcar tornou-se uma planta heterozigótica, com diferentes alelos por lócus gênico (Ma et al., 2003). Essa característica torna-a menos suscetível aos efeitos de mutações deletérias e pode contribuir para o surgimento de novos genes (Ma et al., 2003). Entretanto, essa mesma característica dificulta o processo de melhoramento genético convencional. O processo de seleção por intercruzamento visando a obtenção de variedades de cana-de-açúcar mais adaptadas às condições climáticas, mais resistentes a doenças e pragas e conseqüentemente mais produtivas, é muito lento, podendo levar até 15 anos para ser concluído e para disponibilizar uma nova variedade comercial (Burnquist, 2000). O processo de melhoramento genético pode ser efetivamente reduzido por meio da engenharia genética, através de técnicas modernas de análises e manipulação genética de plantas. As novas técnicas de identificação e isolamento de genes de interesse podem ser utilizadas para gerar variedades mais produtivas e melhor adaptadas às condições ambientais adversas, além de acelerar o processo de melhoramento da cana-de-açúcar. Conseqüentemente, promove o aumento da produtividade e a redução das

perdas no processo de produção da cana-de-açúcar.

## 2. Importância econômica da cana-de-açúcar

Devido a sua extraordinária capacidade de estocar sacarose nos entrenós, a cana-de-açúcar transformou-se em uma importante fonte de energia para o homem e conseqüentemente, de potencial econômico, tornando-se um alvo prioritário para o melhoramento genético (Moore, 2005).

A partir da sacarose acumulada no colmo da cana-de-açúcar é possível produzir o açúcar e o álcool (etanol), produtos estes, amplamente utilizados pela população mundial e cuja demanda vem crescendo ano após ano. Assim, a cana-de-açúcar vem se destacando como uma das mais importantes espécies vegetais cultivadas pelo homem na atualidade, contribuindo com cerca de 65% de todo o açúcar produzido (Ingelbrecht et al., 1999).

A cana-de-açúcar é cultivada em várias regiões do mundo, dentre elas, podemos destacar como mais importantes os países do continente americano, africano e asiático, totalizando cerca de 121 países produtores (http://www.sugarcanecrops.com/) (Figura 1). No Brasil, o cultivo da cana-de-açúcar teve início na primeira metade do século XVI, ocupando o título de monocultura mais antiga sendo explorada no País (Carvalho, 1993) (Figura 2). A cultura se adaptou muito bem as condições edafoclimáticas ocupando diferentes regiões do território nacional, estabelecendo-se principalmente nas regiões Centro-Sul e Nordeste (Carvalho, 1993). É uma cultura de grande importância econômica, especialmente para o Brasil, já que dentre os países que cultivam a cana-de-açúcar, o Brasil ocupa posição de destaque, contribuindo com 34% de toda a produção mundial, seguido pela Índia e pela Austrália.

O Brasil é o maior produtor mundial de cana-de-açúcar com uma safra estimada em 455 milhões de toneladas em 2006, ocupando uma área de cerca de 6 milhões de hectares (GCEA/ IBGE, 2006). Tal montante representa menos de 10% da superfície agrícola do País (Macedo e Nogueira, 2005). Embora os estados de Pernambuco, Alagoas, Rio de Janeiro e Minas Gerais tenham uma contribuição bastante efetiva na produção nacional, com quase 30% do total produzido, somente o estado de São Paulo é responsável por cerca de 52% da safra brasileira (IBGE, 2006), ou seja, pouco mais da metade de toda cana-de-açúcar produzida no País. Há cerca de 50 mil produtores e 308 unidades de processamento industrial, produzindo 17,7 milhões de toneladas de açúcar e 13,7 milhões de $m^3$ de etanol por ano (Saciloto, 2003).


Figura 1. Distribuição da cana-de-açúcar no mundo.

A agroindústria canavieira é responsável por 600 mil empregos diretos e geração de divisas para o país através da exportação de açúcar e álcool, seus principais produtos (Raizer, 1998). Com o grande apelo atual dos combustíveis renováveis e da agroenergia, a expectativa é que a área plantada com cana-de-açúcar no país sofra uma acelerada expansão, principalmente no Estado de São Paulo, estado onde se concentra grande parcela da população brasileira e da frota de veículos e indústrias do País.

A qualidade dos vários produtos e subprodutos gerados com o processamento da cana-de-açúcar é dependente em grande parte da qualidade do insumo que chega até as usinas. Fatores como a variedade de cana-de-açúcar utilizada, condições de solo e clima em que a cana foi cultivada, estágio de maturação da cana no momento da colheita, pureza do mosto, entre outros, pode interferir na qualidade do insumo e conseqüentemente na qualidade do produto final. Deste modo, uma importante estratégia para se obter o aumento da produtividade e da qualidade é a utilização de variedades melhoradas geneticamente.

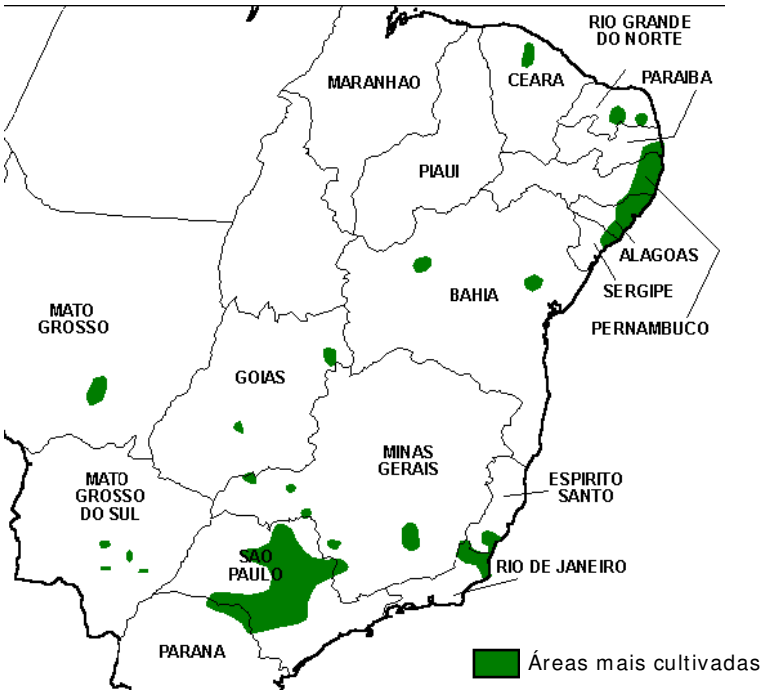



Figura 2. Distribuição da cana-de-açúcar no Brasil. Fonte GCEA/ IBGE, 2006

Atualmente, a demanda mundial por etanol combustível tem gerado novas oportunidades de negócio para o Brasil no mercado externo. Os impactos ambientais provocados pela emissão de gases precursores do efeito estufa na atmosfera tem levado vários países a iniciar programas e estratégias para diversificação da matriz energética, reduzindo a dependência por combustíveis fosséis não renováveis.

**3. A cana-de-açúcar e o Projeto Genoma**

Com o surgimento dos estudos de genômica, a descoberta de genes de interesse tornou-se um processo mais rápido e eficaz na geração de informações úteis ao melhoramento genético (Grivet e Arruda, 2001). Vários projetos cujos objetivos são estudar e entender o transcriptoma de diferentes espécies vegetais em diferentes condições experimentais, já foram e vem sendo conduzidos (Ewing et al. 1999; White et al., 2000; Dong et al., 2003; Ma et al., 2003). Esses projetos, denominados Projeto Genoma EST (*Expressed Sequence Tag*), passam a ser uma valiosa ferramenta para a descoberta de genes que são expressos em tecidos específicos, em fases distintas do desenvolvimento do organismo e em resposta a condições adversas de cultivo da planta. Desta maneira, os ESTs

identificados podem revelar informações sobre centenas ou mesmo milhares de genes de um organismo em uma condição temporal e espacial única, além de permitir a identificação de diferentes isoformas de transcritos (Andrews et al., 2000) e auxiliar o mapeamento genético (Schuler, 1997; Wu et al., 2002).

A importância sócio-econômica da cana-de-açúcar despertou grande interesse na obtenção de dados para um melhor entendimento dos processos fisiológicos e bioquímicos que ocorrem nessa cultura. Deste modo, a cana-de-açúcar passou a ser o modelo de estudo do Projeto SUCEST (*Sugarcane EST*), desenvolvido em rede por vários laboratórios do Estado de São Paulo com apoio financeiro da FAPESP. Esse projeto disponibilizou um conjunto muito grande de informações, identificando cerca de 43 mil prováveis transcritos (*SAS – Sugarcane Assembled Sequences*), obtidos de diferentes tecidos/órgãos vegetais. Destes, 29,7% não apresentam nenhuma homologia com outras seqüências gênicas ou protéicas previamente identificadas em qualquer outro organismo (Vettore et al., 2003).

Os transcritos identificados no Projeto SUCEST podem representar genes de grande importância econômica para cana-de-açúcar, em especial, aqueles relacionados ao metabolismo de sacarose, tolerância a condições adversas de clima e solo e à resistência a doenças e pragas (FAPESP, 2001), pois podem ser utilizados em programas de melhoramento genético biotecnológico, ou seja, pela criação de plantas geneticamente modificadas.

Devido ao crescente número de transcritos identificados através dos projetos ESTs, novas técnicas foram desenvolvidas com o objetivo de avaliar a expressão de milhares de genes simultanemante. Dentre essas técnicas destacam-se o macro e microarranjos de DNA (*macro and microarray)*, técnicas estas que tem sido amplamente utilizadas por grupos de pesquisa em vários locais do mundo. Basicamente, nesta técnica, seqüências conhecidas de DNA (cDNA, produto de PCR ou oligonucleotídeos) são fixadas em lâminas de vidro (chip microarray) ou membranas de náilon (filter macroarray) e hibridadas com sondas de cDNA da amostra que se deseja investigar. Através do pareamento complementar da sonda com o DNA fixado podem ser verificados os padrões de expressão de milhares de genes simultaneamente, comparando-se a intensidade do seu sinal nos diversos tratamentos.

Os macroarranjos são uma alternativa ao uso de microarranjos. Embora o número de genes avaliados simultaneamente no macroarranjo seja bem menor do que nos microarranjos, a vantagem do uso de macroarranjos está no fato de que essa tecnologia emprega metodologias rotineiras, robustas e de fácil implementação na maioria dos laboratórios de biologia molecular, sem a necessidade de investimentos vultosos em

equipamentos e reagentes (Felix et al, 2002).

O conjunto de dados e informações sobre os genes transcritos, gerados a partir do Projeto SUCEST, além de contribuir com as pesquisas que estavam em andamento possibilitou o surgimento de novos estudos e perspectivas com a cana-de-açúcar. Dentre eles, podemos destacar o SUCAST (*Sugarcane Signal Transduction*), que foi criado para identificar e caracterizar as sequências de transcritos, obtidas através do Projeto SUCEST, envolvidos em vias de transdução de sinal na cana-de-açúcar (Souza et al, 2001). O objetivo principal desse projeto foi identificar os ESTs envolvidos na regulação dos processos de crescimento e desenvolvimento da planta (fatores de transcrição, receptores, fosfatases, quinases, etc.). Deste modo, conhecendo melhor os genes envolvidos nas vias de sinalização é possível fazer inferências sobre diferentes aspectos do metabolismo da cana-de-açúcar. A descoberta do modo de ação de cada transcrito dentro das vias metabólicas revela potenciais alvos para o melhoramento dirigido das variedades de cana-de-açúcar.

A partir do banco de dados do Projeto SUCEST também foi possível identificar genes associados ao estresse por frio (Nogueira et al, 2003), estresse oxidativo (Kurama et al, 2002), ao metil jasmonato (De Rosa Jr et al., 2005), genes inibidores de protease associados a resistência a patógeno (Soares-Costa et al, 2002, Falco e Silva-Filho, 2003e Mello et al, 2003), tecido-especificidade envolvida na transdução de sinal (Papini-Terzi et al, 2005) e também avaliar tecido-especificidade de transposons (de Araujo et al, 2005). No momento, outros estudos, como este que visa identificar genes induzidos por etanol, vem sendo desenvolvidos por vários laboratórios de pesquisa nas universidades e centros de pesquisa do Brasil, utilizando a informação gerada, armazenada, organizada e disponibilizada pelo projeto SUCEST.

**4. Variedades de Cana-de-açúcar**

A cana-de-açúcar é uma das espécies domesticadas que vem sendo cultivada há mais tempo pelo homem (6000 a. C.), passando por processos de melhoramento desde o seu primeiro uso (Matsuoka, 1999a). No início, o processo de melhoramento era inconsciente com a seleção de plantas mais vigorosas para o plantio de novas lavouras pelo próprio agricultor. Atualmente, as variedades cultivadas são resultado de cruzamentos entre as espécies *S. spontaneum* e *S. officinarum* (Stevenson, 1965), variedades que foram escolhidas em função das suas características bastante favoráveis para o seu cultivo e produção. A espécie *S. spontaneum* contribui com o seu vigor vegetativo enquanto que *S. officinarum* é responsável pelo aumento no teor de sacarose no colmo (Ming et al, 1998).

Ao longo dos anos, muitos estudos foram realizados com as diversas variedades de cana-de-açúcar originadas nos programas de melhoramento. Nesses estudos, dados sobre produtividade agrícola, característica vegetativa e resistência a pragas e moléstias foram observados. Resultados desses estudos de caracterização botânica deixam claro a necessidade contínua de desenvolvimento de novas cultivares, que devem estar sempre se adaptando as novas realidades de cultivo e produção, tais como novas regiões produtoras, novos tipos de solo, novas técnicas de cultivo e manejo e maior acúmulo de sacarose (Raizer e Vencovsky, 1999).

A obtenção de máxima produtividade em cana-de-açúcar depende principalmente de dois fatores: 1) um correto planejamento de plantio; e 2) um adequado manejo das variedades. Desta forma, o potencial genético é fator determinante para garantir produtividade e qualidade de produção. Atualmente no Brasil, há centenas de variedades de cana-de-açúcar à disposição dos produtores. Cada uma com especificidade para diferentes manejos, épocas de safra e adaptabilidade local. Se a existência de muitas variedades é uma situação vantajosa, por outro lado torna a correta decisão da variedade uma tarefa bastante difícil de ser tomada pelo agricultor. Características agronômicas de três variedades comerciais de cana-de-açúcar, desenvolvidas pela Estação Experimental da COPERSUCAR (hoje denominada Centro de Tecnologia Canavieira, Piracicaba, SP) foram escolhidas para se realizar o presente estudo em função da sua importância e potencial para o setor álcool-açucareiro do Estado de São Paulo. Detalhes fitotécnicos destas variedades estão descritos a seguir:

### 4.1 Variedade SP89-1115

Essa variedade apresenta características como alta produtividade e teor de sacarose, ótima brotação de soqueira além de ser uma planta precoce. A época da colheita deve ser realizada até o meio da safra. Apresenta hábito semi-ereto e com baixa quantidade de fibra. Apresenta um florescimento freqüente, uma característica que não é muito desejada pelos produtores, uma vez que ocorre translocação de sacarose para a inflorescência e diminui o acúmulo no colmo. É bastante resistente ao carvão, mosaico, ferrugem e escaldadura.

### 4.2 Variedade SP90-3414

Essa variedade é resultado de cruzamentos entre as variedades SP80-1079 e SP82-3544. Destaca-se pelo seu porte ereto, por não apresentar florescimento e por sua alta

produção. É uma variedade indicada para colheita do meio para o final da safra. Apresenta teor médio de sacarose e fibras. Com relação às doenças e pragas, é susceptível à escaldadura e possui resistência intermediária ao carvão e broca.

### 4.3 Variedade SP80-3280

Apresenta alto teor de sacarose e boa brotação em soqueira. Devido ao crescimento inicial vigoroso, o seu perfilhamento é intermediário e o fechamento das entrelinhas é bom. É uma variedade que apresenta florescimento. No entanto, apresenta pouca isoporização do colmo. Apresenta resistência média ao carvão, mosaico, ferrugem e broca, porém é tolerante a escaldadura.

## 5. Indutores de expressão gênica

A regulação da expressão dos genes pode ocorrer em diferentes momentos e mecanismos da maquinaria celular. A expressão gênica está relacionada com a taxa de transcrição do DNA, processamento do RNA ou através do processo de tradução do RNA mensageiro (mRNA). A transcrição do DNA depende principalmente da estrutura da cromatina e da ativação do mecanismo de transcrição, envolvendo fatores de transcrição e proteína de polimerização de RNA (enzima RNA polimerase). No processamento do transcrito primário ocorre o revestimento da extremidade 5' (cap), adição da cauda poli-A na extremidade 3' e a correta remoção dos introns. A estabilidade do mRNA maduro na citoplasma da célula está relacionado com a sua taxa de tradução. Qualquer interferência em algum desses mecanismos pode interferir na expressão de um gene.

A expressão gênica é influenciada por diferentes fatores, internos e externos, como hormônios, estágio de desenvolvimento do organismo, estresses bióticos e abióticos.

Estresses abióticos induzem a mudanças morfológicas, bioquímicas e fisiológicas nas plantas (Andjelkovic e Thompson, 2006). Durante a aquisição de tolerância ao estresse também ocorrem mudanças na expressão dos genes. Os produtos gênicos induzidos pelo estresse podem ser classificados de duas maneiras: 1) genes que diretamente protegem contra o estresse e 2) genes que regulam a expressão de outros genes (Bray, 1997; Hasegawa et al., 2000).

Um causador de estresse abiótico em plantas é o alumínio. Na sua forma de maior toxidez ($Al^{3+}$) é capaz de inibir a elongação das raízes após poucos minutos do início da exposição da planta ao Al (Ryan et al., 1993). Em plântulas de milho foi observado que o gene *GST27*.2, uma glutationa S-transferase, é induzido por estresse provocado pelo

alumínio (Cançado et al., 2005). A enzima codificada por *GST27*.2 possui importante papel na detoxificação de compostos tóxicos presentes nas células. A indução desse gene pode estar relacionada ao mecanismo de tolerância ao estresse por Al em milho.

Um dos mais importantes estresses ambientais que afetam o crescimento e a produtividade das plantas é o frio (Xin e Browse, 2000). Plantas variam consideravelmente na sua habilidade de tolerar condições de frio intenso. Enquanto plantas oriundas de regiões de clima temperado resistem a temperaturas inferiores a -5 $^{o}$C, plantas de clima tropical e subtropical apresentam pouca ou nenhuma capacidade de sobrevivência em temperaturas próximas a 0 $^{o}$C (Tomashow, 2001). Cana-de-açúcar, uma planta de clima tropical, é considerada sensível a baixas temperaturas (Tai e Lentini, 1998). Estudos utilizando a técnica de cDNA macroarray identificaram genes de cana-de-açúcar com perfil de expressão alterados em condições de baixa temperatura (Nogueira et al., 2003). Dentre eles, foi identificado um gene pertencente à família de proteínas com domínio NAC (Kikuchi et al., 2000), denominado *SsNAC23*. Estudos com esse fator de transcrição sugerem que em folhas de cana-de-açúcar há um sistema bioquímico que identifica e sinaliza condição de baixa temperatura (Nogueira et al., 2005).

A habilidade de regular a expressão de genes em plantas superiores, através de métodos exógenos, apresenta vantagens consideráveis, tanto para a ciência básica quanto para a ciência aplicada na agricultura (Gatz, 1997). Diferentes métodos químicos de regulação da expressão de transgenes em plantas têm sido relatados (Ayoma e Chua, 1997; Gatz, 1997). Esses métodos permitem que a expressão seja ativada ou desativada, de acordo com a adição ou remoção do químico específico, sendo este um indutor ou inibidor (Wang et al., 2003). Diferentes substâncias podem ser utilizadas como indutores de expressão de transgenes: dexametasona (Ayoma e Chua, 1997), tetraciclina (Gatz e Quail, 1988), ethephon (Van Loon e Antoniw, 1982) e etanol (Caddick et al., 1998) dentre outros.

O etanol é um indutor químico que não apresenta toxicidade nos níveis requeridos para a indução. Além disso, é de fácil utilização, podendo ser aplicado nas plantas por pulverização, submersão das raízes e adicionado ao meio de cultura (Salter et al,1998).

## Objetivo Geral

O objetivo maior deste trabalho foi a identificação de genes expressos em folhas de cana-de-açúcar induzidos por etanol, visando a obtenção de um sistema de indução química da expressão gênica por meio da aplicação de etanol.

## Objetivos Específicos

I. Identificação de genes induzidos por etanol em folhas de cana-de-açúcar;
II. Construção de um cassete de expressão com o sistema *alc* de regulação gênica ativado por etanol e sua utilização para transformação genética de cana-de-açúcar;
III. Avaliação do perfil de expressão do gene SsNAC23 em folhas de cana-de-açúcar tratadas com etanol;
IV. Propor uma técnica de fácil implementação e uso para estimar o erro estatístico promovido pelo agrupamento de dados de arranjos de cDNA, baseado na metodologia de SOM (Self Organizing Maps).

# Identification of Genes Modulated by the Application of Ethanol on Sugarcane Leaves

Identificação de Genes Modulados pela Aplicação de Etanol em Folhas de Cana-de-açúcar

Sandra R. de Camargo, Geraldo M. A. Cançado, Eugênio C. Ulian, Marcelo Menossi

*Plant Cell Reports,* DOI 10.1007/s00299-007-0430-8

Plant Cell Rep
DOI 10.1007/s00299-007-0430-8

GENETICS AND GENOMICS

# Identification of genes responsive to the application of ethanol on sugarcane leaves

**Sandra R. Camargo · Geraldo M. A. Cançado · Eugênio C. Ulian · Marcelo Menossi**

Received: 16 May 2007 / Revised: 22 July 2007 / Accepted: 30 July 2007


**Abstract** The control of gene expression in precise time and space is a desirable attribute of chemically inducible systems. Ethanol is a chemical inducer with favourable features, such as being inexpensive and easy to apply. The aim of this study was to identify ethanol-responsive genes in sugarcane. The cDNA macroarray technique was adopted to identify transcript changes in sugarcane leaves (*Saccharum* spp. cv SP80-3280) exposed to ethanol. The expression profiles of sugarcane genes were analysed using nylon filters containing 3,575 cDNA clones from the leaf roll library of the SUCEST project. Seventy expressed sequence tags (ESTs) presented altered expression patterns, including ESTs corresponding to genes related to transcriptional and translational processes, abiotic stress and others. Several genes of unknown function were also identified. Among the 48 ESTs up-regulated by ethanol, an abiotic stress-responsive protein and an unknown function gene presented rapid induction by ethanol. The macroarray data of selected ethanol-responsive EST were confirmed by RNA-blot hybridisation. The expression profile of the 48 up-regulated genes was compared in two other cultivars: SP89-1115 and SP90-3414. Surprisingly, no gene showed a similar expression profile in the three cultivars. This result suggests that sugarcane plants have a high diversity in their responses to ethanol.

**Keywords** Macroarray · Expression profiling · Chemical inducer · *Saccharum* spp.

**Abbreviations**

| | |
|---|---|
| SUCEST | Sugarcane EST genome project |
| EST | Expressed sequence tag |
| ALCR | Alcohol-regulated transcription factor |
| *alcA* | Alcohol dehydrogenase gene |

## Introduction

Environmental signals can alter gene expressions by altering the activity of transcription factors that interact with regulatory DNA sequences. The identification of these regulatory sequences could be used to control the expression of recombinant genes in plants (Haran et al. 2000).

One of the most important challenges in plant biotechnology is the control of expression through a simple, inexpensive and safe activating mechanism. The activation of transgenic traits only when their expression is needed would prevent the energy wasted by plants and might avoid undesirable effects (Haran et al. 2000).

Some chemical products are used for repression, inactivation or activation of gene expression in plants (Zuo and Chua 2000). Chemicals, such as the steroids dexamethasone and stradiol, benzothiadiazol, methoxyfenozide, herbicides, copper and ethanol, are used to regulate transgene expression (Padidan 2003). Chemical-inducible systems are a very

Communicated by J. Zou.

S. R. Camargo (✉) · M. Menossi
Centro de Biologia Molecular e Engenharia Genética,
Universidade Estadual de Campinas (UNICAMP),
CP 6010, 13083-875 Campinas, SP, Brazil
e-mail: sandrarcamargo@gmail.com

G. M. A. Cançado
Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG),
31170-000 Belo Horizonte, MG, Brazil

E. C. Ulian
Centro de Tecnologia Canavieira, CP 162,
13400 Piracicaba, SP, Brazil

interesting tool because they allow the spacial, temporal and quantitative control of target gene expressions (Padidan 2003; Wang et al. 2003). There is an ethanol-inducible system based on the *alc* regulon of *Aspergillus nidulans* (Felenbok et al. 1988). This system is constituted by an alcohol-regulated transcription factor (ALCR) and a promoter derived from the *alcA* gene. In the absence of ethanol, the ALCR protein is inactive but when ethanol is present, it interacts with the ALCR protein, activating it. The activated ALCR then binds to the *alcA* promoter inducing the transcription of the target gene (Felenbok 1991).

The use of ethanol as an inducer has several advantages: ethanol is a biodegradable molecule, inexpensive, easy to apply and safe, because the amounts required to start induction are considered not toxic (Wang et al. 2003). It is worth noting the opposition that genetically modified organisms have received in several countries. Even considering that most of the concerns of the people have not found scientific support, there are a growing number of initiatives to develop new technologies that address some of the issues raised by society, for instance, the use of marker-free selection systems (De Vetten et al. 2003) and the use of genes from the same species. In the latter case, the development of an ethanol-induced system based on sugarcane genes would be interesting for several applications in sugarcane biotechnology.

Techniques for expression profile analysis on a large scale, such as cDNA array, can be helpful in the discovery of novel genes involved in biotic and abiotic stress (Cushman and Bohnert 2000). There is a large collection of sugarcane expressed sequence tags (SUCEST; Vettore et al. 2001) that allows for the analysis of the gene expression profile following different treatments using cDNA macroarrays. This technique was already successfully used to identify genes differentially expressed in response to MeJA (De Rosa et al. 2005) and cold stress (Nogueira et al. 2003) in sugarcane.

The aim of this work was the identification of genes whose expression in sugarcane leaf was affected by ethanol application. The macroarray technique was used to analyse the expression profile of 3,575 EST clones in sugarcane plants during ethanol exposition. The identification of ethanol-responsive genes would permit the indication of those that are promising for cloning and the characterization of their promoter regions.

## Materials and methods

### Plant material and ethanol treatment

The sugarcane (*Saccharum* spp.) cultivars SP80-3280, SP89-1115 and SP90-3414 were used. Plantlets were obtained by vegetative propagation and acclimated for 40 days in a greenhouse at the Centro de Tecnologia Canavieira (CTC, Piracicaba, SP, Brazil). The plantlets were then transferred to a growth chamber at 26°C with a 16-h photoperiod and photon flux density of 70 $\mu Em^{-2} s^{-1}$. After 1 week, the plantlets were sprayed with a 2.5% (v/v) ethanol solution, whereas control plantlets were sprayed only with water. Ethanol-treated and control leaves were collected 0, 1, 4 and 12 h after pulverization and immediately frozen in liquid nitrogen and stored at −70°C until RNA extraction. Three independent experiments were carried out with the SP80-3280 cultivar, and one experiment was done with the SP89-1115 and SP90-3114 cultivars.

### RNA extraction and cDNA synthesis

Total RNA was isolated from the leaves of treated and untreated sugarcane plantlets using the Trizol Reagent, according to the manufacturer's instructions (Life Technologies, USA). The cDNA probes were synthesized from total RNA as described by Schummer et al. (1999). Reverse transcriptions were performed with Superscript II (Invitrogen, USA) using 3 μM of an oligo-dT18V (V = A, C or G) primer, with 3,000 Ci $mmol^{-1}$ [$^{33}$P]CTP and unlabelled ATP,GTP and TTP (1 mM each) for 20 min at 42°C. Unlabelled CTP was then added to a final concentration of 1 mM, and the reaction continued for another hour. cDNA probes were purified with ProbeQuant G-50 micro-columns according to the manufacturer's instructions (Amersham Biosciences, USA), and the radiolabelled probe intensity was quantified in a 1217 Rack Beta liquid scintillation counter to verify [$^{33}$P]CTP incorporation.

### Hybridization of cDNA macroarray filters

Nylon-filters containing 3,575 EST plasmid clones from a cDNA library of immature sugarcane leaves were acquired from the Brazilian Clone Collection Center (BCCCenter, UNESP, Jaboticabal, Brazil). Each EST was spotted at two different locations on each filter.

Variations in the amount of DNA of each EST were estimated by hybridising the filters with an oligonucleotide probe that recognized the sequence of the $Amp^r$ gene of the pSPORT1 vector (Invitrogen) used in the cDNA libraries (Vettore et al. 2001). This probe was synthesized with the primers 5′-GTGGTCCTGCAACTTTATCCGC-3′ and 5′-TAGACTGGATGGAGCGGATAA-3′ in the presence of [$^{33}$P]CTP. After hybridisation and washing as described before (Mondego et al. 2005), the filters were exposed to

imaging plates for 96 h and then scanned in a phosphorimager FLA3000-G (Fujifilm, Japan). The probe was removed from the filters by boiling in a 0.1% (w/v) SDS solution. The filters were then hybridized with the cDNA probes at 42°C for 18 h and then washed (Mondego et al. 2005). The filters were sealed, exposed to imaging plates for 96 h, and scanned as described above.

### Data analysis of the nylon filter arrays

Quantification of the signal intensity of each spot was done with the Array Vision Software (Imaging Research, Canada). The local background was subtracted automatically from the signal intensity of each spot. To eliminate the variations in signal intensity amongst replicate filters, each spot was normalized as a ratio of the average intensity of all the spots on each filter. The normalized data were analysed using the SAM software (Significance Analysis of Microarrays, Tusher et al. 2001) using the 1.5 minimum fold-induction or repression relative to the control and 1% or less *False Discovery* parameters. Three separate analyses were carried out with each data set from the three cultivars experiments.

The expression profile data from those genes selected by the SAM software were processed as follows. The expression data set from each gene for each ethanol treatment was divided by the value detected in the reference condition (control treatment), and converted into the logarithm (base 2). Positive and negative values refer to up- and down-regulated ESTs, respectively, with respect to the reference condition.

### RNA-blot analysis

The total RNA (10 µg) from control and ethanol-treated sugarcane leaves was eletrophoresed on a 1% (w/v) agarose gel containing formaldehyde and transferred to Hybond XL nylon filters (Amersham Biosciences) by capillarity, according to the manufacturer's manual. The RNA filters were hybridized with a 26S rRNA probe to confirm equal RNA loading except for the RNA blots with samples of sugarcane treated and untreated with ethanol, whose RNA loading was checked only with EtBr. The filters were stripped and re-hybridized with cDNA radio-labelled probes synthesized using the Ready To Go kit (AP Biotech, USA) and [$^{32}$P]CTP as described by Sambrook et al. (1989). After hybridisation, the filters were washed, sealed and exposed to imaging plates for 24 h. Detection was carried out using a FLA3000 phosphorimager (Fujifilm) and data analysis done using the ImageGauge software (Fujifilm).

## Results

### Identification of ethanol-responsive genes with nylon arrays

Plantlets from the sugarcane variety SP80-3280 were sprayed with a 2.5% (v/v) ethanol solution. After 0, 1, 4 and 12 h, the leaves were cut and used to produce cDNA probes, which were hybridized onto nylon filters containing 3,575 sugarcane ESTs. The average signal of the replicated spots was used for statistical analyses using the SAM software (for details see "Material and methods"). The selection of ethanol-responsive genes was done using the statistical analysis of the microarray software (Tusher et al. 2001), comparing the expression levels between each time after the ethanol spray and the control treatment ($T = 0$). In addition, to be considered as induced by ethanol, the ratio between the intensity of the genes in each ethanol treatment and their respective intensities in the control had to be at least 1.5. Using these two cut off parameters, 70 ESTs were selected. The ESTs whose expression was significantly altered by the ethanol treatment thus represented about 2% of the total ESTs originally spotted on the arrays filters (Table 1). The selected genes encoding known proteins were grouped into a class with similar biological functions (1) nucleic acid and protein metabolism; (2) cellular dynamics; (3) and stress response (Table 1). Thirty-one of the selected ESTs had unknown functions, of which 16 encoded proteins presenting no hit in the GenBank database.

Among the ESTs identified, 48 (68.6%) were up-regulated and 22 (31.4%) were down-regulated (Table 1). Histones, ribosomal proteins, early responsive to dehydration protein (ERD) and peroxidase are examples of genes identified in this work as inducible by ethanol. The genes repressed by ethanol included a ring finger protein, a receptor-like serine/threonine kinase and a RNA helicase. Most of the ESTs identified had their expression altered after 1 or 4 h of ethanol treatment. After 12 h, 31 genes showed lower expression than in the first hours (Fig. 1).

### Evaluation of gene expression profiles in other sugarcane cultivars

The selection of candidate genes to be used in an ethanol-induced system must take into account the possible influence of genotype on the gene expression profile. To get an insight into this topic, two sugarcane varieties, SP89-1115 and SP90-3414, were evaluated. Around 50% of the genes induced in the SP80-3280 variety were also induced in one of the other two varieties, and only 14 genes (29.1%) were induced in the three varieties (indicated in italics of Table 1, Fig. 2). Considering that only a single

**Table 1** Average expression ratios and sequence similarities of ethanol-responsive sugarcane ESTs

| | Accession[a] | Blast Hit (E-value)[b] | Description[c] | Ratios[d] | | |
|---|---|---|---|---|---|---|
| | | | | T1/T0 | T4/T0 | T12/T0 |
| | Genes up-regulated in SP80-3280 variety | | | | | |
| | Gene regulation | | | | | |
| 1 | CA293457 | AAM94159 (3E-04) | Putative RGA protein | 1.4 | 3.2 | 2.9 |
| 2 | CA129983 | C86382 (2E-56) | 60S acidic ribosomal protein | 1.2 | 2.1 | 1.8 |
| 3 | CA294535 | XP_475510 (6e-68) | 50S ribosomal protein L10 | 1.5 | 1.5 | 1.5 |
| *4* | *CA119317* | *AAF64312 (6e-19)* | *Plastid ribosomal protein L19 precursor* | *1.3* | *1.7* | *1.6* |
| 5 | CA128696 | NP_909185 (1e-57) | 60S ribosomal protein L26 | 1.8 | 1.8 | 1.7 |
| *6* | *CA119363* | *AAX96401 (3e-29)* | *60S ribosomal protein L38* | *1.7* | *1.1* | *1.2* |
| 7 | CA294757 | CAB88393 (3E-22) | Trypsin inhibitor | 3.8 | 2.1 | 0.8 |
| | Cellular dynamics | | | | | |
| *8* | *CA116806* | *P40267 (3e-16)* | *Histone H1* | *0.7* | *1.5* | *2.5* |
| 9 | CA118922 | CAA40362 (2e-20) | Histone H1 | 0.8 | 1.1 | 1.6 |
| 10 | CA124635 | P62786 (3e-27) | Histone H4 | 1.1 | 1.3 | 1.6 |
| *11* | *CA124643* | *P62786 (6e-36)* | *Histone H4* | *1.0* | *1.3* | *1.7* |
| 12 | CA120539 | P62786 (3e-27) | Histone H4 | 1.0 | 1.2 | 1.6 |
| 13 | CA128654 | P62786 (7e-39) | Histone H4 | 0.9 | 1.4 | 1.6 |
| 14 | CA129817 | P62786 (6e-36) | Histone H4 | 1.0 | 1.2 | 1.8 |
| 15 | CA123642 | CAA74307 (3e-65) | Calmodulin | 1.7 | 1.1 | 1.1 |
| 16 | CA120001 | AAR29963 (2e-16 ) | Cellulose synthase catalytic subunit | 1.6 | 2.2 | 2.0 |
| | Stress response | | | | | |
| 17 | CA295973 | CAH69372 (2e-84) | Class III peroxidase | 2.4 | 1.7 | 1.3 |
| 18 | CA296014 | BAB63915 (2e-65) | ERD 4 | 4.4 | 1.5 | 0.9 |
| *19* | *CA300472* | *AAX18700 (1e-30)* | *NPR1* | *6.5* | *2.7* | *2.6* |
| | Others | | | | | |
| 20 | CA119441 | NP_921405 (4e-30 ) | Putative CGI-203 homolog | 1.0 | 2.1 | 2.6 |
| 21 | CA190212 | BAD11349 (7E-43) | BRI1-KD interacting protein | 1.3 | 1.6 | 1.4 |
| 22 | CA117139 | NP_912465 (9e-17) | NAC | 1.1 | 1.3 | 1.7 |
| *23* | *CA120594* | *AAC39501 (1e-31)* | *Nodulin-like protein* | *0.8* | *1.3* | *2.3* |
| | Unknown and no hit proteins | | | | | |
| 24 | CA123859 | AAU44092 (2e-29) | Unknown protein | 1.4 | 3.6 | 3.1 |
| 25 | CA126100 | AAX96578 (3e-59) | Unknown protein | 1.0 | 2.1 | 2.4 |
| 26 | CA294259 | NP_566113 (4e-35 ) | Unknown protein | 1.4 | 1.9 | 1.7 |
| *27* | *CA122671* | *NP_566593 (5e-31)* | *Unknown protein* | *1.1* | *1.4* | *1.7* |
| 28 | CA130287 | XP_463043 (2e-14) | Unknown protein | 1.1 | 1.3 | 1.6 |
| 29 | CA300395 | XP_463043 (1E-10) | Unknown protein | 1.2 | 1.4 | 1.5 |
| *30* | *CA117495* | *XP_550628 (6e-11)* | *Unknown protein* | *1.5* | *1.1* | *2.0* |
| 31 | CA119831 | AAM34403 (4e-05) | Unknown protein | 1.2 | 1.6 | 1.4 |
| 32 | CA296215 | XP_470272 (2e-16) | Unknown protein | 1.2 | 1.9 | 1.4 |
| *33* | *CA118738* | *XP_479628 (4e-14 )* | *Unknown protein* | *1.3* | *1.7* | *1.3* |
| 34 | CA128320 | CAG06679 (0.13) | Unknown protein | 1.4 | 1.5 | 1.5 |
| *35* | *CA123802* | *XP_464722 (3e-10)* | *Unknown protein* | *2.3* | *1.7* | *1.3* |
| *36* | *CA118497* | *BAD46590 (3e-76)* | *Unknown protein* | *1.7* | *0.8* | *0.8* |
| 37 | CA128221 | no hit | | 1.4 | 1.5 | 1.0 |
| 38 | CA130306 | No hit | | 1.5 | 1.7 | 0.8 |
| 39 | CA129408 | No hit | | 3.0 | 1.4 | 1.0 |
| *40* | *CA128874* | *No hit* | | *2.5* | *9.5* | *3.0* |
| *41* | *CA119555* | *No hit* | | *0.8* | *1.5* | *2.5* |

**Table 1** continued

| | Accession[a] | Blast Hit (E-value)[b] | Description[c] | Ratios[d] | | |
|---|---|---|---|---|---|---|
| | | | | T1/T0 | T4/T0 | T12/T0 |
| *42* | *CA125651* | *No hit* | | *0.7* | *1.6* | *2.1* |
| 43 | CA121993 | No hit | | 1.2 | 1.2 | 1.9 |
| 44 | CA125789 | No hit | | 1.0 | 1.2 | 1.7 |
| 45 | CA121251 | No hit | | 1.1 | 1.9 | 1.4 |
| 46 | CA122935 | No hit | | 1.3 | 1.4 | 1.6 |
| 47 | CA120904 | No hit | | 0.7 | 1.2 | 1.6 |
| 48 | CA295623 | no hit | | 1.8 | 1.4 | 1.4 |
| | Genes down-regulated in SP80-3280 variety | | | | | |
| | Gene regulation | | | | | |
| 49 | CA124254 | AAG23693 (4e-62) | 40S ribosomal protein S24 | 1.3 | 1.3 | 1.4 |
| 50 | CA122198 | P25866 (6e-74) | Ubiquitin carrier protein | 1.0 | 0.8 | 1.2 |
| *51* | *CA117179* | *CAA09199 (5e-34)* | *RNA helicase* | *0.6* | *0.6* | *1.1* |
| 52 | CA125259 | AAU44196 (9e-27) | RING-H2 finger protein | 0.5 | 0.4 | 0.5 |
| 53 | CA121094 | CAA57636 (1e-19) | Small subunit ribosomal protein S28 | 1.1 | 1.4 | 1.4 |
| 54 | CA121946 | BAA22083 (5e-13) | RNA binding protein | 0.8 | 1.3 | 1.4 |
| 55 | CA301169 | ABE87449 (4e-12) | Reverse transcriptase | 1.3 | 1.0 | 1.2 |
| | Cellular dynamics | | | | | |
| *56* | *CA121162* | BAE20250 (4e-45) | *Histone H3* | *1.0* | *1.1* | *1.2* |
| 57 | CA130347 | AAB36498 (4e-42) | Histone H3 | 0.9 | 1.0 | 1.2 |
| 58 | CA121135 | XP_469838 (1e-73) | Putative ankyrin | 1.4 | 0.7 | 0.7 |
| | Stress response | | | | | |
| 59 | CA130162 | JS0731(1e-14) | Wound-induced basic protein | 1.2 | 1.2 | 1.4 |
| 60 | CA116742 | BAD29543 (6e-13) | DRE/CRT binding factor 1 | 0.2 | 0.2 | 0.3 |
| | Others | | | | | |
| 61 | CA120110 | XP_480472 (4e-40) | Auxin efflux carrier protein-like | 0.8 | 0.9 | 1.1 |
| 62 | CA296079 | BAA35120 (5e-85) | NADH dependent glutamate synthase | 1.2 | 1.2 | 0.6 |
| *63* | *CA127352* | *XP_480583 (4e-23)* | *Receptor-like serine/threonine kinase* | *0.5* | *1.2* | *0.8* |
| 64 | CA125401 | AAV25244 (2e-14) | Putative phosphate translocator | 0.8 | 1.0 | 1.3 |
| | Unknown and no hit proteins | | | | | |
| 65 | CA122078 | AAO72578 (2e-32) | Unknown protein | 1.0 | 1.0 | 1.2 |
| 66 | CA189734 | BAD68838 (1e-07) | Unknown protein | 0.1 | 0.2 | 0.1 |
| 67 | CA121449 | No hit | | 0.8 | 0.8 | 1.1 |
| 68 | CA300807 | No hit | | 0.6 | 0.8 | 1.0 |
| 69 | CA118219 | No hit | | 0.5 | 0.5 | 0.5 |
| *70* | *CA123017* | *No hit* | | *0.4* | *0.4* | *0.4* |

Accession numbers in italics represent the ESTs up-regulated and down-regulated in all sugarcane cultivars

[a] GenBank accession number

[b] The best hit from BlastX analysis

[c] Description indicates the putative functions of the gene products expected from similarity

[d] Expression ratios of Ethanol-treated intervals (1, 4 and 12 h) in relation to its control (0 h): T1/T0 = 1 h after treatment/control; T4/T0 = 4 h after treatment/control; T12/T0 = 12 h after treatment/control

hybridisation was done at each time point after the ethanol spray in the case of SP89-1115 and SP90-3414, we can speculate that the fraction of genes induced in the three genotypes could be overestimated, even considering the use of an algorithm specific for single hybridisation statistics (Drummond et al. 2005).

## RNA-blot analysis

Three ethanol-induced ESTs, CA128874 (no hit), CA296014 (ERD 4) and CA300472 (NPR 1), were used as probes in RNA-blot hybridisation to validate the macroarray data. These genes were selected due to their profile of

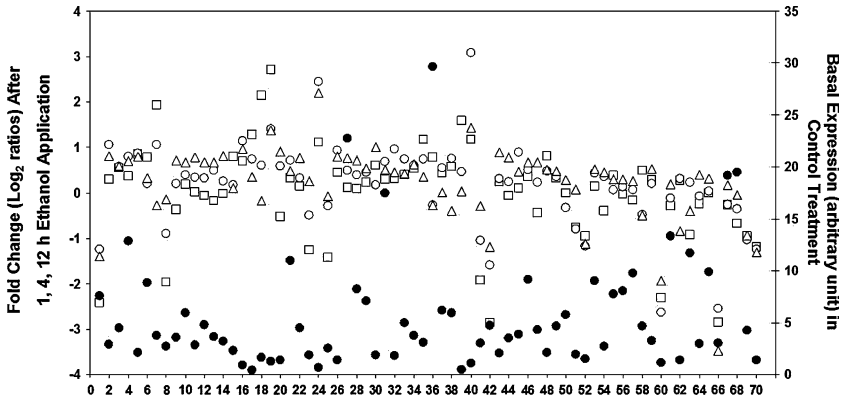



**Fig. 1** Expression profiles of 70 ethanol-responsive ESTs identified in the macroarray analysis of the SP80-3280 cultivar. The *numbers* on the *X*-axis refer to each EST described in Table 1. The *right Y*-axis shows the basal expression in the control treatment and the *left Y*-axis shows the fold change corresponding to the $\log_2$ (ratios) for each EST. The *black circles* show gene expressions in the control condition (basal expression axis); the *white squares*, *white circles*, and *white triangles* show gene expression changes after 1, 4, and 12 h of ethanol treatment, respectively (fold change axis). The fold changes are expressed as $\log_2$ of the ratio of the expression value in the treatment condition to the value in the control experiment

induction after treatment with ethanol. Figure 3 compares the expression profiling of these genes using the macroarray approach and RNA-blot. The expression profiles observed using both techniques were very similar for all the genes tested, indicating consistency between the two data sets. Figure 4 compares the expression profiling of the ESTs CA128874 (no hit), CA296014 (ERD 4), and CA300472 (NPR 1) in samples of RNA extracted from sugarcane leafs untreated and treated with ethanol.

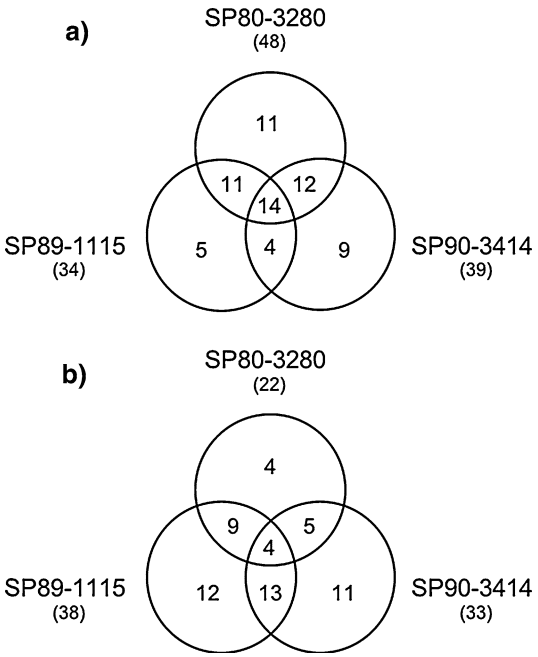



**Fig. 2** A comparison of 70 differential expression genes obtained from the SP80-3280 sugarcane cultivar with two other cultivars: SP89-1115 and SP90-3414. **a** Venn diagram of 66 up-regulated genes in three sugarcane cultivars (SP80-3280, SP89-1115, and SP90-3414) exposed to ethanol. **b** Venn diagram of 58 down-regulated genes in three sugarcane cultivars. *Values* indicate the number of genes

## Discussion

### Effect of ethanol in the expression profiling of sugarcane leaves

High-density filters were used to identify alterations in the gene expression profile in the ESTs of sugarcane plantlets after exposure to ethanol. Seventy genes were identified with their expressions significantly altered by ethanol, in the SP80-3280 sugarcane cultivar (Table 1). These genes are involved in different biological process, such as nucleic and protein metabolism, dynamic cell processes and stress conditions, indicating that several metabolic processes were altered during the ethanol treatment of sugarcane plants.

Several genes related to nucleic acid and protein metabolism were up- or down-regulated by ethanol stress in the SP80-3280 sugarcane variety. Six genes encoded different ribosomal proteins, of which five were up-regulated in plants treated with ethanol, including L10 (accession CA294535 and CA129983), L19 (CA119317), L26 (CA128696) and L38 (CA119363), and one, the ribosomal protein S24 (CA124254), was down-regulated. Ribosomal proteins are involved in the translation of mRNA into a polypeptide chain during protein synthesis. The fine regulation of this process is thought to play an important regulatory role in the translational control of gene expression.

Another gene involved in gene regulation and up-regulated by ethanol encoded a trypsin inhibitor (CA294757).

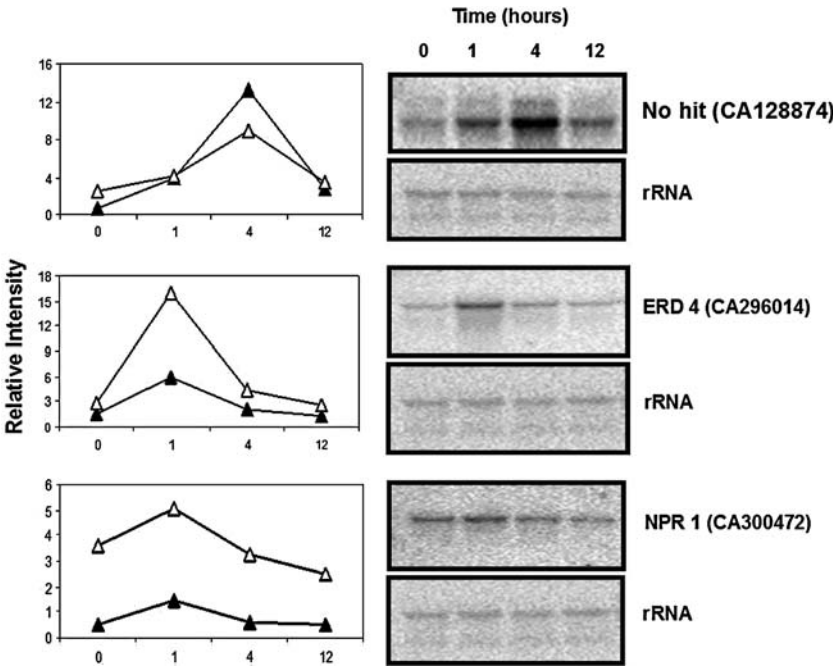



**Fig. 3** A comparison between the EST macroarray and RNA-blot analyses for ethanol-regulated sugarcane ESTs. The graphs show the expression profile observed in macroarray (*black triangle*) and RNA-blots (*white triangle*). In RNA gel blots, each lane was loaded with total RNA isolated from sugarcane leafs not-treated with ethanol (control) and sugarcane leafs treated with ethanol after 1, 4, and 12 h

This gene showed high induction by ethanol after the first hour of plant treatment (Table 1). As the time of plant exposition to ethanol increased, the induction of the trypsin inhibitor was decreased. The gene encoding a putative RNA helicase protein (CA117179) was down-regulated by the ethanol treatment. This enzyme shows dsRNA unwinding activity and is associated with several RNA metabolism processes, such as transcription, splicing, ribosome biogenesis, RNA transport to the cytoplasm, translation, RNA degradation and organelle gene expression (de la Cruz et al. 1999).

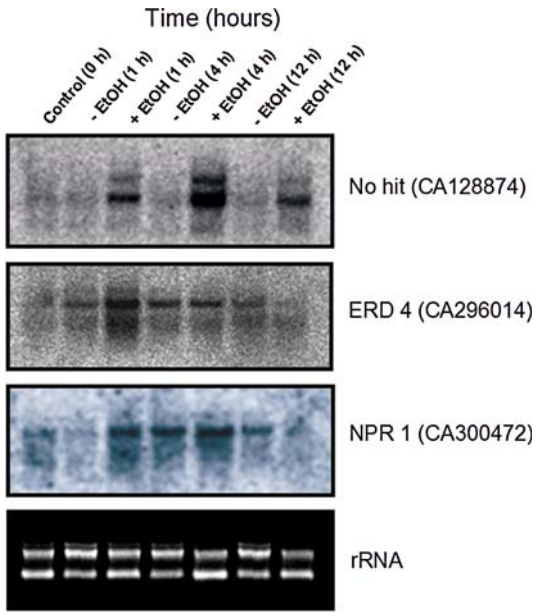



**Fig. 4** RNA-blot analyses of CA128874 (*no hit*), CA296014 (*ERD 4*), and CA300472 (*NPR 1*). In the *RNA* gel blots, each lane was loaded with total RNA isolated from sugarcane leafs not-treated with ethanol (control); sugarcane leafs not-treated with ethanol after 1, 4, and 12 h of water application; and sugarcane leafs treated with ethanol after 1, 4, and 12 h of 2.5% ethanol solution application

Genes involved in cell dynamics also presented changes in their expression profiles due to ethanol. Genes that encoded histone proteins were represented in this group, although these proteins were also related to gene regulation. In this study, the histones H1 (CA116806 and CA118922), H3 (CA130347 and CA121162) and H4 (CA124635, CA120539, CA128654, CA129817 and CA124643) were identified as ethanol-responsive. H1 and H4 were induced by ethanol during the three times of plant exposition, with greater induction after 12 h of treatment, indicating that the chromatin structure may be modified in response to ethanol.

In this study, we also identified genes responsive to stress conditions, including a wound-induced protein (CA130162), a peroxidase class III (CA295973), NPR1 (CA300472) (nonexpresser of pathogenesis-related gene) and ERD 4 (CA296014) (early responsive to dehydration). Peroxidases are present at all stages of plant life, due to several stressful conditions that plants meet. Under chemical stress, peroxidases are often induced at the beginning of an event, and then slowly decrease with time (Klumpp et al. 2000). This type of response was found in the present study, in which peroxidase had an elevated expression in the first hour of ethanol exposition and subsequently decreased. The NPR1 gene is a key component of the salicylic-acid-regulated PR (pathogenesis-related) gene expression and disease resistance (Cao et al. 1997; Shah et al. 1997). A sequence analysis of NPR1 revealed four ankyrin repeats, which are found in proteins with diverse

biological functions, and represent domains interacting with other proteins (Bork 1993; Michaely and Bennet 1992). This protein interacts with bZIP transcriptional factors regulating the expression of genes associated with salicylic acid induction (Zhang et al. 1999). NPR1 was induced during the whole (12 h) ethanol treatment, with a peak of expression of up to six times in the first hour. After 4 and 12 h of ethanol exposition the expression was about 2.5-fold. The ERD 4 gene also showed an induction peak after 1 h of ethanol exposition, presenting 4.4 times more expression than found in control plants. This gene has been shown to be related to dehydration stress. The ERD gene produced metabolic proteins to protect cells and also showed a function in regulating genes that were involved in transducing the stress response signals (Thomashow 1999; Ingran and Bartels 1996; Steponkus et al. 1998; Bohnert et al. 1999). The higher expression of genes encoding an ERD suggested that ethanol caused dehydration, and probably its expression would protect plants against this stress.

A gene encoding a calmodulin protein was up-regulated after 1 h of ethanol exposition (CA123642). Calmodulin (CaM) is a calcium ($Ca^{2+}$) signal transducer involved in controlling many biochemical processes in both plants and animals (Snedden and Fromm 2001; Yang and Pooviah 2003). Biotic and abiotic stimuli alter intracellular $Ca^{2+}$ fluxes, triggering a signalling cascade (Sanders et al. 2002; Berrige et al. 2003). CaM serves as a sensor to detect and translate $Ca^{2+}$ signals regulating the activities of target proteins, which allow organisms to adapt to the changing environment (Baum et al. 1996). Calmodulin could act perceiving the ethanol stimulus and thus modulate the expression of other sugarcane genes.

The EST encoding a RING-H2 finger protein (CA125259) was repressed after ethanol exposure. The ring finger domain is found in numerous proteins that participate in the regulation of many growth and development processes in eukaryotic organisms (Salinas-Mondragón et al. 1999). At the molecular level it may be involved in transcriptional regulation, vesicular transport, peroxisomal biogenesis and in protein degradation (Borden 2000 and Freemont 2000). RING-H2 domains have been described as part of proteins that are involved in a ubiquitination pathway, associated with E3 ubiquitin ligases (Jackson et al. 2000). DRE/CRT binding factor 1 (CA116742) is a transcription factor induced by dehydration and low-temperatures and controls the expression of several genes related to stress tolerance (Kazuo and Kazuko 2000). However, this EST was down-regulated by ethanol stress.

NADH-glutamate synthase is a glutamate synthase form that uses NADH as the electron donor for ammonium ion incorporation into glutamate molecule by conversion of L-glutamine plus 2-oxoglutarate to L-glutamate. Glutamate is used for the synthesis of amino acids, nucleic acids and in protein biosynthesis (Suzuki and Knaff 2005). Since the application of ethanol repressed the expression of NADH-glutamate synthase, we may suggest that this compound might affect DNA replication, transcription and translation.

The expression profile of the 70 up- and down-regulated genes was compared in two other sugarcane cultivars, SP89-1115 and SP90-3414, using a single cDNA array hybridisation. Figure 2 illustrates the comparison of the gene expression profile between the three sugarcane cultivars. Of the 48 up-regulated genes in the SP80-3280 cultivar, only 14 were up-regulated in the other two cultivars. A similar result occurred for the down-regulated genes; in SP80-3280, 22 genes were down-regulated. Amongst these, only 4 genes presented the same characteristics in the SP89-1115 and SP90-3414 cultivars. This result suggests that different sugarcane varieties show great diversity in their response to ethanol application, probably because of their different foliar architecture, cuticle thickness, etc. In this sense, our data highlight an unexpected difficulty in the construction of an ethanol-based system using sugarcane genes.

The data presented in this work can help in the identification of ethanol-induced systems similar to the *alc* system of *A. nidulans* (Salter et al. 1998). Some genes, such as CA300472 (similar to NPR1 gene) and CA296014 (ERD 4), present interesting characteristics for this purpose. Both have low levels of basal expression, rapid induction after ethanol exposition and elevated expression levels. The one encoding an ERD 4 protein was particularly interesting, because it was induced in both SP80-3280 and SP89-1115. We are starting the identification of the promoter region of both promoters, and future work will evaluate their usefulness in an ethanol-induced system.

Vreugdenhil et al. (2006) verified in wild *Solanum tuberosum* that application of solution with low concentration of ethanol was responsible for down-regulation of the expression of at least some of the genes involved in cell division and in storing reserves in this species. In our studies we also showed that application of low concentration of ethanol was responsible by alteration in gene expression of non-transformed sugarcane plants but the genes that presented alteration of expression did not result in perceptible toxic symptoms or developmental alterations in the sugarcane plants tested. Although the results obtained in *S. tuberosum* indicate deleterious effects of ethanol in the tubers development, it is not possible to extrapolate the same effects to sugarcane, because both species are distantly related. Until now, there is no report about deleterious effects promoted by ethanol application on sugarcane leafs. New experiments will be carried out to check the physiological and biochemical effects of ethanol application in sugarcane leaf.

**Acknowledgments** The authors would like to thank J. M. Felix, E. R. Santos and R. D. Drummond for excellent technical assistance. This work was developed with financial support from FAEPEX/ UNICAMP (grant 519.292 to M.M.) and the Centro de Tecnologia Canavieira (CTC). S.R.C. was recipient of a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), G.M.A.C. was recipient of a fellowship from Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), and M.M. was recipient of a fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## References

Baum G, Lev-Yadun S, Fridmann Y, Arazi T, Katsnelson H, Zik M, Fromm H (1996) Calmodulin binding to glutamate decarboxylase is required for regulation of glutamate and GABA metabolism and normal development in plants. EMBO J 15:2988–2996

Berrige MJ, Bootman MD, Roderick HL (2003) Calcium signalling: dynamics, homeostasis and remodelling. Nat Rev Mol Cell Biol 4:517–529

Bohnert HJ, Su H, Shen B (1999) Molecular mechanisms of salinity tolerance. In: Shinozaki K, Yamaguchi-Shinozaki K (eds) Molecular responses to cold, drought, heat, and salt stress in higher plants. R.G. Landes, Austin, pp 29–62

Borden KLB (2000) RING domains: máster builders of molecular scaffolds? J Mol Biol 295:1103–1112

Bork P (1993) Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? Protein Struct Funct Genet 17:363–374

Cao H, Glazebrook J, Clarke JD, Volko S, Dong X (1997) The Arabidopsis *NPR1* gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. Cell 88:57–63

Cushman JC, Bohnert HJ (2000) Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 3:117–124

De la Cruz J, Kressler D, lLnder P (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. Trends Biochem Sci 24:192–198

De Rosa Junior VE, Nogueira FTS, Menossi M, Ulian EC, Arruda P (2005) Identification of methyl jasmonate-responsive genes in sugarcane using cDNA arrays. Braz J Plant Physiol 17:173–180

De Vetten N, Wolters AM, Raemakers K, Van Der Meer I, Ter Stege R, Heers E, Heers P, Visser R (2003) A transformation method for obtaining marker-free plants of a cross-pollinating and vegetatively propagated crop. Nat Biotechnol 21:439–442

Drummond RD, Pinheiro A, Rocha CS, Menossi M (2005) ISER: selection of differentially expressed genes from DNA array data by non-linear data transformations and local fitting. Bioinformatics 21:4427–4429

Felenbok B (1991) The ethanol utilization regulon of *Aspergillus nidulans*: the alcA-alcR system as a tool for the expression of recombinant proteins. J Biotechnol 17:11–17

Felenbok B, Sequeval D, Mathieu M, Sibley S, Gwynne DI, Davies RW (1988) The ethanol regulon in *Aspergillus nidulans*: characterization and sequence of the positive regulatory gene alcR. Gene 73:385–396

Freemont PS (2000) RING for destruction? Curr Biol 10:R84–R87

Haran S, Logendra S, Seskar M, Bratanova M, Raskin I (2000) Characterization of Arabidopsis acid phosphatase promoter and regulation of acid phosphatase expression. Plant Physiol 124:615–626

Ingran J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol Plant Mol Biol 47:377–403

Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, Kaiser BK, Reimann JDR (2000) The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. Trends Cell Biol 10:429–439

Kazuo S, Kazuko YS (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217–223

Klumpp G, Furlan CM, Domingos M, Klumpp A (2000) Response of stress indicators and growth parameters of *Tibouchina pulchra* Cogn. exposed to air and soil pollution near the industrial complex of Cubatao, Brazil. Sci Total Environ 246:79–91

Michaely P, Bennet V (1992) The ANK repeat: a ubiquitous motif involved in macromolecular recognition. Trends Cell Biol 2:127–129

Mondego JMC, Guerreiro-Filho O, Bengtson MH, Drumond RD, Felix JM, Duarte MP, Ramiro D, Maluf MP, Sogayar MC, Menossi M (2005) Isolation and characterization of Coffea genes induced during coffe leaf miner (Leucoptera coffeella) infestation. Plant Sci 169:351–360

Nogueira FTS, de Rosa Junior VE, Menossi M, Ulian EC, Arruda P (2003) RNA expression profile and data mining of sugarcane response to low temperature. Plant Physiol 132:1811–1824

Padidan M (2003) Chemically regulated gene expression in plants. Curr Opin Plant Biol 6:169–177

Salinas-Mondragon RE, Garciduenas-Pina C, Guzman P (1999) Early elicitor induction in members of a novel multigene family coding for highly related RING-H2 proteins in *Arabidopsis thaliana*. Plant Mol Biol 40:579–590

Salter MG, Paine JA, Riddell KV, Jepson I, Greenland AJ, Caddick MX, Tomsett AB (1998) Characterisation of the ethanol-inducible *alc* gene expression system for transgenic plants. Plant J 16:127–132

Sambrook J, Fritsch EF, Maniats TA (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sanders D, Pelloux J, Brownie C, Harper J (2002) Calcium at the crossroads of signaling. Plant Cell 14:S401–S417

Schummer M, Ng VLV, Baumgarner RE, Nelson PS, Schummer B, Bednarski DW, Hassell L, Baldwin RL, Karlan BY, Hood L (1999) Comparative hybridization of an array of 21,500 ovarian cDNAS for the discovery of genes overexpressed in ovarian carcinomas. Gene 238:375–385

Shah J, Tsui F, Klessig DF (1997) Characterization of a salicylic acid insensitive mutant (*sai1*) of *Arabidopsis thaliana*, identified in a selective screen utilizing the SA-inducible expression of the *tms2* gene. Mol Plant Microbe Interact 10:69–78

Snedden WA, Fromm H (2001) Calmodulin as a versatile calcium signal transducer in plants. New Phytol 151:35–66

Steponkus PL, Uemura M, Joseph RA, Gilmour SJ, Thomashow MF (1998) Mode of action of the COR15a gene on the freezing tolerance of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 95:14570–14575

Suzuki A, Knaff DB (2005) Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. Photosynth Res 83:191–217

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 50:571–599

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121

Vettore A, da Silva FR, Kemper EL, Arruda P (2001) The libraries that made SUCEST. Genet Mol Biol 24:1–7

Vreugdenhil D, Claassens MMJ, Verhees J, Van der Krol AR, Van der Plas L (2006) Ethanol-inducible gene expression: non-

transformed plants also respond to ethanol. Trends Plant Sci 11:9–11

Wang R, Zhou X, Wang X (2003) Chemically regulated expression systems and their applications in transgenic plants. Transgenic Res 12:529–540

Yang T, Pooviah B (2003) Calcium/calmodulin-mediated signal network in plants. Trends Plant Sci 8:505–512

Zhang Y, Fan W, Kinkema M, Li X, Dong X (1999) Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the *PR-1* gene. Proc Natl Acad Sci USA 96:6523–6528

Zuo J, Chua NH (2000) Chemical-inducible systems for regulated expression of plant genes. Curr Opin Plant Biol 11:146–151

# SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane

SsNAC23, um membro da família de proteínas com domínio NAC, está associada a frio, herbivoria e estresse hídrico em cana-de-açúcar

Fábio T.S. Nogueira, Paulo S. Schlögl, Sandra R. Camargo, Jorge H. Fernandez, Vicente E. De Rosa, Patrícia Pompermayer e Paulo Arruda

*Plant Sciences, 169: 93-106 (2005)*

Available online at www.sciencedirect.com
SCIENCE DIRECT
Plant Science 169 (2005) 93–106
PLANT SCIENCE
www.elsevier.com/locate/plantsci

# SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane

Fábio T.S. Nogueira [a,b], Paulo S. Schlögl [a,b], Sandra R. Camargo [a,b], Jorge H. Fernandez [c], Vicente E. De Rosa Jr. [a,b], Patrícia Pompermayer [d], Paulo Arruda [a,b,*]

[a] *Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas (UNICAMP) 6010, 13083-970 Campinas, SP, Brazil*
[b] *Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), 13083-970 Campinas, SP, Brazil*
[c] *Embrapa, Universidade Estadual de Campinas (UNICAMP), 13083-970 Campinas, SP, Brazil*
[d] *Departamento de Genética, Escola Superior Luiz de Queiroz (ESALQ-USP), 13400-970 Piracicaba, SP, Brazil*

Received 29 September 2004; received in revised form 15 February 2005; accepted 1 March 2005
Available online 25 March 2005

**Abstract**

The NAC (*N*AM/*A*TAF1/2/*C*UC) domain protein family is widely distributed in plants, and some of their members are regulated by biotic and abiotic stress. Using the SUCEST sugarcane EST database, we have identified 26 members of the NAC domain protein family, referred to collectively as *Saccharum* sp. NACs (SsNACs). Several *SsNACs* genes displayed a similar pattern of tissue/organ expression. One of these, *SsNAC23*, was found to be induced by chilling stress. SsNAC23 showed strong similarity to the *Arabidopsis* ATAF1 and rice OsNAC6 proteins, which have been phylogenetically classified into dicot and monocot ATAF sub-subfamilies, respectively. Molecular modeling of the SsNAC23 3-D structure and its nuclear localization suggest that this protein has conserved transcription factor domains. The expression of *SsNAC23* was strongly induced at 4 °C but not at 12 °C, indicating that this gene is involved in the response to extreme low temperature stress. *SsNAC23* was also induced by water stress and herbivory. The role of *SsNAC23* in sugarcane plants subjected to stress stimuli is discussed.


*Keywords:* NAC domain protein; *Saccharum* sp.; Abiotic stress; Biotic stress; Transcription factor

## 1. Introduction

Plants vary considerably in their ability to survive under stress conditions. Stress-regulated genes play an important role in stress tolerance mechanisms when plants are exposed to hostile environments [1]. Chilling, for instance, is a common abiotic stress that unfavorably affects plant growth, especially those of tropical and subtropical origin, such as sugarcane [2]. Molecular studies using cDNA arrays have identified several genes that are upregulated by biotic and abiotic stress [3–8]. Many of these genes encode proteins involved in signaling pathways, including mitogen-activated protein kinases (MAPKs), histidine kinases, $Ca^{2+}$-dependent protein kinases (CDPKs), SOS3 $Ca^{2+}$ sensor family, as well as transcription factors. The sequencing of entire plant genomes has revealed that transcription factors often belong to large gene families, such as MYB (MYB DNA-binding domain), ERF (EREBP/ethylene response factor), bZIP, and WRKY [9,10]. Some members of these families play a role in regulating defense responses to biotic and environmental stress while others appear to coordinate plant developmental pathways [9,11].

An interesting family of transcription factors, unique to plants, is the NAC domain protein family [12,13]. The name derives from a highly conserved N-terminal region originally associated with the *NO APICAL MERISTEM* (*NAM*) gene in *Petunia* and the *ATAF1*, *ATAF2* and *CUP-*

* Corresponding author at: Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas (UNICAMP) 6010, 13083-970 Campinas, SP, Brazil. Tel.: +55 19 3788 1137; fax: +55 19 3788 1089.
*E-mail address:* parruda@unicamp.br (P. Arruda).

0168-9452/$ – see front matter 
doi:10.1016/j.plantsci.2005.03.008

*SHAPED COTYLEDON* (*CUC*) of *Arabidopsis thaliana* [12]. In contrast to the conserved N-terminal region, the carboxy terminus of NAC domain proteins is highly variable in length and amino acid sequence [14]. Additionally, the NAC domain protein family shows no sequence homology to any other proteins in the GenBank nr (non-redundant) database [15]. Recently, it was shown that NAC domain proteins in rice (*Oryza sativa*) and *A. thaliana* constitute a large superfamily whose members have different roles in plant development [16]. For instance, the *Arabidopsis NAP* (*NAC-like activated by AP3/PI*) was found to be upregulated in flower organ primordia by the two MADs box proteins, APETALA3 (AP3) and PISTILLATA (PI) [17]. *NAP* was expressed mainly underneath the inflorescence meristem and in the pedicels, as well as in developing sepals of young flowers. Other examples include the *Arabidopsis NAC1* and *CUC* genes. *NAC1* mediates auxin signaling to promote lateral root development [15]. *CUC1* is required for complete penetrance of the cupped cotyledon phenotype [18] whereas *CUC3* seems to be involved in establishing the boundary regions containing cells with low proliferation and/or differentiation rates [19].

NAC domain genes have also been implicated in biotic and abiotic stress responses. The potato *StNAC* and the *Arabidopsis ATAF1* and *ATAF2* were found to be induced by pathogen attack and wounding [20]. More recently, it was found that *Brassica napus BnNAC* family members were differentially regulated in response to biotic and abiotic stress [21]. Finally, it was demonstrated that the *Arabidopsis* TIP (TCV-INTERACTING PROTEIN) interacts specifically with the capsid protein of turnip crinkle virus (TCV), suggesting that TIP may participate in virus resistance mechanism [22]. However, the correlation between family classification and the putative physiological roles of NAC domain proteins in stress responses has not been elucidated.

Two lines of evidence suggest that NAC domain proteins are involved in transcriptional regulation. First, *A. thaliana* lines over-expressing *NAC1* showed upregulation of the auxin-responsive gene *TIR1* (*TRANSPORT INHIBITOR RESISTANT1*) with a corresponding downregulation being observed in *NAC1* anti-sense lines [15]. Second, *BnNACs* were capable of activating gene expression in yeast and recognize a *cis*-acting element within the CaMV 35S promoter sequence that is also found in some plant promoter sequences [21].

Here, we report the identification of 26 non-redundant sugarcane genes encoding NAC domain proteins, including a chilling-inducible sugarcane *NAC*, *SsNAC23* [7]. RNA gel-blot analysis showed that *SsNAC23* was strongly induced by extreme chilling stress, water stress and herbivory. Prediction of the three-dimensional structure of the NAC domain and in vivo analysis showed that SsNAC23 has the conserved characteristics of NAC domain transcription factors. The role of *SsNAC23* in response to stress stimuli is discussed.

## 2. Methods

### 2.1. Plant growth and treatments

Sugarcane plantlets (*Saccharum* sp. cv. SP80-3280), propagated axenically in vitro [7], were maintained at 26 °C on a 16 h/8 h day/night cycle with a photon flux density of 70 μE $m^{-2}$ $s^{-1}$. To compare the *SsNAC23* expression profiles at 4 and 12 °C, 3-month-old plantlets were transferred to 4 or 12 °C under the same photoperiod conditions. Control plantlets were maintained at 26 °C. The leaves of control and chilling-treated plantlets were harvested after 0, 1, 3, 6, 12, and 48 h of treatment (six plantlets per time point). Plants were also treated with abscisic acid (ABA) or polyethylene glycol (PEG) by transferring sugarcane plantlets to rooting medium [7] supplemented with 100 μM ABA or 20% (p/v) PEG 6000, a polymeric osmolyte that does not penetrate plan cells and is relatively non-toxic [23]. For the ABA treatment, plants were sampled at 0, 1, 3, 6, 12, 24 and 48 h while for the treatment with PEG, the leaf samples were sampled after 0, 3, 6, 12, 24, and 48 h. The expression pattern of *SsNAC23* in apical meristem tissues, leaves, roots, and stem was examined using one-month-old sugarcane plantlets cultivated in a greenhouse. Flowers were obtained from sugarcane plants grown at the Copersucar experimental station (http://www.ctc.com.br). For the bioassay with sugarcane borer, third and fourth instar *Diatraea saccharalis* larvae were starved overnight prior to the experiments. Larvae were placed on the mature terminal leaves of three-month-old sugarcane plants and then removed after 0, 4, 12 and 24 h. Tissue samples were collected from sites close to where larvae had fed and were immediately frozen in liquid nitrogen.

### 2.2. Identification and sequence analysis of sugarcane NAC domain proteins

*A. thaliana* and *O. sativa* NAC domain proteins [16] were used as drivers to identify and classify a set of non-redundant sugarcane assembled sequences (SASs) [24] sharing similarities to NACs in the SUCEST database (http://sucest.lad.ic.unicamp.br/public). SASs with an e-value $\leq 10^{-30}$ were considered to represent putative sugarcane NAC domain-containing genes (referred to hereafter as *SsNACs*; *Saccharum* sp. *NACs*). Subsequently, the identified sequences were inspected for the presence of a conserved N-terminal NAC domain sequence using Pfam database [25] and BLASTP algorithm. To identify gene redundancy among sequences retrieved from the SUCEST database, the DNA sequences of all *SsNAC* candidates were aligned using the CLUSTALX program [26]. Those SASs whose DNA sequences showed more than 95% identity were considered redundant.

An un-rooted, neighbor-joined phylogenetic tree (1000 bootstraps) was inferred by using the MEGA2 program [27]

and the protein sequences corresponding to the NAC domain amino acids to ensure the most conserved alignment. Motifs present in the C-terminal region of *Arabidopsis,* rice, *Brassica napus*, and orthologous sugarcane NAC domain proteins belonging to sub-subfamilies I and II were detected using the pattern prediction program package MEME-MAST, version 3.0 ([28]; http://www.sdsc.edu/MEME). Peptide sequences common to proteins that are targeted for rapid turnover were identified with PEST-FIND software (http://www.at.embnet.org/embnet/tools/bio/PESTfind).

Sequences from the following list (accession numbers in parentheses) were used for the alignment and phylogenetic analysis of sub-subfamilies I–V: OsNAC6 or ONAC048 (AK068392), OsNAC5 or ONAC009 (AK063399), OsNAC4 or ONAC068 (AK073848), OsNAC3 or ONAC067 (AK073667), ONAC010 (AK063406), ONAC002 (AK104712), NAP or ANAC029 (CAA10955), jasmonate-regulated NAC1 or ANAC056 (JaNAC1, NP_188170), jasmonate-regulated NAC2 or ANAC055 (JaNAC2, AAM61076), AtNAM or ANAC018 (AAD17314), ANAC or ANAC019 (NP_175697), ATAF1 or ANAC002 (NP_171677), ATAF2 or ANAC036 (CAC35884), ANAC025 (NP_564771), ANAC032 (NP_177869), ANAC047 (NP_187057), ANAC081 (NP_680161), ANAC102 (NP_201184), BnNAC5-8 (AY245883), BnNAC5-7 (AY245882), BnNAC5-11 (AY245884), BnNAC5-1 (AY245881), BnNAC3 (AY245880), BnNAC18 (AY245886), BnNAC14 (AY245885), BnNAC1-1 (AY245879), SsNAC19 (CA140418), SsNAC14 (CA132543), and SsNAC8 (CA120159).

### 2.3. Identification of the SsNAC23 and SsMAPK (*Saccharum sp.* mitogen-activated protein kinase) full-length sequences

To obtain the full-length cDNA sequences of *SsNAC23* and *SsMAPK*, the longest EST of each SAS was re-sequenced using an ABI PRISM 3700 DNA sequencer and M13 forward and reverse primers. Sequences were analyzed using the PHRED-PHRAP package (http://www.phrap.org/). The identified full-length sequences of *SsNAC23* and *SsMAPK* were submitted to GenBank database under accession numbers AY742218 and AY738652, respectively.

### 2.4. RNA gel-blot analysis

RNA was isolated using Trizol reagent (Invitrogen, USA) as described by the manufacturer. Fifteen micrograms of total RNA were electrophoresed in a 1% (w/v) agarose gel containing formaldehyde and transferred to a Hybond-N+ filters (Amersham Pharmacia Biotech, USA) as described by Sambrook et al. [29]. RNA gel-blots were hybridized at 42 °C for 16 h with a 3′-fragment of *SsNAC23* or full-length cDNA of *SsMAPK*, both labeled with $\alpha$-$^{32}$P dCTP [29]. The blots were then washed at high stringency and exposed to imaging plates. Each RNA gel-blot experiment was repeated twice. Later, the blots containing total RNA from sugarcane organs/tissues and different treatments were re-hybridized under the same conditions with a $\alpha$-$^{32}$P-labeled sugarcane full-length wheat [30] ortholog actin cDNA (accession number AY742219).

### 2.5. Nuclear localization of SsNAC23

The *SsNAC23* coding region (891 bp) was amplified by PCR using internal primers (forward: 5′-GGATCCGCG-ATTGAGCGAGGGAGGAA-3′ and reverse: 5′-GGAT-CCGAACGGCTTGCCCCAGTACA-3′). The single PCR product was isolated from agarose gel using the Concert™ rapid gel extraction system (Invitrogen, USA) and then subcloned into the pGEM-T EASY vector (Promega, USA). The vector was subsequently digested at the *Bam*HI restriction site included at the 3′- and 5′-ends and ligated into the pBI221 plant expression vector (Clontech, USA) downstream of the CaMV 35S RNA promoter and fused in-frame upstream of the *β-glucuronidase* (*GUS*) reporter gene. The construct was checked by appropriate digestion and DNA sequencing.

A helium biolistic gene transformation system (Bio-Rad, USA) was used to transiently transform onion (*Allium cepa*) epidermal cells with the *SsNAC23*::*GUS* fusion construct. Onion cells were also transformed with pBI221 as a control. Inner epidermal single cell layers were peeled from the onions and placed on MS solid medium [31]. Five micrograms of plasmid DNA purified on a column (Qiagen, USA) and dialyzed against water was precipitated onto 1.6 mm gold particles (Bio-Rad, USA) using 2.5 M $CaCl_2$ and 1 M spermidine (Sigma, USA). DNA-coated particles were rinsed with absolute ethanol, re-suspended in ethanol, and used to bombard the onion cells at 1300 psi. The bombarded tissue was incubated for approximately 24 h in the dark before histochemical analysis. The SsNAC23::GUS fusion protein was detected in transformed cells by histochemically staining the bombarded onion epidermis with X-gluc [32]. The position of the nucleus in each cell was determined by staining with 4′,6-diamidino-2-phenylindole (DAPI, [32]). X-gluc staining was detected and photographed in transformed cells using a Nikon Eclipse E600 microscope (Nikon, Japan) with bright-field optics, and the DAPI-stained nuclei of the same cells were visualized using fluorescence optics.

### 2.6. Determination of the SsNAC23 DNA-binding NAC domain three-dimensional structure

A multiple peptide alignment was generated with ClustalW [33] and displayed with BOXSHADE (http://helix.nih.gov/apps/bioinfo/boxshade.html). The structure of the DNA-binding NAC domain of *Arabidopsis* ANAC (abscisic-acid-responsive NAC, pdb 1ut7, [34]) was used to predict the three-dimensional structure of the SsNAC23. We used the 68% sequence identity between ANAC (chain

A) and SsNAC23 to map the SsNAC23 NAC domain sequence in the three-dimensional structure described for chain A of ANAC in the 1ut4 structure. The atomic coordinates of the 1ut7 pdb structure, solved at 1.7 Å resolution, were used as a template for comparative modeling based on the satisfaction of spatial restraints [35] implemented in the program MODELLER 6v2 [36]. Twenty-five different models of the structure of SsNAC23 were generated. The quality of the predicted fold was evaluated in MODELLER using the score of the variable target function method [36]. The stereochemical quality of the five best scoring models was assessed by PROCHECK software [37] at the same resolution as the 1ut7 structure. The final model was selected based on the overall stereochemical quality.

To refine the final molecular model of the SsNAC23 structure, additional energy minimization and equilibrating molecular dynamics simulations were done using the program GROMACS 3.14 [38] on a Linux workstation. The initial protein model was submitted to a steepest-descent (SD) energy minimization (5000 steeps) to remove bad van der Waals contacts. For further relaxation of the structure, an un-restrained molecular dynamics simulation was done for 100 peak seconds (p.s.) in water (spc216 model) with Berendsen-type temperature (312 K) and pressure (1 atm) coupling in a $30 \times 30 \times 30$ simulation cell [38,39], using the PME (particle mesh Ewald) method [40]. An unrestrained, multiple step, conjugate-gradient (CG) minimization process (to 0.1 kJ $mol^{-1}$ $nm^{-1}$) was used to obtain the minimum energy structure of the protein. A combinatorial extension method (CE; [41]) was used to superimpose the target and modeled backbones. The SsNAC23 model obtained is available for non-profit use directly from the authors.

## 3. Results

### 3.1. Identification of sugarcane NAC domain containing-sequences and phylogenetic analysis of stress-responsive NAC proteins

An ordered set of NAC domain protein sequences from *Arabidopsis* (ANAC001 to ANAC105) and rice (ONAC001 to ONAC075), described by Ooka et al. [16], was used as driver to identify and classify sugarcane NAC domain-containing sequences encoded by SASs produced by the assembly of 237,954 ESTs [24]. Queries consisting of the full-length protein sequence of each member of the *Arabidopsis* and rice NAC domain protein families were used to screen the SUCEST database. Candidate SASs were selected based upon an e-value $\leq 10^{-30}$ and the presence of at least four out of five conserved N-terminal NAC subdomains (A through D, [12]). Using this approach, we identified 35 EST clusters corresponding to 26 non-redundant *SsNACs* (Table 1).

A phylogenetic analysis of selected SsNAC sequences containing the entire 160 amino acid-long NAC domain sequence (subdomains A through E, [12]) along with *Arabidopsis* and rice NAC domain proteins [16] was performed using the neighbor-joining method. Twenty SsNACs were separated into 14 previously described NAC subfamilies (data not shown). Of the identified *SsNACs*, *SsNAC23* was previously shown to be regulated by chilling stress [7]. SsNAC23, together with other three sugarcane NAC domain proteins (SsNAC8, SsNAC14 and SsNAC19), belong to the ATAF, NAP, and OsNAC3 subfamilies, respectively (Fig. 1A). The members of these subfamilies have been proposed to be involved in the response to biotic and abiotic stress [16].

Detailed phylogenetic analysis showed that these three subfamilies could be further divided into five sub-subfamilies, namely dicot ATAF (I), monocot ATAF (II), OsNAC3 (III), ANAC (IV), and NAP (V) (Fig. 1A). To determine the reliability of the classification of these five sub-subfamilies using the neighbor-joining method, we also performed a phylogenetic analysis using the maximum parsimony (MP) method (PROTPARS in Felsenstein's program package PHYLIP, version 3.5c). The reliability of the MP results was assessed by examining the level of significance of the MP analysis using bootstrap testing with 1000 replicates and similar results to the neighbor-joining method were observed (data not shown). In addition, searching of published expression profile data [5,7,8,20,21,42,45] revealed that several biotic and abiotic stress-responsive *ANACs* and *ONACs* occur in all five sub-subfamilies, mainly in sub-subfamilies I and II (Fig. 1B). For instance, wounding induces the expression of *ATAF1* and *ATAF2*, although *ATAF1* is also induced by cold and drought stress, as well as treatment with ABA (Fig. 1B; [5,20]). Other *Arabidopsis* NAC domain genes have been reported to be induced or repressed by biotic and abiotic stress, including the already described *AtNAM* (*Arabidopsis NAC DOMAIN PROTEIN NAM*) and *NAP* (Fig. 1B; [8]). Of the proteins classified in the monocot ATAF sub-subfamily, OsNAC6 and SsNAC23 are encoded by stress-induced genes (Fig. 1B; [7,42]). Finally, besides *ANACs*, *ONACs*, and *SsNACs*, the *Brassica NACs* classified in the dicot ATAF sub-subfamily have also been reported to be responsive to biotic and abiotic stress (Fig. 1B; [21]).

To complement the phylogenetic data of previously described stress-induced *NAC* domain genes of sub-subfamilies I and II, which include *SsNAC23*, we identified conserved motifs in the carboxy terminus (transcriptional activation region; [12]) of the proteins of these sub-subfamilies (Fig. 1C). Fig. 1C illustrates the position of each C-terminal motif found in representative proteins of sub-subfamilies I and II. Based on these domains, the dicot ATAF sub-subfamily could still be divided into ATAF1- and ATAF2-like proteins that share transcriptional activation regions with serine-rich (M1) and acidic amino acid-rich (M2) motifs (Fig. 1D). Moreover, while most ATAF2-like

Table 1
Summary of the *SsNACs* identified in the SUCEST database

| Name | SAS ID | Accession no.[a] | EST no. | Highest identity[b] | Accession no.[c] |
|---|---|---|---|---|---|
| SsNAC1 | SCAGRT2039C10 | CA137402 | 6 | ONAC040 | BAD09611 |
| SsNAC2 | SCBGLR1118D05 | CA119152 | 15 | ANAC014 | NM_202225 |
| SsNAC3 | SCCCCL4009H05 | CA094553 | 2 | OsNAC7 | AB028186 |
| SsNAC4 | SCCCCL5072D12 | CA095553 | 3 | ANAC033 | NM_202449 |
| SsNAC5 | SCCCCL6005A08 | CA096752 | 6 | OsNAC8 | BAA89802 |
| SsNAC6 | SCCCFL4092A07 | CA235099 | 1 | *Triticum* sp.GRAB2 | AJ010830 |
| SsNAC7 | SCCCLB1024F12 | CA111892 | 5 | ANAC079 | NM_120850 |
| SsNAC8 | SCCCLR1070C08 | CA120159 | 19 | OsNAC3 | AB028182 |
| SsNAC9 | SCCCLR1C04H01 | CA189789 | 18 | ANAC083 | NM_121321 |
| SsNAC10 | SCCCRZ1004G12 | CA147133 | 3 | ANAC036 | NM_127259 |
| SsNAC11 | SCCCST3005F09 | CA180484 | 4 | ANAC073 | NM_118992 |
| SsNAC12 | SCEPLB1042D02 | CA112161 | 22 | ANAC053 | NM_111885 |
| SsNAC13 | SCEPRT2048G05 | CA138286 | 3 | ANAC021 | NM_179486 |
| SsNAC14 | SCEQRT1024G08 | CA132543 | 16 | OsNAC5 | AB028184 |
| SsNAC15 | SCEQRT1028D09 | CA132861 | 1 | PhNAM2 | AAM34765 |
| SsNAC16 | SCEQRT1031E06 | CA133129 | 11 | ONAC007 | AK062952 |
| SsNAC17 | SCEQRT2030D12 | CA138754 | 6 | ANAC041 | NM_128908 |
| SsNAC18 | SCEZRT2019E05 | CA140061 | 3 | ANAC087 | NM_203071 |
| SsNAC19 | SCEZRT2023G12 | CA140418 | 8 | *Triticum* sp. GRAB1 | CAA09371 |
| SsNAC20 | SCJFRT1009A08 | CA133624 | 2 | ANAC100 | NM_125536 |
| SsNAC21 | SCJLRT1015C08 | CA135079 | 2 | ANAC056 | NM_112419 |
| SsNAC22 | SCJLRT1018B06 | CA135405 | 1 | *P. hybrida* NAC protein | AC007454 |
| SsNAC23 | SCCCCL4007D12 | CA094345 | 28 | OsNAC6 protein | AB028185 |
| SsNAC24 | SCMCLB2082D05 | CA271774 | 1 | *N. tabacum* NAC protein | AB021178 |
| SsNAC25 | SCMCRT2106G06 | CA142777 | 1 | ANAC021 | NM_179486 |
| SsNAC26 | SCQGRT3042B03 | CA260833 | 1 | ANAC104 | NM_125849 |

[a] Accession numbers refer to SAS containing the NAC domain sequence.
[b] Highest identity based on [16] when applicable.
[c] Accession numbers refer to NAC domain proteins homologous to identified SsNACs.

proteins have a glutamine-rich motif (M3), ATAF1-like proteins include a motif containing proline-rich regions (M4) as well as a hydrophobic amino acid-rich motif (M5, Fig. 1C and D). The motifs found in the proteins belonging to the monocot ATAF sub-subfamily were similar, except for motif M2 that is present only in OsNAC6 and SsNAC23. Motifs M1, M2, and M5 were conserved in both sub-subfamilies (dicot and monocot ATAF, Fig. 1C).

To assess the expression profile of all *SsNACs* in several sugarcane tissue/organs, we used the cDNA library information available in the SUCEST database (http://sucest.lad.ic.unicamp.br/public). EST counting was used to access information on gene expression levels [43]. To avoid potentially misleading interpretations of gene expression levels, the records of EST sequences were individually inspected and clone IDs were retrieved and displayed in a catalog such that EST sequences obtained from opposite ends of the same cDNA clone were scored as one. Non-normalized cDNA libraries were grouped into library pools (Table 2) and used for further analysis. Additionally, a SAS encoding an actin (accession number CA192497, referred to hereafter as *Ssactin*) was included in the analysis as an internal control. The relative abundance of ESTs was calculated as the number of *SsNAC* and *Ssactin* ESTs in a given library pool divided by the total number of ESTs in that pool (Fig. 2). Most *SsNACs* showed a similar pattern of expression in the root pool, except for *SsNAC14* and *SsNAC23*, which showed higher expression than the other members. Thirty percent (8/26) of *SsNACs* displayed tissue/organ-enriched expression, while the remaining were represented in several tissue/organs analyzed (Fig. 2).

### 3.2. Structure and cellular localization of SsNAC23

Since *SsNAC23* was shown to be strongly upregulated in response to chilling stress in sugarcane [7], we focused on characterizing the SsNAC23 protein in detail. The longest cDNA clone representing SsNAC23 in the SUCEST database (Table 1) was completely sequenced. A 1.56 kb cDNA was identified and found to contain an 891 bp open reading frame (ORF) encoding a 296 amino acid protein with features characteristic of known NAC domain proteins, including the five N-terminal A through E NAC subdomains (Fig. 3) and a predicted molecular mass of 34 kDa. SsNAC23 presented 82% identity and 83% similarity at the amino acid level to OsNAC6 [12] (Table 1). Sequence analysis of the SsNAC23 transcriptional activation region (Fig. 1C) identified a non-cannonical PEST domain (peptidic sequence enriched in proline, glutamic acid, serine, and threonine [44]) that is present in a motif conserved among dicot and monocot ATAF sub-subfamily members (M1; Fig. 1D). The PEST domains are common in regulatory proteins that are rapidly targeted for degradation by the ubiquitin-26S proteosome pathway [44].

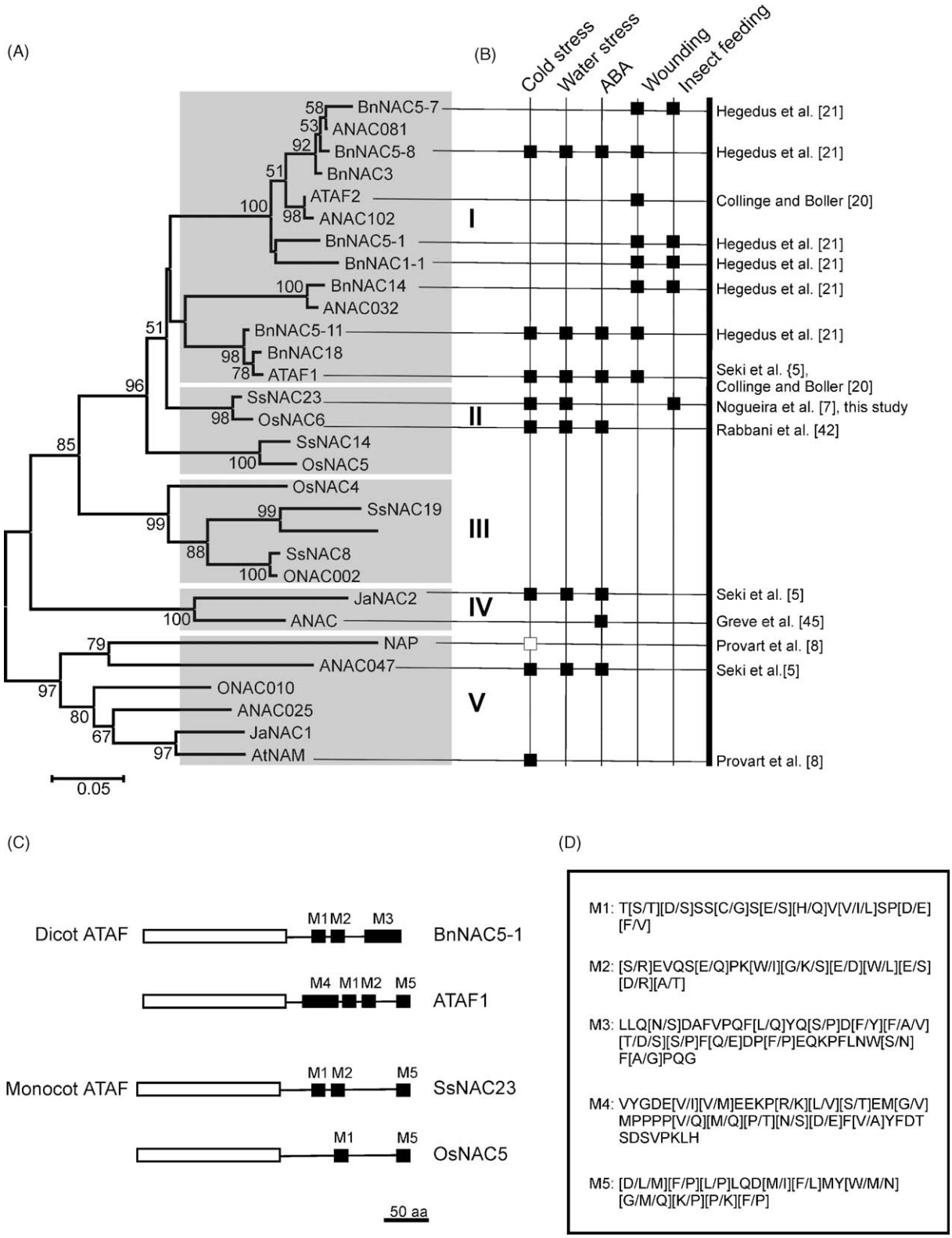



Fig. 1. Comparison of deduced amino acid sequences and expression patterns of *NAC* domain genes and conserved motifs in transcriptional activation regions. (A) Un-rooted phylogram of *Arabidopsis*, *B. napus*, rice and sugarcane NAC domain proteins. The five sub-subfamilies distinguished within the ATAF, NAP and OsNAC3 subfamilies [16] are shown in gray rectangles. The phylogram was generated as a consensus of 1000 bootstrap replicates by the neighbor joining method (the bootstrap values are indicated close the branch divisions, when >50%). The scale bar indicates the relative amount of change along branches. The accession numbers are given in Section 2. (B) Comparison of the expression profiles of *AtNACs*, *OsNACs*, *BnNACs*, and *SsNACs* and their respective references. Full squares represent an induction at the transcription level, whereas empty squares represent downregulated mRNA levels in response to the stress/treatment under consideration. (C) NAC domains are identified by white rectangles while motifs (M1 to M5) in transcriptional activation regions are identified by black

Table 2
Summary of the tissue/organ-specific library pools from the SUCEST database

| EST library ID[a] | Tissue/organ-specific library pools[b] | Total no. of ESTs |
|---|---|---|
| CL6 | Calli | 5518 |
| AM1, AM2, LB1, LB2 | Meristematic tissues | 39116 |
| FL1, FL3, FL4, FL6, FL8 | Floral organs | 52430 |
| LV1, LR1, LR2 | Leaves | 19676 |
| RT1, RT2, RT3 | Roots | 25302 |
| ST1, ST3, SB1, RZ1, RZ2, RZ3 | Stem | 49785 |

[a] Identification of each EST library within the SUCEST database [24].
[b] EST libraries were grouped into seven library pools according to the source tissues from which the library was derived.

Certain basic residues in the NAC domain sequence involved in DNA-binding are conserved in *Arabidopsis* [34]. The basic amino acid-rich extremities of the β-strands 4 and 5 that protrude from the face of the NAC dimer are likely to be important at the DNA-protein interface [34]. This is in agreement with biological data observed in yeast expressing deletion mutants of AtNAM [13]. A sequence alignment using 15 NAC sequences from different plant species showed that the basic residues conserved among *Arabidopsis* NACs are also present in other plants (Fig. 3). Moreover, since the most conserved parts of the primary structure among the NAC domains analyzed coincided with the secondary elements described in the ANAC structure [34], we hypothesized that the overall folding of the secondary structure element might be conserved in all NAC proteins.

Based on this reasoning, molecular modeling was used to obtain the putative structure of the NAC domain of the SsNAC23 protein. A Ramachandran plot of the backbone angles of modeled SsNAC23 structure showed that no residues lay in the disallowed regions, and stereochemical parameters checked with Procheck software [37] were as good as or better than expected at the 95% confidence level (data not shown). A natural consequence of the homology modeling procedure is the targeting of the initial 3-D coordinates from the initial used structure (ANAC) to the obtained model (SsNAC23). To avoid the unnatural identity of the obtained initial model to ANAC structure in the refinement process of our model, we performed 100 p.s. of un-restrained molecular dynamics in the initial model of the SsNAC23 using GROMACS [38], with additional energy minimization of the final structure.

The final SsNAC23 model fold maintained the overall structure of a twisted anti-parallel β-sheet packed against an N-terminal α-helix [34]. The secondary structure of the modeled SsNAC23 was almost identical to the ANAC structure, with minor differences (Fig. 4A). The backbone superimposition of the SsNAC23 model and ANAC structures

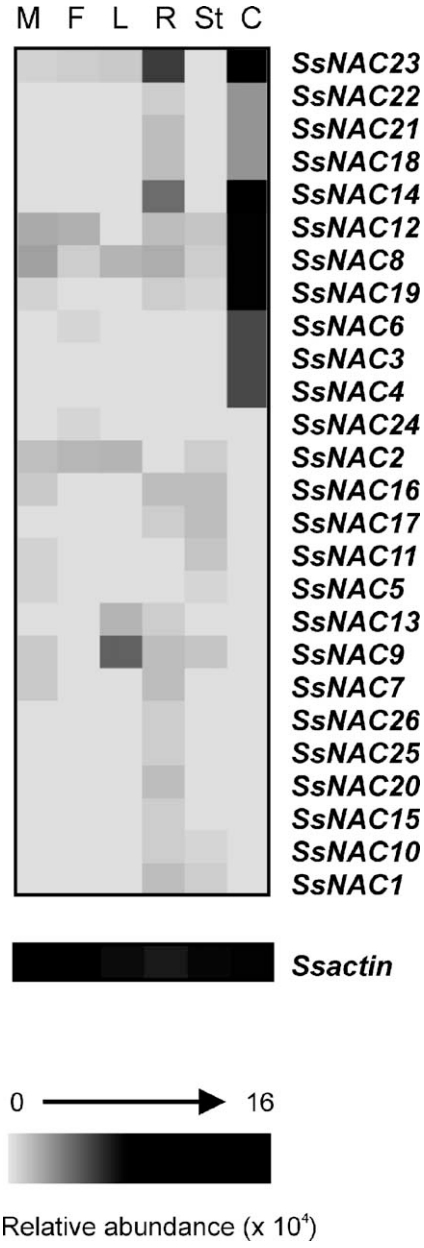



Fig. 2. Relative abundance of *SsNAC*-representing clusters in sugarcane tissue/organ-enriched library pools. Each *SsNAC* is represented by a single row and each tissue/organ library pool is represented by a single column. Since *actin* is reported to be present in most plant tissues/organs, an SAS encoding sugarcane actin (Ssactin) was included as an internal control. The relative abundance of ESTs was calculated as the number of ESTs in a given library pool divided by the total number of ESTs in that pool (Table 2). The normalized values were analyzed using the CLUSTER program [52] and the image was generated by Tree View software downloaded from http://rana.lbl.gov. The *SsNAC* and *Ssactin* names are shown on the right and the color intensity scale representing the relative abundance is indicated at the bottom. M, meristematic tissues; F, floral organs; L, leaves; R, roots; St, stem; C, calli.

rectangles. M1, serine-rich motif; M2, acidic amino acid-rich motif; M3, glutamine-rich motif; M4, proline-rich motif; M5, hydrophobic amino acid-rich motif. (D) Amino acid sequences of all motifs found by the MEME-MAST program [28].

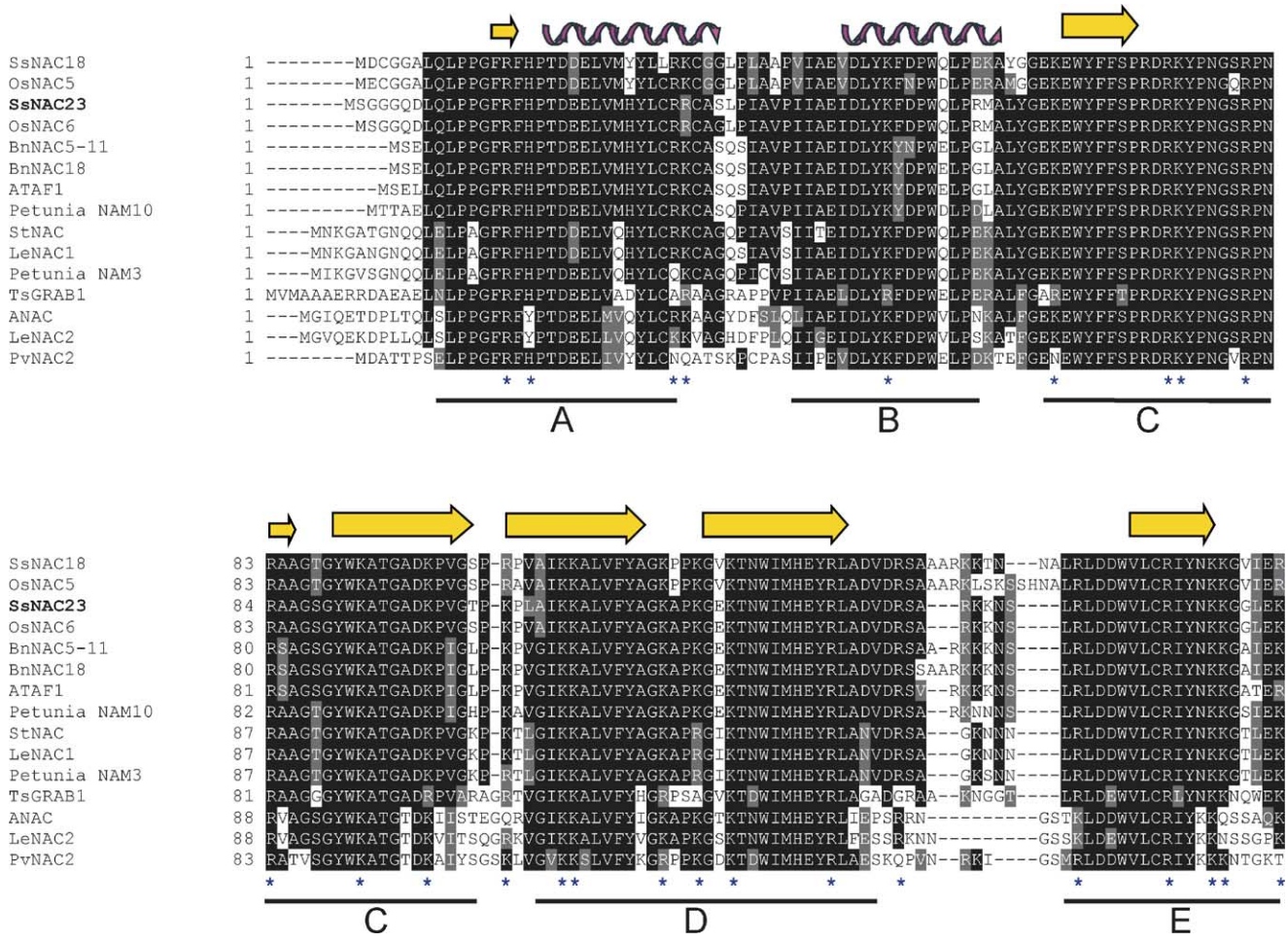



Fig. 3. Alignment of 15 NAC domain protein sequences from different plant species. Only amino acids belonging to the NAC domain sequence were used in the alignment. Subdomains A through E are shown by lines below the sequences. Amino acid identification: white on black, identical residues and white on light gray, conserved residues. Arg, Lys and His residues, conserved in these sequences, are marked by a blue star and the secondary elements described in the ANAC structure [34] are shown at the top of the alignment, where curved arrows represent α-helices and dark yellow arrows represent β-strands. The accession numbers for each protein sequence (in parentheses) are given in Section 2, except for petunia NAM3 and NAM10 (AAM34766 and AAM34773, respectively), potato StNAC (CAC42087), tomato LeNAC1 and LeNAC2 (AAR88435 and AAF04915, respectively), wheat TsGRAB1 (CAA09371), and *Phaseolus* PvNAC2 (AAK84884). SsNAC23 protein is in bold.

showed backbone root mean standard deviation (R.M.S.D.) equal to 1.5 Å, indicating high conservation in the domain structure (Fig. 4A). The SsNAC23 dimer structure had a shape fold configuration (Fig. 4B) with a positively charged internal surface rich in Arg, Lys and His. The residues that allow the formation of the interdomain salt bridge (Arg19 and Glu26 in Fig. 4A) and the Glu67, which is important for coordinating the Arg19 position at the domain interface, were all conserved. All of the basic residues at the DNA-contacting surface were conserved when compared with the structure of ANAC (Fig. 4C), indicating that the SsNAC23 molecular model retained all of the important features that characterize DNA-contacting proteins.

A degenerate bipartite nuclear localization signal (NLS), IKKALVFYAGKAPKGEKTN [45], conserved in several NAC domain proteins among different plant species, was identified in the SsNAC23 protein (Fig. 5A). The presence of this sequence suggested that SsNAC23 could be targeted to the nucleus. To test this hypothesis, the *SsNAC23* ORF was fused to the N-terminal of the *Escherichia coli GUS* gene using the vector pBI221 (Section 2). Recombinant DNA constructs encoding the SsNAC23::GUS fusion and a GUS protein alone were introduced into onion epidermal cells by particle bombardment [32]. Most of the transiently expressed SsNAC23::GUS fusion protein was targeted to the nucleus (Fig. 5B). In contrast, the GUS protein alone was detected throughout the cytosol (Fig. 5B). These results agree with previous findings for the NAC1 and ANAC proteins in *Arabidopsis* [15,45].

### 3.3. SsNAC23 gene expression under chilling stress

Since NAC domain genes have a high degree of sequence conservation in the 5′-coding region, an *Xba*I-digested 3′-DNA fragment of *SsNAC23* cDNA was generated as a specific probe for all RNA gel-blot analyses. In general,

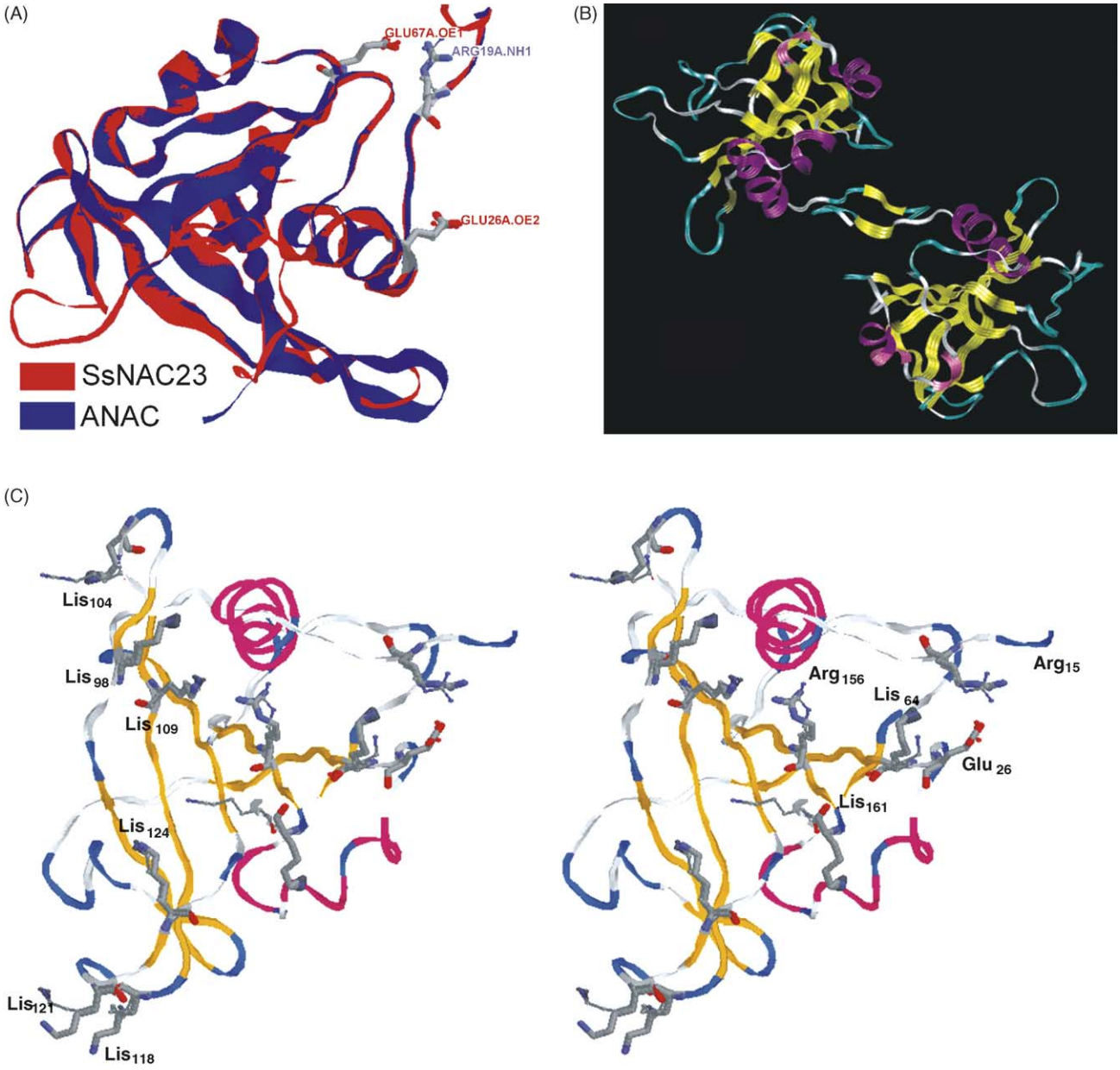



Fig. 4. Structural conservation of the NAC domain of the SsNAC23 protein. (A) Structural alignment of the obtained SsNAC23 model (in red) and the ANAC structure (in blue) backbones. The conserved Arg19 and Glu26 residues responsible for interdomain salt bridges are represented in CPK (thick lines for the SsNAC23 model and thin lines for the ANAC structure). (B) Structure of the SsNAC23 backbone dimer model represented and colored according to the secondary structure presented in Fig. 3. (C) Stereo representation of the DNA-contacting basic surface of the SsNAC23 structure colored according to the secondary structure presented in Fig. 3. Conserved basic residues in the internal surface, important for the DNA-binding function (Arg and Lys), were structurally aligned and represented in CPK (thick lines for the SsNAC23 model and thin lines for the ANAC structure).

RNA gel-blot results for tissue/organ-enriched expression profiling confirmed the findings of the in silico analysis (Fig. 2). *SsNAC23* was moderately expressed in roots and young leaves of non-stressed, 1-month-old sugarcane plantlets, was only marginally detected in stems and was not detected in apical meristematic tissue and floral organs (Fig. 6).

To examine the expression of *SsNAC23* in response to low temperature in detail, 3-month-old sugarcane plantlets were kept at 4 °C for 48 h (Fig. 7A). The *SsNAC23* expression was induced within 1 h and increased steadily in the 48 h after the onset of chilling treatment. Since sugarcane growth is strongly affected by temperatures below 16 °C [2], we evaluated the *SsNAC23* expression profile under moderately low temperature. Interestingly, *SsNAC23* expression was not induced or repressed in leaves of sugarcane plantlets exposed to 12 °C (Fig. 7A). These results showed that *SsNAC23* is induced at 4 but not at 12 °C. Similar findings have been reported for the rice *lip19* gene encoding a bZIP-type DNA-binding protein, the expression of which is

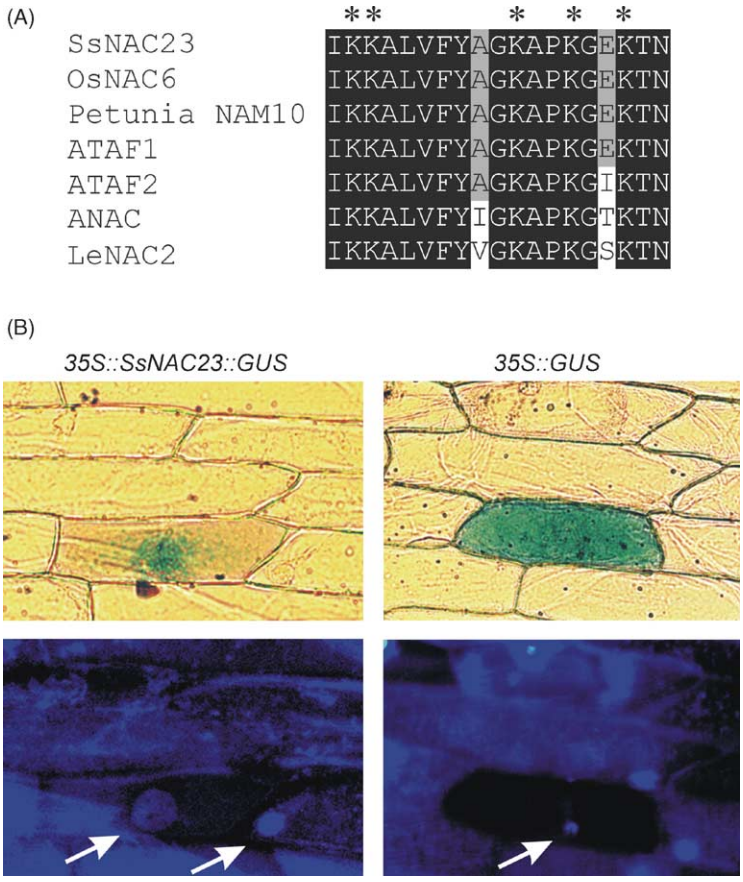



Fig. 5. Nuclear localization of SsNAC23 in onion epidermal cells. (A) Alignment of the degenerate bipartite nuclear localization motif from selected NAC domain proteins using the CLUSTALW program [26]. The accession numbers of SsNAC23, ATAF1, ATAF2, ANAC and OSNAC6 are specified in Section 2 while those of *Petunia* NAM10 and *Lycopersicum esculentum* NAC domain protein 2 (LeNAC2) are given in Fig. 3. Amino acid identification: white on black, identical residues and white on light gray, conserved residues. Asterisks indicate the presence of the basic amino acid lysine (K). (B) Histochemical staining of GUS activity in cells 24 h after particle bombardment (upper panel) and DAPI staining of the same cells to identify nuclei (lower panel). The cells were photographed with a Nikon Eclipse E600 microscope (400×). Arrows indicate the nuclei.

induced at 5 °C [46] but not at 12 °C [47]. The *SsNAC23* mRNA levels remained high as long as the plantlets were kept at 4 °C and returned within 3 h to basal levels when 48-h-cold treated plantlets were transferred to 26 °C (data not shown).

The expression of a rice mitogen-activated protein kinase gene (*OsMAP1* or *OsMAPK5*) was induced at 12 °C but not at 4 °C [47], suggesting that different pathways are triggered in response to distinct ranges of low temperature in rice. These data, together with the distinct regulation of *SsNAC23* at 4 and 12 °C (Fig. 7A), led us to the hypothesis that sugarcane might also show a differential response to extreme (4 °C) and moderate (12 °C) chilling stress at the molecular level. To test this hypothesis, we have identified a cDNA in the SUCEST database that encodes a protein very similar to OsMAPK5 (94% identity at the amino acid sequence level; [48]). Fig. 7A (lower panel) shows a marked accumulation of *SsMAPK* transcripts at 4 and at 12 °C, but with distinct expression profiles. These results, together with the differential expression of *SsNAC23* in response to a specific range of low temperature, suggested that there could be distinct signaling systems that perceive and transduce temperature signals in sugarcane leaves.

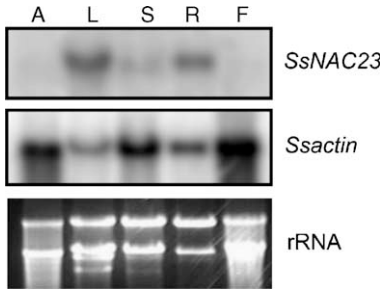



Fig. 6. RNA gel-blot analysis of organ-enriched expression of *SsNAC23*. A, apical meristem; L, leaves; S, stem; R, roots; F, floral organs. All filters were subsequently re-hybridized with the $^{32}$P-labeled sugarcane full-length cDNA (*Ssactin*) as a control.

### 3.4. Effects of water stress, ABA, and herbivory on the expression of SsNAC23

Some cold-inducible genes also respond to water deficit [5] and herbivory [48]. We therefore examined the effects of water stress and herbivory on the expression of *SsNAC23*.

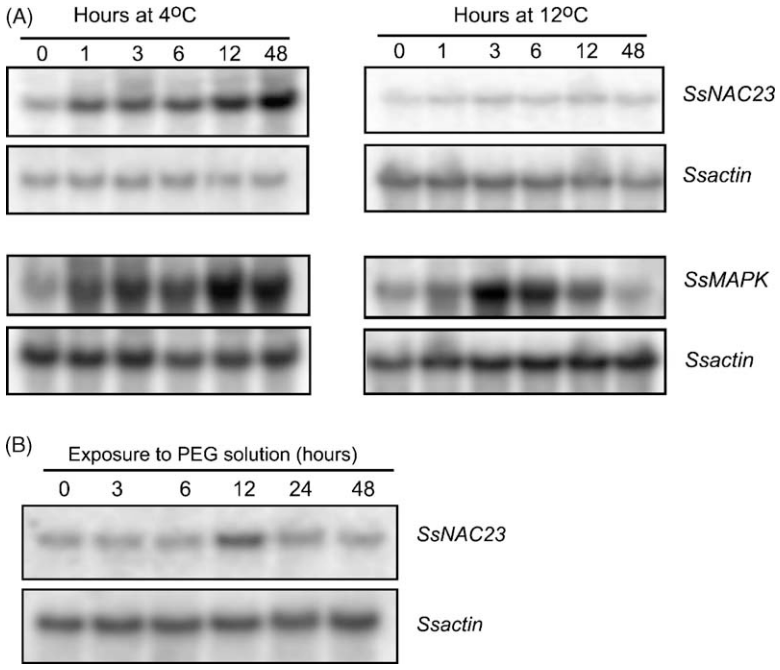



Fig. 7. RNA gel-blot analysis of *SsNAC23* and *SsMAPK* expression. (A) Time course of *SsNAC23* (upper panel) and *SsMAPK* (lower panel) expression in sugarcane plantlets in response to low temperature. Plantlets were exposed to 4 or 12 °C for the indicated periods (h). (B) Plantlets were subjected to PEG-mediated water stress for 3, 6, 12, 24 and 48 h and the blots were hybridized with the $^{32}$P-labeled 3′-fragment of *SsNAC23*. All filters were subsequently re-hybridized with the $^{32}$P-labeled sugarcane full-length cDNA (*Ssactin*) as a control.

Fig. 7B shows that the expression of *SsNAC23* was transiently induced by PEG-mediated water stress (simulated by a solution of PEG 6000; [49]), as also observed for the *SsNAC23* orthologues *ATAF1* and *OsNAC6* [5,42]. Interestingly, *SsNAC23* did not show a well-defined responsiveness to exogenous ABA (data not shown). This finding differs from the previously identified *SsNAC23* orthologous genes *ATAF1* and *OsNAC6* from *Arabidopsis* and rice, respectively (Fig. 1A), which are responsive to exogenous ABA [5,42]. This may be attributable to the different plant species used, as well as the stress treatment and conditions of plant growth. *SsNAC23* transcripts also accumulated transiently after 4 h of tissue damage caused by herbivory (feeding by *Diatraea saccharalis*), with the maximum transcript accumulation at 12 h (Fig. 8). Together, these results suggest a complex regulation of the *SsNAC23* gene during biotic and abiotic stress.

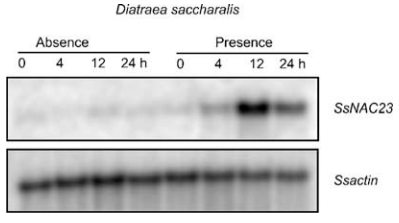



Fig. 8. RNA gel-blot analysis of *SsNAC23* expression in response to insect feeding. Plantlets were subjected to feeding by sugarcane borers for the indicated periods (h). All filters were subsequently re-hybridized with the $^{32}$P-labeled sugarcane full-length cDNA (*Ssactin*) as a control.

## 4. Discussion

### 4.1. The sugarcane NAC domain protein family

In this work, we identified 26 members of sugarcane NAC domain proteins (SsNACs, Table 1) using *Arabidopsis* and rice protein sequences as drivers to probe the SUCEST database. Phylogenetic analysis using protein sequences from four SsNACs and members of the ATAF, NAP and OsNAC3 subfamilies [16], along with *B. napus* NAC domain proteins, showed that these three subfamilies actually represented five sub-subfamilies (Fig. 1A). The inclusion of two dicot and two monocot species improved the resolution of the evolutionary relationships in this set of proteins. The ATAF subfamily showed a clear distinction between dicots and monocots (sub-subfamilies I and II), while the sub-subfamilies ANAC and OsNAC3 contained only dicot and monocot members, respectively (Fig. 1A). Nevertheless, additional information, such as common motifs, expression patterns, and map positions is required to support this orthology [50]. MEME-based analysis of the transcriptional activation regions revealed common motifs (M1, M2 and M5) among proteins belonging to dicot and monocot ATAF sub-subfamilies (Fig. 1C), suggesting that these regions were conserved in parallel with NAC domain

structures. Therefore, it is possible that both NAC subdomains and transcriptional activation regions are involved in determining similar functions in dicot and monocot plant species. Moreover, several members of ATAF sub-subfamilies I and II have been reported to be regulated by biotic and abiotic stress (Fig. 1B). Eight *B. napus* NAC domain genes were differentially regulated in response to wounding, insect feeding, cold stress, and dehydration [21]. *ATAF2* is wounding-inducible, whereas *ATAF1* is predicted to be induced by biotic and abiotic stress, as observed for its orthologues *OsNAC6* and *SsNAC23* [5,7,20,42]. These findings suggest that stress-related functions could be conserved among dicot and monocot members of ATAF sub-subfamilies.

Most *SsNACs* were expressed at low levels in sugarcane tissue/organs (Table 1, Fig. 2), but seven of them were expressed at relatively high levels, especially *SsNAC23*, which is the highest expressed sugarcane *NAC* gene (Table 1). Few *SsNACs* displayed organ/tissue-enriched expression (Fig. 2). This may indicate that most sugarcane NAC domain proteins have similar or even redundant roles in distinct plant organs. In fact, it has been demonstrated that, in *Arabidopsis*, *CUC1* and *CUC2* are functionally redundant genes that are involved in regulating shoot apical meristem (SAM) formation and organ separation [18]. Most *SsNACs* had at least one EST from the root pool, which suggested an important role for these proteins in this organ. Indeed, when *Arabidopsis NAC1* was expressed in the anti-sense orientation, lateral root development was reduced [15].

### 4.2. Stress-inducible SsNAC23 encodes a putative DNA-binding protein that is targeted to the nucleus

SsNAC23 shares 82% identity with the rice OsNAC6 and shows extensive similarity with other NAC domain homologs identified in plants, including *Arabidopsis* ANAC (Fig. 3). The secondary structural elements described in the ANAC structure [34] coincided with the highly conserved domains of the alignment, including certain positively charged residues that were conserved in all plant species studied (Figs. 3 and 4). ANAC is currently the only NAC domain protein for which the DNA-binding domain X-ray structure has been determined [34]. In our modeling study, the NAC domain from SsNAC23 showed a very similar structural fold to the NAC domain of ANAC protein (Fig. 4A). Moreover, the positions of residues that participate in the formation of the dimer through salt bridges were conserved, and there were numerous basic residues in the surface that contact with DNA (Fig. 4B). Furthermore, the conservation of the β-sheet scaffold in the NAC primary sequences among several plant species (Fig. 3), together with the extensive similarity in structure between SsNAC23 and ANAC (Fig. 4), suggested that these proteins can use β-sheet structures not only for binding DNA but also for interacting with other proteins, as proposed for ANAC [45]. Indeed, some BnNACs have been reported to interact with isolated NAC domains from other NAC domain family members [21] and the NAC domain have been found to be sufficient for the interaction between *Arabidopsis* RING-H2-type RHA2a protein and ANAC [45]. Finally, we found that SsNAC23 contains a degenerate bipartite nuclear localization signal that can direct reporter GUS fusions to the nucleus in onion epidermal cells (Fig. 5B). Both *Arabidopsis* NAC1 [15] and ANAC [45] were also able to target green fluorescent protein (GFP) and GUS fusions to the nucleus, suggesting that the predicted degenerate bipartite nuclear localization signal is highly conserved among NAC domain proteins (Fig. 5A).

*SsNAC23* was induced by chilling to 4 °C, but not by exposure a moderately low temperature (12 °C) (Fig. 7A, upper panel). In contrast, the expression of *SsMAPK*, another sugarcane gene possibly involved in cold stress, showed long-term induction during exposure to 4 °C but transient induction at 12 °C (Fig. 7A, lower panel). The distinct expression profiles of *SsMAPK* and *SsNAC23* in response to 4 and 12 °C suggest that sugarcane has mechanisms to distinguish these two ranges of low temperature. Other plant species, such as *Arabidopsis* and rice, have been described as having distinct molecular responses to different ranges of cold stress [8,47]. A possible reason for the existence of distinct transcript accumulation patterns in response to extreme and moderate chilling stress may be related to the fact that exposure to 12 °C for 2 days only includes responses that allow sugarcane metabolism to acclimate to growth at low temperature.

*SsNAC23* may be involved in dehydration-related responses since it was also induced by polyethylene glycol-mediated water stress (Fig. 7B). In addition, sugarcane plantlets exposed to feeding by *D. saccharalis* for varying times showed a high accumulation of *SsNAC23* transcript after 12 h of exposure to the larvae (Fig. 8). Although biotic and abiotic stress is very complex and still not fully understood [1], our results suggest that there is a similar molecular mechanism in the plant responses to these stimuli. Herbivory induced plant responses that possibly involve a jasmonic acid-independent pathway and cross-talk with pathways related to abiotic stress, such as water stress [51]. Hence, it is possible that sugarcane *NAC23* is involved in a pathway of cross-talking between severe chilling (4 °C), PEG-mediated water stress, and herbivory. The functional characterization of SsNAC23 will provide new insights in the role of this protein in any of these stress conditions.

## Acknowledgments

The authors thank A.S. Zanca, S. Meire, D. Althmann and A.L. Beraldo for excellent technical assistance, Dr. Lo Leggio for sharing the ANAC PDB data before publication, and Dr. F. Papes for critical reading of this manuscript. This work was supported by grants to P.A. from FAPESP (no. 98/

12250-0). F.T.S.N., P.S.S., S.R.C., J.H.F., V.E.R.J., and P.P. were recipients of fellowships from FAPESP.

## References


[1] L. Xiong, K.S. Schumaker, J.K. Zhu, Cell signaling during cold, drought, and salt stress, Plant Cell 14 (2002) S165–S183.

[2] P.Y.P. Tai, R.S. Lentini, Freeze damage of Florida sugarcane, in: D.L. Anderson (Ed.), Sugarcane Handbook, Florida Cooperative Extension Service, University of Florida, FL, 1998, pp. 1–3.

[3] K. Maleck, A. Levine, T. Eulgem, A. Morgan, J. Schmid, K.A. Lawton, J.L. Dangl, R.A. Dietrich, The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance, Nat. Gen. 26 (2000) 403–410.

[4] S. Kawasaki, C. Borchert, M. Deyholos, H. Wang, S. Brazille, K. Kawai, D. Galbraith, H.J. Bohnert, Gene expression profiles during the initial phase of salt stress in rice, Plant Cell 13 (2001) 889–905.

[5] M. Seki, M. Narusaka, J. Ishida, T. Nanjo, M. Fujita, Y. Oono, A. Kamiya, M. Nakajima, A. Enju, T. Sakurai, M. Satou, K. Akiyama, T. Taji, K. Yamaguchi-Shinozaki, P. Carninci, J. Kawai, Y. Hayashizaki, K. Shinozaki, Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray, Plant J. 31 (2002) 279–292.

[6] S. Fowler, M.F. Thomashow, *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway, Plant Cell 14 (2002) 1675–1690.

[7] F.T.S. Nogueira, V.E. De Rosa Jr., M. Menossi, E.C. Ulian, P. Arruda, RNA expression profiles and data mining of sugarcane response to low temperature, Plant Physiol. 132 (2003) 1811–1824.

[8] N.J. Provart, P. Gil, W. Chen, B. Han, H.-S. Chang, X. Wang, T. Zhu, Gene expression phenotypes of *Arabidopsis* associated with sensitivity to low temperatures, Plant Physiol. 132 (2003) 893–906.

[9] H. Jin, C. Martin, Multifunctionality and diversity within the plant MYB-gene family, Plant Mol. Biol. 41 (1999) 577–585.

[10] K.B. Singh, R.C. Foley, L. Onate-Sanchez, Transcription factors in plant defense and stress responses, Curr. Opin. Plant Biol. 5 (2002) 430–436.

[11] S. Robatzek, I.E. Somssich, A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes, Plant J. 28 (2001) 123–133.

[12] K. Kikuchi, M. Ueguchi-Tanaka, K.T. Yoshida, Y. Nagato, M. Matsusoka, H.-Y. Hirano, Molecular analysis of the NAC gene family in rice, Mol. Gen. Genet. 262 (2000) 1047–1051.

[13] M. Duval, T.-F. Hsieh, S.Y. Kim, T.L. Thomas, Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily, Plant Mol. Biol. 50 (2002) 237–248.

[14] M. Aida, T. Ishida, H. Fukaki, H. Fujisawa, M. Tasaka, Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant, Plant Cell 9 (1997) 841–857.

[15] Q. Xie, G. Frugis, D. Colgan, N.H. Chua, *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development, Genes Dev. 14 (2000) 3024–3036.

[16] H. Ooka, K. Satoh, K. Doi, T. Nagata, Y. Otomo, K. Murakami, K. Matsubara, N. Osato, J. Kawai, P. Carninci, Y. Hayashizaki, K. Suzuki, K. Kojima, Y. Takahara, K. Yamamoto, S. Kikuchi, Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*, DNA Res. 10 (2003) 239–247.

[17] R.W.M. Sablowski, E.M. Meyerowitz, A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*, Cell 92 (1998) 93–103.

[18] S. Takada, K.-I. Hibara, T. Ishida, M. Tasaka, The *CUP-SHAPED COTYLEDON* gene of *Arabidopsis* regulates shoot apical meristem formation, Development 128 (2001) 1127–1135.

[19] C.W. Vroemen, A.P. Mordhorst, C. Albrecht, M.A.C.J. Kwaaitaal, S.C. de Vries, The *CUP-SHAPED COTYLEDON3* gene is required for boundary and shoot meristem formation in *Arabidopsis*, Plant Cell 15 (2003) 1563–1577.

[20] M. Collinge, T. Boller, Differential induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by *Phytophthora infestans* and to wounding, Plant Mol. Biol. 46 (2001) 521–529.

[21] D. Hegedus, M. Yu, D. Baldwin, M. Gruber, A. Sharpe, I. Parkin, S. Whitwill, D. Lydiate, Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress, Plant Mol. Biol. 53 (2003) 383–397.

[22] T. Ren, F. Qu, T.J. Morris, *HRT* gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus, Plant Cell 12 (20001917-1925).

[23] O. Chazen, P.M. Neumann, Hydraulic signals from the roots and rapid cell wall hardening in growing maize leaves, are primary responses to PEG induced water deficits, Plant Physiol. 104 (1994) 1385–1392.

[24] A.L. Vettore, F.R. da Silva, E.L. Kemper, G.M. Souza, A.M. da Silva, M.I.T. Ferro, F. Henrique-Silva, E.A. Giglioti, M.V.F. Lemos, L.L. Coutinho, et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane, Genome Res. 13 (2003) 2725–2735.

[25] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, K.L. Howe, E.L.L. Sonnhammer, The Pfam protein families database, Nucl. Acids Res. 28 (2000) 263–266.

[26] J.D. Thompson, D.G. Higgins, T.J. Gibson, Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucl. Acids Res. 22 (1994) 4673–4680.

[27] S. Kumar, K. Tamura, I.B. Jakobsen, M. Nei, MEGA2: molecular evolutionary genetics analysis software, Bioinformatics 17 (2001) 1244–1245.

[28] T.L. Bailey, M. Gribskov, Combining evidence using p-values: application to sequence homology searches, Bioinformatics 14 (1998) 48–54.

[29] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.

[30] T. Wicker, N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf, B. Keller, Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution, Plant J. 26 (2001) 307–316.

[31] T. Murashige, F. Skoog, A revised medium for rapid growth and biossays with tobacco tissue culture, Physiol. Plant 15 (1962) 471–497.

[32] M.J. Varagona, R.J. Schmidt, N.V. Raikhel, Nuclear localization signal (s) required for nuclear targeting of the maize regulatory protein Opaque-2, Plant Cell 4 (1992) 1213–1227.

[33] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucl. Acids Res. 25 (1997) 4876–4882.

[34] H.A. Ernst, A.N. Olsen, K. Skriver, A. Larsen, L.L. Leggio, Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors, EMBO Rep. 5 (2004) 1–7.

[35] A. Sali, T.L. Blundell, Comparative protein modeling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[36] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.

[37] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26 (1993) 283–291.

[38] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: A Package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317.

[39] T.M. Louie, H. Yang, P. Karnchanaphanurach, X.S. Xie, L. Xun, FAD is a preferred substrate and an inhibitor of *Escherichia coli* general NAD(P)H:flavin oxidoreductase, J. Biol. Chem. 277 (2002) 39450–39455.

[40] U. Essman, L. Perela, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8592.
[41] I.N. Shindyalov, P.E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, Prot. Eng. 11 (1998) 739–747.
[42] M.A. Rabbani, K. Maruyama, H. Abe, M.A. Khan, K. Katsura, Y. Ito, K. Yoshiwara, M. Seki, K. Shinozaki, K. Yamaguchi-Shinozaki, Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses, Plant Physiol. 133 (2003) 1755–1767.
[43] S. Mekhedov, O.M. de Ilarduya, J. Ohlrogge, Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis, Plant Physiol. 122 (2000) 389–402.
[44] S. Rogers, R. Wells, M. Rechsteiner, Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis, Science 234 (1986) 364–368.
[45] K. Greve, T.L. Cour, M.K. Jensen, F.M. Poulsen, K. Skriver, Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization, Biochem. J. 371 (2003) 97–108.
[46] K. Aguan, K. Sugawara, N. Suzuki, T. Kusano, Low-temperature-dependent expression of a rice gene encoding a protein with a leucine-zipper motif, Mol. Gen. Genet. 240 (1993) 1–8.
[47] J.Q. Wen, K. Oono, R. Imai, Two novel mitogen-activated protein signaling components, OsMEK1 and OsMAP1, are involved in a moderate low-temperature signaling pathway in rice, Plant Physiol. 129 (2002) 1880–1891.
[48] L. Xiong, Y. Yang, Disease resistance and abiotic stress tolerance in rice are inversely modulated by an abscisic acid-inducible mitogen-activated protein kinase, Plant Cell 15 (2003) 745–759.
[49] N.P. Money, Osmotic pressure of aqueous polyethylene glycols: relationship between molecular weight and vapor pressure deficit, Plant Physiol. 91 (1989) 766–769.
[50] M. Vincentz, C. Bandeira-Kobarg, L. Gauer, P. Schlögl, A. Leite, Evolutionary pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein, J. Mol. Evol. 56 (2003) 105–116.
[51] F. Zhang, L. Zhu, G. He, Differential gene expression in response to brown plant hopper feeding in rice, J. Plant Physiol. 161 (2004) 53–62.
[52] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 14863–14868.

***SsNAC23*: um gene de cana-de-açúcar induzido por etanol**

O gene *SsNAC23* pertence a uma família de fatores de transcrição que possuem domínio NAC. Proteínas pertencentes a essa família apresentam região N-terminal conservada, enquanto que a região C-terminal é bastante variável em seqüência de aminoácidos e tamanho (Aida et al., 1997). Membros da família NAC estão envolvidos na regulação transcricional, durante o desenvolvimento vegetal (Ooka et al., 2003) e em resposta a estresses bióticos e abióticos (Collinge et al., 2001; Hegedus et al., 2003).

Nogueira et al. (2003) constataram que o gene *SsNAC23* de cana-de-açúcar apresentou alterações de expressão em resposta a condições de baixa temperatura. A expressão de *SsNAC23* foi fortemente induzida a 4 $^{o}$C, embora o mesmo não fosse observado para temperatura de 12 $^{o}$C. Esse resultado indicou que o gene poderia estar envolvido com mecanismos de resposta bioquímica e/ou fisiológica ao estresse provocado por baixa temperatura em cana-de-açúcar. Outros 25 membros da família de proteínas com domínio NAC também foram identificados nesse estudo, embora a maioria não apresentasse comportamento semelhante ao *SsNAC23* em reposta ao estresse por frio. Análises de RNA-blot (conforme descrito em Nogueira et al., 2005), utilizando amostras de RNA extraídas de folhas de cana-de-açúcar submetidas a estresse hídrico, demonstraram que *SsNAC23* também é induzido por este tipo de estresse. Do mesmo modo, *SsNAC23* apresentou elevada indução da sua expressão após exposição ao ataque de larvas de *Diatrea saccharalis* (Nogueira et al., 2005), indicando que *SsNAC23* é possivelmente um gene com resposta abrangente para vários estresses abióticos e bióticos.

A partir da premissa de que o gene *SsNAC23* apresenta alterações de expressão em resposta a diferentes estresses, resolvemos verificar se o etanol também poderia atuar como indutor da expressão de *SsNAC23* em folhas de cana-de-açúcar.

No estudo realizado com macroarranjos de cDNA (Capitulo I) observamos que um membro da família de proteínas com domínio NAC, um EST com elevada similaridade para a proteína OsNAC de arroz, presente nas membranas utilizadas, não teve sua expressão

significativamente alterada pela aplicação de etanol nas folhas da cana-de-açúcar. O fato desse gene não ter sua expressão alterada por etanol, no entanto, não nos desmotivou a testar o perfil de expressão de *SsNAC23* em folhas de cana-de-açúcar tratadas com etanol.

Membranas de náilon foram confeccionadas com amostras de RNA total extraído de folhas de cana-de-açúcar coletadas 1, 4 e 12 horas após a aplicação de solução 2,5 % (v/v) de etanol e água (controle). As membranas foram hibridizadas com a sonda radioativa sintetizada a partir da região 3' não traduzida do gene *SsNAC23*. Os resultados obtidos no RNA-blot indicam que o gene *SsNAC23* tem sua expressão induzida pelo etanol, principalmente no período de 1 hora após a aplicação. Nos períodos de 4 e 12 horas, também se observou aumento na expressão em relação ao controle embora esse aumento não tenha sido tão marcante como no período de 1 hora após o tratamento com o etanol (Figura 1).

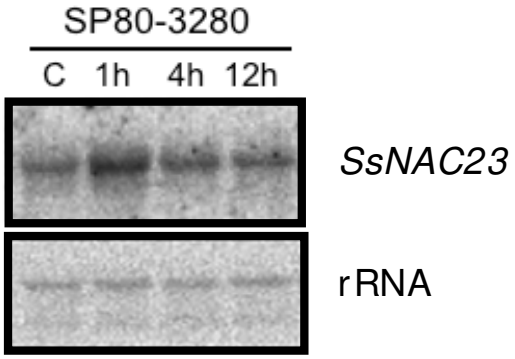



**Figura 1-** Análise de expressão do gene *SsNAC23* em reposta ao etanol. (C) controle com água; (1h) 1 hora após aplicação de etanol; (4h) 4 horas após aplicação de etanol; (12h) 12 horas após aplicação de etanol; (SP80-3280) variedade SP80-3280 de cana-de-açúcar da Coopersucar, utilizada neste estudo.

Em um trabalho recente com *Arabidopsis thaliana* (Tran et al., 2007), foi observado que proteínas com domínio NAC (ANAC019, ANAC055 e ANAC072) atuam como amplificadores da expressão do gene *ERD 1* (Early Responsive to Dehydration Stress). Essa informação está de acordo com resultados obtidos neste estudo, visto que um gene da família ERD foi identificado como altamente induzido em plântulas de cana-de-açúcar, 1h

após a aplicação foliar de solução 2,5 % de etanol, conforme demonstrado no experimento com macroarranjos de cDNA (Capítulo I). Assim, como o gene *SsNAC23* foi induzido por etanol, poderíamos especular que este gene em cana-de-açúcar tenha entre suas funções, a semelhança do que foi observado em *Arabidopsis*, o papel de amplificador ou ativador da expressão do gene *ERD* em cana-de-açúcar.

**Referências**


Aida M., Ishida T., Fukaki H., Fujisawa H., Tasaka M. (1997) Genes involved in organ separation in *Arabidopsis* an analysis of the cup-shaped cotyledon mutant. **Plant Cell** 9: 841-857

Collinge M., Boller T (2001) Differencial induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by Phytophthora infestans and wounding. **Plant molecular Biology** 46: 521-529

Hegedus D., yu M., Baldwin D., Gruber M., Sharpe A., Parkin I., Whitwill S., Lydiate D. (2003) Molecular characterizationof Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress. **Plant Molecular Biology** 53: 383-397

Nogueira F.T.S., De Rosa Jr V.E., Menossi M., Ulian E.C., Arruda P. (2003) RNA expression profiles and data mining of sugarcane response to low temperature. **Plant Physiology** 132: 1811-1824

Nogueira F.T.S., Schlogl P.S., Camargo S.R., Fernandez J.H., De Rosa Jr V.E., Pompermayer P., Arruda P. (2005) SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane. **Plant Science** 169: 93-106

Ooka H., Satoh K., Dói K., Nagata T., Otomo Y., Murakami K., Matsubara K., Osato N., Kawai J., Carninci P., Hayashizaki Y., Suzuki K., Kojima K., Takahara Y., Ya,a,oto K., Kikushi S. (2003) Comprehensive analysis of NAC family genes in *Oryza sativa* ana *Arabidopsis* thaliana. **DNA Research** 10: 239-247

Tran L.P., Nakashima K., Sakuma Y., Osakabe Y., Qin F., Simpson S.D., Maruyama K., Fujita Y., Shinozaki K., Yamaguchi-Shinozaki K. (2007) Co-Expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the *ERD1* gene in Arabidopsis. **The Plant Journal** 49: 46-63

# Avaliação da eficiência do sistema *alc* em plantas transgênicas de cana-de-açúcar (*Sacharum* ssp.)

Sandra R. de Camargo, Eugênio C. Ulian, Sabrina M. Chabregas, Maria Cristina Falco, Marcelo Menossi

**Avaliação da eficiência do sistema *alc* em plantas transgênicas de cana-de-açúcar (*Sacharum* ssp.)**

Sandra R. Camargo[1], Eugênio C. Ulian[2], Sabrina M. Chabregas[2], Maria Cristina Falco[2] e Marcelo Menossi[1]

[1] Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), CP 6010, 13083-875, Campinas, SP, Brasil

[2] Centro de Tecnologia Canavieira, CP 162, 13400, Piracicaba, SP, Brasil.

**Resumo**

O controle da expressão gênica de forma temporal e espacial é uma importante característica para sistemas químicos de indução. Sistemas de indução que promovem a ativação ou inativação da expressão gênica têm se tornado importante ferramenta para a análise genética, estudos bioquímicos e fisiológicos, além de possuírem inúmeras aplicações práticas para o setor biotecnológico. O sistema de indução de expressão gênica através de etanol, chamado *sistema alc*, tem sido utilizado com sucesso em espécies de plantas como *Arabidopsis*, tabaco e batata dentre outras. Resumidamente, o sistema *alc* se caracteriza pela ação de um fator de transcrição, ativado pelo etanol, sobre um promotor induzível. O etanol utilizado como indutor químico apresenta características favoráveis, tais como baixo custo e facilidade de manuseio e aplicação. O objetivo deste trabalho foi desenvolver um vetor de expressão com o sistema *alc* ativando o gene repórter *gus* para ser testado em plantas transgênicas de cana-de-açúcar.

Palavras-chave: expressão gênica, sistema *alc*, etanol, *Saccharum* spp., vetor de expressão

**Abstract**

The spatial and temporal control of genetic expression is an important feature of systems based in chemical induction. Since these systems are able to promote the activation or inactivation of genes it might be useful as tool for genetic analysis, biochemical studies and physiological evaluation. Besides that, genetic systems based in chemical induction have great potential to practical use in the biotechnology field. The system of gene expression based in ethanol induction, often denominated as *alc* system, has been efficiently used in plant species such as *Arabidopsis*, tobacco, and potato. The *modus operandis* of the *alc* system is based in the action of one transcription factor, activated by ethanol, that binds to inducible promoter. The use of ethanol as a chemical inductor involves several advantages such as low cost and easy application and handling. The aim of this work was the development of a genetic construction with the system *alc* controlling the expression of the gene reporter *gus* to be used in the genetic transformation of sugar cane plants.

Key words: gene expression, *alc* system, ethanol, *Saccharum* sp., expression vector

## Introdução

O estudo da expressão de transgenes em plantas é uma das estratégias mais úteis para a caracterização e compreensão da função de um gene e de seu papel na fisiologia da planta (Roslan et al, 2001). Com o recente desenvolvimento de novas técnicas e estratégias aplicadas à engenharia genética, hoje já é possível controlar a expressão de genes de interesse utilizando promotores ativados por indutores químicos. Sistemas de expressão que utilizam esses promotores apresentam importantes vantagens: permitem que a atividade gênica seja induzida em determinadas fases do desenvolvimento da planta; permite modular a expressão do gene e então avaliar o fenótipo da planta para diferentes níveis do produto gênico; verificar se o efeito fenotípico é reversível ou permanece estável em estágios específicos do desenvolvimento por meio da exposição alternada e temporal ao indutor (Garoosi et al, 2005). Em função destes benefícios, os sistemas de expressão gênica que fazem uso da indução química têm sido desenvolvidos e aprimorados para utilização em plantas (Gatz e Lenk, 1998; Jepson et al, 1998; Zuo and Chua, 2000).

O etanol, embora seja um hidrocarboneto de baixo peso molecular, volátil e inflámavel, é considerado como um indutor químico praticamente sem efeitos tóxicos para plantas nos níveis necessários para promover a indução, que variam entre 0,1 a 5% (Salter et al., 1998). Além disso, é de fácil utilização, podendo ser aplicado nas plantas por pulverização, submersão das raízes ou adicionado ao meio de cultura (Salter et al,1998).

Salter et al. (1998) desenvolveram um sistema de expressão controlado por etanol utilizando o sistema gênico *alc*, do fungo *Aspergillus nidulans*, que usa o etanol como indutor. A indução por etanol ocorre pela ação do gene *alcR*, que codifica um fator de transcrição responsável pela ativação da expressão dos genes *alcA* e *aldA,* que por sua vez codificam as proteinas álcool desidrogenase I e aldeído desidrogenase, respectivamente (Pateman et al, 1983). Na presença do etanol a proteína ALCR muda sua conformação estrutural tornando-se ativa e capaz de se ligar nas regiões promotoras dos genes *alcA* e *aldA,* e desta forma, induzindo a transcrição dos mesmos. O funcionamento do sistema *alc*

desenvolvido por Salter et al. (1998) está esquematizado na figura 1. O gene *alcR* é controlado por um promotor geralmente constitutivo, tal como o CaMV35S, para garantir altos níveis da proteína ALCR. Essa proteína por sua vez só é ativada após a ligação com etanol. Já o gene de interesse é controlado pelo promotor do gene *alcA*, cuja atividade é dependente da proteína ALCR. Desta forma, o gene de interesse só será ativado quando na presença de etanol (revisto por Li et al., 2005).

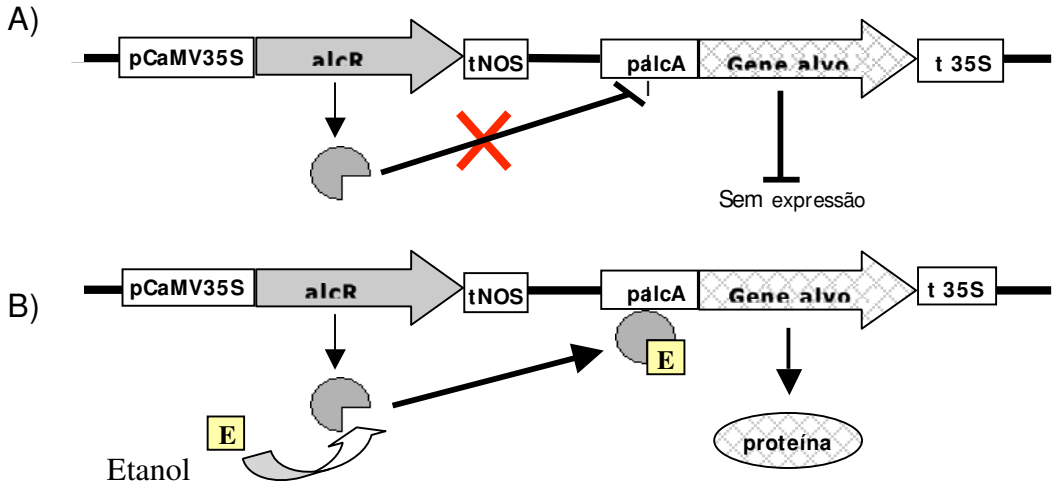



**Figura 1-** Funcionamento do sistema *alc*. a) sem etanol, a proteína ALCR é produzida, mas fica na sua forma inativa, e não há produção da proteína de interesse. b) a aplicação de etanol ativa a proteína ALCR, que aciona o promotor de *alcA*, sendo então, sintetizada a proteína do gene alvo (mofificado a partir de Li et al., 2005).

Uma vantagem do sistema *alc* é que o mesmo é muito sensível à presença de etanol, com rápida indução mesmo em concentrações muito baixas. Salter et al. (1998) observaram que foi necessário a adição de apenas 0,01% (1,7 mM) de etanol no meio de cultura para iniciar a expressão do gene *CAT* (*chloranphenicol acetyl transferase*) após 4 horas de exposição. A indução de expressão gênica usando o sistema *alc* também tem sido estudada através da absorção do etanol pelas raízes ou pulverização das folhas. Esses métodos têm sido utilizados com sucesso em plantas transgênicas de tabaco, no qual a

expressão do gene repórter *CAT* foi observada depois de 2 horas de aplicação atingindo um pico após 96 horas por absorção radicular e 24 horas depois da pulverização das folhas (Salter et al., 1998). De acordo com Sweetman et al. (2002), baixas concentrações de vapor de etanol também induzem de forma eficiente o sistema de expressão do gene *alc* em plantas transgênicas de tabaco, batata (*Solanum tuberosum)* e sementes de oleaginosas (*Brassica napus)*. Tratamento com vapor de etanol em tubérculos de batata geneticamente alteradas com o sistema *alc* induziu a expressão uniforme do gene β-glucoronidase (GUS) (Sweetman et al., 2002). O uso de vapor de etanol como um indutor químico também tem sido relatado em estudos com *Arabidopsis* (Roslan et al., 2001) e tomate (Garoosi et al, 2005). Adicionalmente, o sistema *alc* tem sido usado para promover a expressão espacialmente controlada, através da combinação do gene *alcR* com promotores tecidos específicos que permitem a expressão apenas no tecido alvo (Devaux et al., 2003; Maizel e Weigel, 2004).

Partindo da premissa de que o sistema *alc* pode ser aplicado de forma eficiente na cultura da cana-de-açúcar, o presente trabalho teve por objetivo transformar plantas de cana-de-açúcar com um vetor de expressão contendo o sistema *alc* controlando a expressão do gene repórter *gus* e a partir de plantas de cana-de-açúcar geneticamente transformadas de forma estável com o sistema *alc*, validar a estratégia de controle temporal e espacial utilizando etanol como indutor químico.

## Materiais e Métodos

### 1 - Construção dos vetores

1.1 – Vetor pACNGUS

O gene *gus* e o terminador NOS foram retirados do vetor pBI101.2 (Jefferson et al., 1987) através de digestão enzimática com as enzimas *Bam*HI e *Eco*RI. O inserto foi então clonado no vetor pACN (Syngenta®, EUA; Figura 3A) digerido com as mesmas enzimas (eliminando o gene CAT e o terminador NOS original). O novo vetor foi então denominado pACNGUS (Figura 4A).

1.2 – Vetor pUBILN

Inicialmente, o vetor pAHC17 (Christensen e Quail, 1996) foi modificado para a adição de um sítio de múltipla clonagem logo após o sitío *Bam*HI (Figura 2). Assim, dois oligonucleotídeos complementares foram sintetizados com os sítios de restrição das enzimas *Eco*RV, *Sac*I, *Mlu*I e *Kpn*I (PLSense: 5' GATCCCGATATCGAGCTCACGCG TGGTACC3'; e PLAnti-sense: 5' GATCGGTACCACGCGTGAGCTCGATATCGG 3'). Para a formação da fita dupla 1 nmol de cada um dos oligonucleotídeos foi adicionado a solução contendo 10 mM de Tris-HCl (pH 7,5), 100 mM NACl, 1 mM EDTA. Essa solução permaneceu em banho-maria a 94 $^{o}$C durante 2 minutos e então, foi resfriada gradualmente a temperatura ambiente (Schlögl et al., 2004). O fragmento de DNA dupla fita originado (Figura 2) foi posteriormente ligado ao vetor pAHC17 (linearizado com a enzima *Bam*HI) na proporção de 100:1 (Sambrook et al., 1989). A partir destas alterações o vetor pAHC17 passou a ser denominado de pAHC17L1.

Para a construção do vetor pUBILN o vetor pAHC17L1 foi digerido com as enzimas *Hind*III e *Kpn*I. O produto dessa digestão (promotor UBI:intron:sítio de múltipla clonagem) foi recuperado do gel de agarose 0,8 % e purificado com o kit "GFX DNA and Gel Band Purification" (Amersham Bioscience®, EUA). O terminador NOS foi amplificado por PCR para adição dos sítios de restrição *Kpn*I na extremidade 5' (oligo sense 5'GGTACCGATCGTTCAAACATTTGGC3', o sitio de *Kpn*I aparece sublinhado) e *Hind*III e *Sph*I na extremidade 3' (oligo anti-sense 5' AAGCTTGCATGC CCGATCTAGTAACATAGATGAC3', em sublinhado estão os sítios de *Hind*III e *Sph*I, respectivamente). Os dois insertos (promotor UBI e terminador NOS da nopalina sintase) foram clonados juntos no plasmideo pRT104 (Topfer et al., 1987) digerido com *Hind*III. O novo vetor foi denomidado de pUBILN (Figura 4B).


**Figura 2-** Seqüência de DNA inserida no vetor pAHC17. Os nucleotídeos em azul representam o sítio da enzima *Bam*HI, reconstituído após a ligação com o vetor linearizado; os nucleotídeos em verde representam o sítio da enzima *Eco*RV; os nucleotídeos em vermelho representam o sítio da enzima *Sac*I; os nucleotídeos em preto representam o sítio da enzima *Mlu*I e os nucleotídeos em rosa representam o sítio da enzima *Kpn*I.

1.3 – Vetor pUBILNalcR

O gene *alcR* foi retirado do vetor pbinSRNACatN (Syngenta®, EUA; Figura 3C) através de amplificação por PCR utilizando a enzima DNA polimerase *proofreading* (Platinum Pfx taq DNA polymerase, Invitrogen®, EUA) e oligonucleotídeos específicos adicionados de enzimas de restrição (sense 5´ ACGCGTCAGATACGCGC3´ com sítio para *Mlu*I e anti-sense 5´ GAATTCCTACAAAA A GCTGTCAACTTTCC3´ com sítio para *Eco*RI). O produto da PCR, de aproximadamente 2,6Kb, foi recuperado do gel de agarose e subclonado no vetor pGEMT-Easy (Promega®, EUA) após reação de adenilação (tampão Taq 1x, ATP 0.2 mM, 5 unidades de Taq DNA polymerase, incubação por 30 minutos a 70 °C). Posteriormente, o gene *alcR* foi retirado do vetor pGEMT-Easy, através da digestão com a enzima *Mlu*I, e clonado no vetor pUBILN clivado com a mesma enzima.

1.4 – Vetor pALCs

A construção pUBILN*alcR* (Figura 4C) foi digerida com a enzima *Hind*III, liberando o cassete "promotor ubiquitina:alcR:NOS". Assim, o inserto "promotor ubiquitina:*alcR*:terminador NOS3´" foi recuperado e clonado no vetor pACNGUS linearizado com a mesma enzima, *Hind*III. O vetor resultante, contendo os dois cassetes de expressão, recebeu a denominação de pALCs (Figura 4D). A confirmação da identidade das construções foi obtida através de digestões enzimáticas seguida de eletroforese em gel de agarose e também por seqüenciamento.

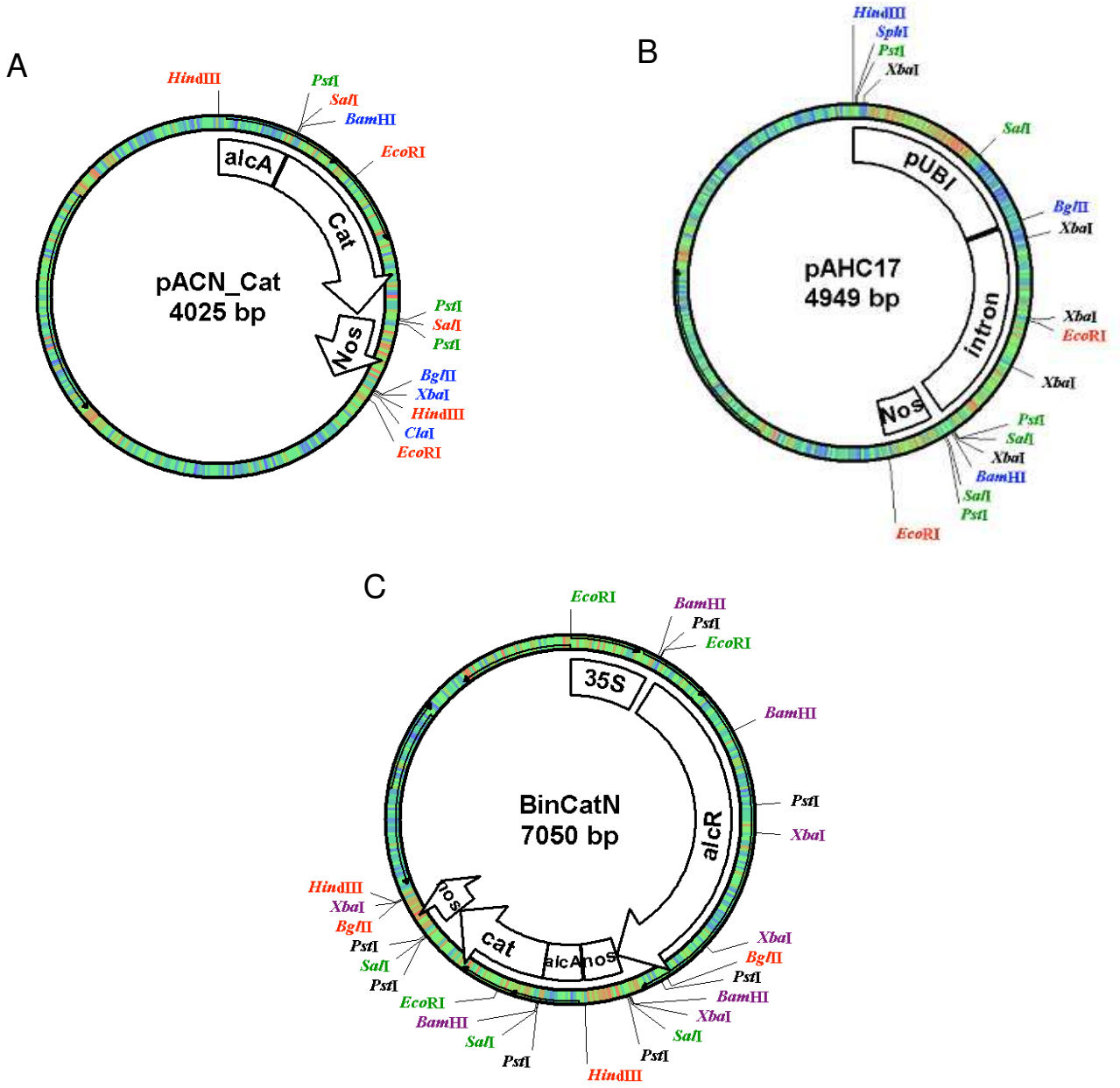



**Figura 3-** Representação esquemática de vetores de expressão utilizados. a) vetor pACN; b) vetor pAHC17 e c) vetor binSRNACatN.

**2 – Material vegetal e transformação de cana-de-açúcar**

Plantas transgênicas de cana-de-açúcar foram produzidas em colaboração com o Centro de Tecnologia Canavieira (CTC, Piracicaba - SP), utilizando-se o genótipo Q117. A transferência dos cassetes de expressão para as células vegetais foi realizada mediante a técnica de biolística, seguindo os protocolos e procedimentos de rotina adotados pelo laboratório de transformação genética da cana-de-açúcar pertencente ao CTC e descritos em detalhes por Saciloto (2003). Sucintamente, para obtenção dos calos foi utilizada a região meristemática apical de planta com idade de 3 meses. O meristema foi então seccionado e colocado em placas com meio sólido CI-3clav (Saciloto, 2003) e posteriormente, foram produzidos calos friáveis adequados para transformação por biolística. Os plasmídeos a serem utilizados (pALCs, pACNGUS e pHA9) foram purificados e concentrados a 1 µg/µL e então, cerca de 10 µg de plasmídeo foram precipitados por centrifugação a 11.200 g por 1 min sob partículas de tungstênio em presença de solução de cloreto de cálcio 2,5 M e espermidina 'free base" 0,1 M. Posteriormente, o "pellet" foi lavado duas vezes em etanol absoluto e duas vezes em etanol 95%, sendo ressuspendido em 30 µL de etanol. Quatro microlitros desta mistura foram utilizados para o bombardeamento de cada placa de Petri. O bombardeamento foi realizado conforme descrito por Rech e Aragão (1998) com modificações citadas em Saciloto (2003). Logo após o bombardeamento, os calos foram colocados em meio de seleção contendo geneticina.

**3 – Isolamento de DNA genômico e amplificação por PCR**

DNA genômico foi isolado a partir de discos foliares utilizando o método CTAB (Doyle & Doyle, 1987). Aproximadamente 50 ng de DNA isolado foram utilizados como molde para reações de amplificação por PCR com o objetivo de confirmar a presença do transgene. Diferentes combinações de pares de oligonucleotídeos foram utilizados para amplificar diferentes regiões do cassete de expressão: região do gene *gus*; região do gene *alcR*; região do terminador NOS e região do gene endógeno *SsNAC23* (Tabela 1). A condição usada para 30 ciclos de amplificação foi a seguinte: desnaturação a 94 $^{o}$C por 1 min, anelamento de *primer* entre 52 a 60 $^{o}$C por 1 min, polimerização a 72 $^{o}$C por 1-1,5 min. O

produto de PCR foi visualizado em luz UV após eletroforese em gel de agarose 0,8 % e coloração com EtBr.

**Tabela 1-** Descrição dos oligonucleotídeos utilizados nas reações de PCR

| Alvo | Par de óligos nº | Sequência (5'-3') | Produto (pb) | Presença |
|---|---|---|---|---|
| GUS | 1 | SENSE – ACGGAGAGCCACAAACG (alcA)<br>ANTI-SENSE – ACGTTCTACAGGACGGACGA (GUS) | ~ 500 | pALCs e pACNGUs |
| GUS | 2 | SENSE – ATGTTACGTCCTGTAGAAACCC (GUS)<br>ANTI-SENSE – GGGGTACCTCATTGTTTCCCT (GUS) | ~ 2.000 | pALCs e pACNGUS |
| tNOS | 3 | SENSE – GGTACCGATCGTTCAAACATTTGGC (NOS)<br>ANTI-SENSE – CCGATCTAGTAACATAGATGAC (NOS) | ~ 260 | PALCs e pACNGUS |
| alcR | 4 | SENSE – CATCTATTCATATGCTCTAACCTTG (Ubi-1)<br>ANTI-SENSE – CAATCCTTGTTCCAACGC (alcR) | ~ 400 | pALCs |
| SsNAC23 | 5 | SENSE – CTACAAGTTCGATCCATGGC (NAC23)<br>ANTI-SENSE – GTCGCCACGAACGGCAACTG (NAC23) | ~ 470 | Todas as plantas |

**4 – Indução por etanol**

Plantas transgênicas e selvagens de cana-de-açúcar foram utilizadas nos experimentos de indução por etanol. Essas plantas foram cultivadas em condição de casa-de-vegetação até atingirem cerca de 2 meses de idade. Dois experimentos de indução foram realizados. No primeiro experimento amostras das folhas das plantas tratadas foram colhidas 5 horas após pulverização com etanol, para detectar a presença da proteína GUS. Em um segundo experimento, essas mesmas plantas foram novamente pulverizadas com etanol, desta vez utilizando uma dose de 4% (v/v). Neste segundo experimento a coleta de amostras de folhas foi feita após 24 horas da aplicação do etanol. Em ambos experimentos, amostras das folhas foram coletadas antes da pulverização com etanol, tanto em plantas transgênicas quanto em plantas selvagens, para serem utilizados como controle negativo para aplicação de etanol.

**5 – Detecção de atividade de GUS**

Para ensaio histoquímico de GUS, os tecidos foliares foram incubados a 37°C, em solução contendo 100 mM de $NaH_2PO_4.H_2O$; 0.5 mM de $K_4Fe(CN)_6.3H_2O$, 10 mM $Na_2EDTA.2H_2O$, 0,1 % de Triton X-100, 1 mM de X-Gluc (McCabe et al., 1988). Após cerca de 40 horas, essa solução foi substituída por etanol 70% e em seguida, as amostras foram observadas visualmente e com auxílio de lupa estereoscópica. A atividade da β-Glucuronidase *in vitro* foi determinada por ensaio fluorimétrico com o substrato 4-methyllumberlliferil β-D-glucuronide (4-MUG) (Jefferson et al., 1987). A liberação de 4-methyl umbelliferone (4-MU) foi medida através de um fluorímetro após 10, 30 e 60 minutos de reação, além disso, a fluorescência das amostras foi visualizada na presença de luz UV. Para determinar a concentração total de proteína presente nos extratos vegetais das amostras foi utilizado o método colorimétrico descrito por Bradfort (1976).

**Resultados e Discussão**

**O sistema *alc* em cana-de-açúcar**

O sistema *alc* de indução de expressão gênica apresenta duas peças fundamentais para o seu funcionamento: o promotor do gene *alcA* e o gene *alcR*. Além disso, para que se torne ativo necessita do contato direto com a molécula do etanol, seu indutor químico. Diversos estudos com o sistema *alc* foram realizados em plantas (Li e al., 2005). A maioria desses estudos utilizam o promotor constitutivo CaMV35S para controlar a expressão do gene *alcR*. No entanto, apesar do promotor CaMV35S ser funcional em folhas de cana-de-açúcar, ele apresenta baixa atividade em monocotiledôneas (Jang et al., 2002; Tadesse et al., 2003), o que poderia reduzir a eficácia da nossa estratégia. Por este motivo, o promotor CaMV35S foi substituido pelo promotor *Ubi1* do gene que codifica a proteína ubiquitina em milho. O promotor *Ubi1* é um promotor forte e capaz de expressar constitutivamente genes em monocotiledôneas, de forma que proteínas recombinantes possam ser produzidas em altos níveis (Cornejo et al., 1993).

**Construção do vetor de expressão para transformação de cana-de-açúcar**

Os vetores pACN e binSRNACatN (Syngenta©, EUA), contendo parte do sistema *alc* e o sistema *alc* completo, respectivamente, foram a base para a construção de um novo vetor de expressão para utilização em monocotiledôneas, como a cana-de-açúcar. Três vetores preliminares foram desenvolvidos para que construção final contendo os cassetes UBI:alcR:NOS e alcA:GUS:NOs pudesse ser realizada.

*Vetor pACNGUS*

A partir do vetor pACN, o gene que codifica a proteína cloranfenicol acetil transferase (CAT) foi substituído pelo gene da β-glucuronidase. Desse modo, o promotor do gene *alcA*, quando ativo, passa a ser o reponsável pela expressão do gene repórter *gus* (Figura 4A).

*Vetor pUBILN*

O vetor pUBILN foi obtido pela modificação do vetor pAHC17 (Christensen e Quail, 1996) que contém o promotor ubiquitina de milho e o terminador NOS da nopalina sintase. O vetor pAHC17 possui apenas um sítio de restrição (*Bam*HI) capaz de favorecer a inserção do gene de interesse entre o promotor e a região terminadora, tornando o mecanismo de clonagem bastante limitado (Figura 3B). Para inserção de outros sítios de restrição nesse vetor dois oligonucleotídeos foram desenhados (PL-S e PL-AS), de modo que um fosse complementar ao outro, formando uma fita dupla de DNA. Esses óligos possuem as seqüências de reconhecimento das enzimas *Bam*HI, *Eco*RV, *Sac*I, *Mlu*I e *Kpn*I, que funcionou como um sítio múltiplo de clonagem. Desta maneira, o vetor pAHC17 foi primeiramente digerido com a enzima *Bam*HI e em seguida o vetor foi ligado ao fragmento de DNA contendo os sítios de restrição. Após a ligação, o sítio de *Bam*HI foi reconstituído em apenas uma das extremidades, uma vez que a seqüência após o sítio *Kpn*I é capaz de parear com a extremidade protuberante do plasmídeo linearizado, mas não reconstitui o sítio *Bam*HI. Desta forma, o sítio *Bam*HI segue como sítio único. O vetor resultante foi chamado então de pAHC17/L1.

Para facilitar a retirada do cassete de expressão, visto que em alguns casos é necessário a transferência do cassete para outro vetor, foi feita uma nova modificação logo ao final do terminador NOS, com a remoção do sítio da enzima de restrição *Eco*RI e a adição dos sítios para *Sph*I e *Hind*III. Sítios de restrição para essas duas enzimas também estavam presentes antes do promotor *Ubi1*, permitindo assim, que a digestão com uma dessas duas enzimas fosse capaz de liberar o cassete de expressão do vetor pUBILN (Figura 4B).

*Vetor pUBILNalcR*

O vetor pUBILNalcR foi resultado da clonagem do gene *alcR* no vetor pUBILN formando o cassete de expressão UBI:alcR:NOS (Figura 4C).

*Vetor pALCs*

O vetor pALCs foi utilizado para transformação genética de plantas de cana-de-açúcar e foi obtido pela união do vetor pACNGUS com o cassette de expressão do vetor pUBILNalcR, resultando em um vetor único contendo os dois cassetes de expressão do sistema *alc* (Figura 4C). Embora relatos afirmem não haver influência na orientação dos cassetes na eficácia do sistema, os mesmos foram clonados em orientações contrárias para evitar problemas durante a transcrição.

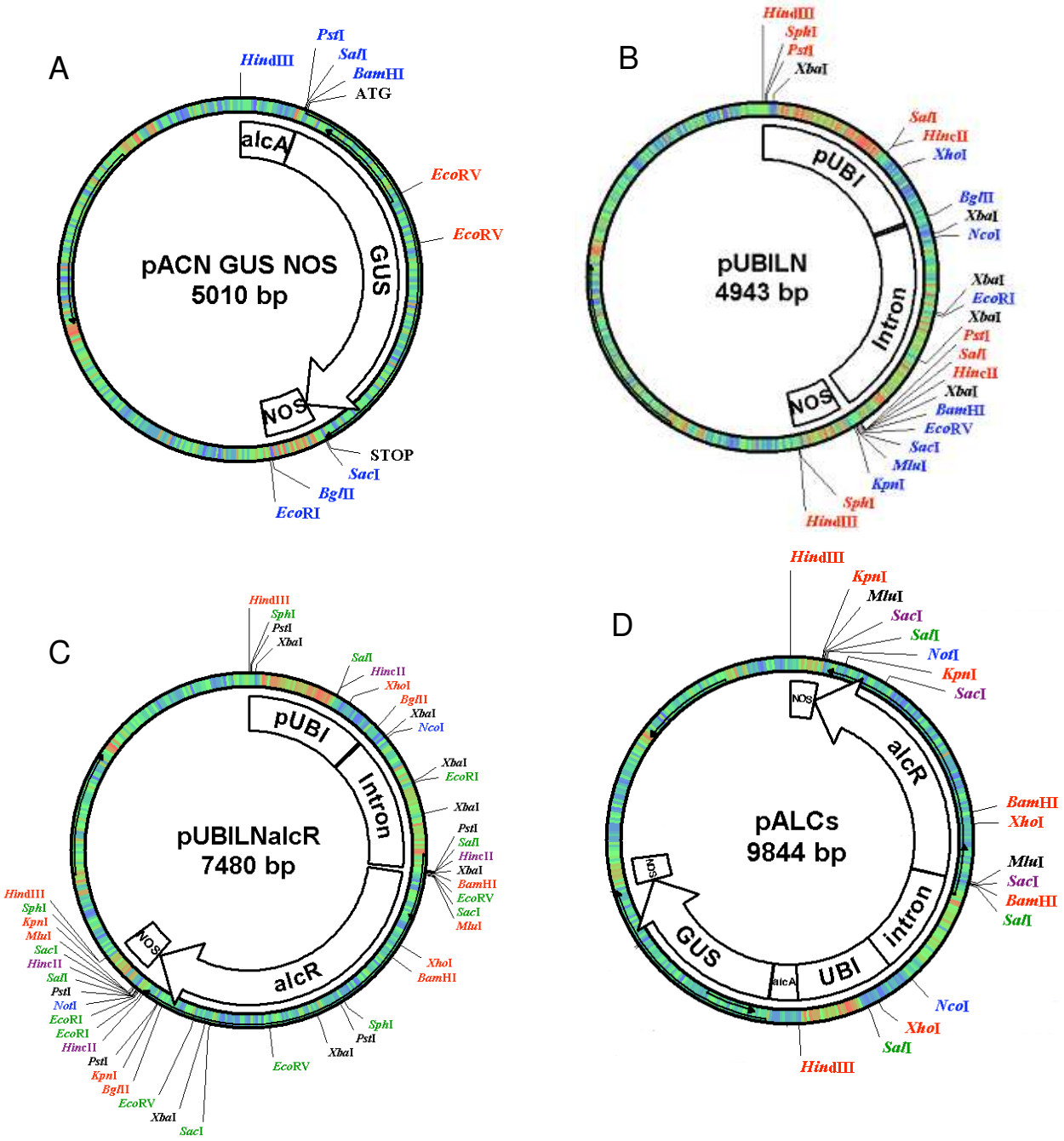



**Figura 4- Representação esquemática de vetores de expressão construídos**. a) vetor pACNGUS; b) vetor pUBILN; c) vetor pUBILNalcR e d) vetor pALCs.

***Geração de plantas transgênicas de cana-de-açúcar***

Os vetores pALCs e pACNGUS foram utilizados separadamente para a transformação genética de calos friáveis de cana-de-açúcar por meio de biolística no laboratório do Centro de Tecnologia Canavieira (CTC, Piracicaba –SP). Vinte e uma plântulas transgênicas de cana-de-açúcar bombardeadas com o vetor pALCs foram regeneradas e outras 55 plântulas com o vetor pACNGUS foram obtidas após seleção em meio de cultura com agente seletivo.

***Avaliação das plantas transgênicas***

Uma caracterização preliminar foi realizada com uma parcela das plântulas regeneradas para identificar os eventos positivos. A partir de DNA genômico extraído das folhas foram realizadas análises PCR utilizando várias combinações de iniciadores (descritos em Materias e Métodos).

Inicialmente 15 plântulas pALCs e 10 plântulas pACNGUS foram analisadas por PCR. Para a reações de PCR que utilizaram o par de óligonucleotídeos 1, esperava-se um produto de aproximadamente 500 pb nas plantas pALCs e pACNGUS. Os eventos testados apresentaram um fragmento de tamanho esperado, porém de intensidade muito fraca (Figura 5). Um fragmento de tamanho menor (~450pb) também foi amplificado e com melhor intensidade. Além disso, houve amplificação em amostra de DNA de planta selvagem. Assim, o produto amplificado pode ser resultado de anelamento inespecífico dos iniciadores utilizados na reação de PCR com regiões do genoma da cana-de-açúcar. Algumas modificações no protocolo de reação, temperatura de anelamento e programação dos ciclos foram realizadas, mas os resultados não apresentaram alterações significativas que apontassem para a confirmação da introgressão do cassete gênico. Resultados semelhantes foram obtidos com os demais pares de iniciadores testados. Os resultados obtidos com a técnica de PCR não foram satisfatórios e conclusivos, uma vez que não foi possível determinar com exatidão se plantas transformadas eram eventos positivos ou não para o sistema *alc* ou se limitações da técnica de PCR para o sistema e organismo em questão impediram a identificação dos eventos positivos. Novos testes utilizando as técnicas de Southern blot e Northern blot serão realizados futuramente dando continuidade ao trabalho.

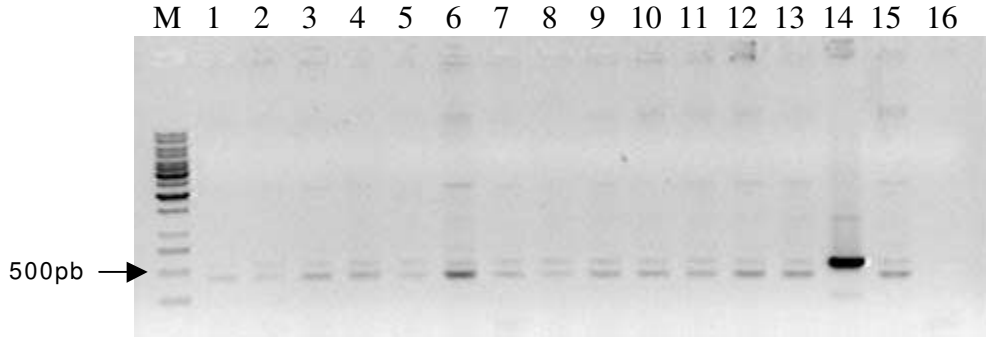



**Figura 5-** Eletroforese em gel de agarose 1% do produto de PCR a partir de DNA de plantas potencialmente transformadas de cana-de-açúcar utilizando o par de óligos 1. M – marcador de peso molecular de 1Kb (Fermentas); 1) pAlCs-1; 2) pALCs-2; 3) pALCs-3; 4) pALCs-4; 5) pALCs-5; 6) pALCs-6; 7) pALCs-7; 8) pALCs-8; 9) pALCs-9; 10) pACNGUS-1; 11) pACNGUS-2; 12) pACNGUS-3; 13) pACNGUS-4; 14) controle positivo para a reação utilizando o vetor pALCs; 15) controle negativo - planta selvagem; e 16) controle negativo da reação - água. A seta indica o tamanho aproximado de 500 pb esperado para a banda ampificada.

***Ensaios com indutor químico e detecção da atividade de GUS***

Devido aos resultados preliminares obtidos por PCR não serem conclusivos quanto a transgenia das plântulas de cana-de-açúcar regeneradas, resolvemos então verificar *in situ* a atividade da proteína GUS. Vinte e uma plântulas obtidas para o vetor pALCs e outras 10 plântulas regeneradas com o vetor pACNGUS foram tratadas com etanol 2,5% (v/v). Passados 5 horas após a pulverização do etanol, a atividade da proteína GUS foi avaliada através reação histoquímica com substrato X-Gluc. Como esperado, plantas selvagens de cana-de-açúcar não apresentaram atividade de GUS. Plantas potencialmente transgênicas para o vetor pACNGUS também não apresentaram atividade detectável para a proteína GUS. Esse resultado foi esperado, uma vez que o promotor do gene *alcA* não é ativado apenas pela presença de etanol e sim através do fator de transcrição *AlcR* ativo. Uma vez que o sistema *alc* está completo no vetor pALCs, o fator de transcrição *AlcR* deve ser constantemente expresso na planta devido ao promotor constitutivo Ubi-1. A proteína ALCR disponível na célula e a presença de etanol exógeno deveriam proporcionar a atividade do

promotor do gene *alcA* e consequentemente a transcrição do gene *B-glucuronidase*. Os resultados obtidos com as plantas transformadas com o vetor pALCs foram contrários ao que era esperado. Não foi detectada atividade de GUS em nenhuma das 21 plantas analisadas. Como controle positivo desse experimento foram utilizadas folhas de petúnia transgênica contendo o gene GUS (Ulian et al., 1993; material cedido pelo CTC), as quais apresentam forte atividade desta enzima (resultados não mostrados).

Um segundo experimento de indução por etanol foi realizado com as mesmas plântulas de cana-de-açúcar. Desta vez, aumentou-se a dose de etanol aplicado de 2,5 para 4% (v/v) e as folhas foram coletadas 24 horas após a aplicação. Como o ensaio histoquímico é um método qualitativo e os resultados obtidos no experimento anterior não foram satisfatórios, neste segundo experimento foi realizado um ensaio fluorímetrico *in vitro* que é mais sensível do que o método histoquímico além de permitir a quantificação da atividade da proteina GUS por meio da fluorescência emitida pelo produto da reação com o substrato MUG (4-metilumbeliferil ß-D-glucuronídeo), a metilumbeliferona (Jefferson et al., 1987). Em concordância com os resultados obtidos nos testes histoquímios, os extratos proteicos das plântulas transformadas com pALCs e pACNGUS, assim como de plantas selvagens de cana-de-açúcar (controle negativo), todas não tratadas com etanol, não apresentaram atividade fluorimétrica produzida pela proteína GUS. Plântulas transformadas com pACNGUS e com pALCs tratadas com etanol também não apresentaram fluorescência. A figura 6 ilustra os resultados obtidos pelo ensaio fluorimétrico com substrato MUG.

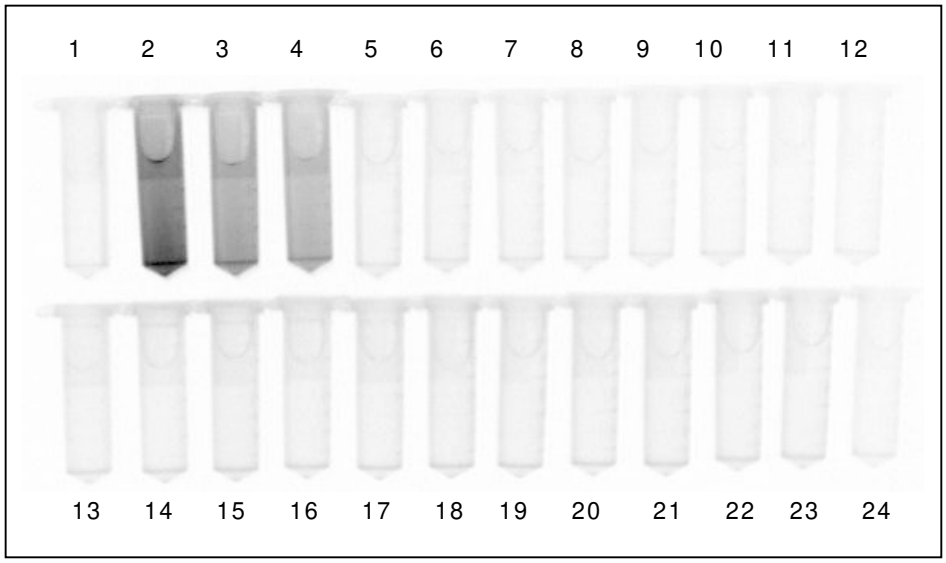



**Figura 6-** Ensaio fluorimétrico de fluorescência da metilumbeliferona sob luz ultravioleta em plantas potencialmente transformadas de cana-de-açúcar. 1) controle negativo; 2) *E. Coli* expressando GUS – controle positivo I; 3) *Arabidopsis* transgênica expressando GUS – controle positivo II; 4) petúnia transgênica expressando GUS – controle positivo III; 5) planta pALCs 1 sem etanol; 6) planta pALCs 1 com etanol; 7) planta pALCs 2 sem etanol; 8) planta pALCs 2 com etanol; 9) planta pALCs 3 sem etanol; 10) planta pALCs 3 com etanol; 11) planta pALCs 4 sem etanol; 12) planta pALCs 4 com etanol; 13) planta pALCs 5 sem etanol; 14) planta pALCs 5 com etanol; 15) planta pALCs 6 sem etanol; 16) planta pALCs 6 com etanol; 17) planta pACNGUS 1 sem etanol; 18) planta pACNGUS 1 com etanol; 19) planta pACNGUS 2 sem etanol; 20) planta pACNGUS 2 com etanol; 21) planta pACNGUS 3 sem etanol; 22) planta pACNGUS 3 com etanol; 23) planta selvagem sem etanol; e 24) planta selvagem com etanol. A coloração escura nos tubos 2, 3 e 4 indicam atividade da enzima GUS.

**Conclusões**

Até o momento, os resultados obtidos não foram suficientes para comprovar a transformação estável das plantas regeneradas com os cassetes de expressão e a funcionalidade do sistema *alc* em cana-de-açúcar. A cana-de-açúcar por ser um organismo de genoma bastante complexo, poliplóide, pode interferir ou mesmo impedir a expressão ou tradução do gene repórter. Em adição aos testes de PCR, histoquímico e fluorimétrico, novos testes com as técnicas de *Southern blot* e *Northern blot* serão realizados para analisar as plântulas de cana-de-açúcar transformadas com os vetores pALCs e pACNGUS.

**Referências Bibliográficas**


Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry** 112: 195-203

Christensen AH, Quail PH (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. **Transgenic Research** 5:213-218

Deveaux Y, Peaucelle A, Roberts GR, Coen E, Simon R, Mizukami Y, Traas J, Murray JAH, Doonam JH, Laufs P (2003) The ethanol switch: a toll for tissue-specific gene induction during plant development. **The Plant Journal** 36:918-930

Doyle JJ & Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin** 19: 11-15

Felenbok B, Sequeval D, Mathieu M, Sibley S, Gwynne DI, Davies RW (1988) The ethanol regulon of *Aspergillus nidulans:* characterisation and sequence of the positive regulatory gene alcR. **Gene** 73:386-396

Fillinger S, Panozzo C, Mathieu M, Felenbok B (1995) The basal level of transcription of the alc genes in the ethanol regulon in Aspergillus nidulans is controlled both by the specifif transactivator AlcR and the general carbon catabolite repressor CreA. **FEBS Letters** 368:547-550

Garoosi GA, Salter MG, Caddick MX, Tomsett AB (2005) Characterization of the ethanol-inducible alc gene expression system in tomato. **Journal of Experimental Botany** 56 (416): 1635-1642

Gatz C & Lenk I (1998) Promotors that respond to chemical inducers. **Trends in Plant Science** 3:352-358

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. **EMBO Journal** 6: 3901-3907

Jepson I, Martinez A, Sweetman JP (1998) Chemical inducible gene expression system. Plant **Molecular Biology Report** 56:360-367

Li R, Jia X, Mao X (2005) Ethanol-inducible gene expression system and its applications in plant functional genomics. **Plant Science** 169:463-469

Maizel A, Weigel D (2004) Temporally and spacially controlled induction of gene expression in *Arabidopsis thaliana*. **The Plant Journal**. 38:164-171

McCabe DE, Swain WF, Martinell BJ, Christou P (1988) Stable transformation of soybean (Glycine max) by particle bombardment. **Biotechnology** 6: 923-926

Pateman JH, Doy CH, Norris U, Creaser EH, Hynes M (1983) Regulation of alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (AldDH) in *A nidulans.* **Proccedings of the Royal Society of London** 217:243-264

Rech EL, Aragão FJL (1998) Biolística. In: **Manual de Transformação Genética de Plantas**. Brasileiro, ACM e Carneiro, VTC (ed.). Brasília: Embrapa-SPI/Embrapa-Cenargem. P 51-64

Roslan HA, Salter MG, Wood CD, White MRH, Croft KP, Robson F, Coupland G, Doonan J, Laufs P, Tomsett AB, Caddick MX (2001) Characterisation of the ethanol-inducible alc gene expression system in *Arabidopsis thaliana.* **The Plant Journal** 28:225-235

Saciloto RFZ (2003) **Inserção do gene *PR5K* em cana-de-açúcar visando induzir resitência ao fungo da ferrugem *Puccinia melanocephala***. Dissertação de Mestrado. Escola Superior de Agricultura Luiz de Queiroz, Piracicaba – SP, 74p.

Salter MG, Paine JA, Riddle KV, Jepson I, Greenland AJ, Caddick MX, Tomsett AB (1998) Characterisation of ethanol-inducible alc gene expression system for transgenic plants. **The Plant Journal**. 16:127-132

Sambrook J, Fritsch EF, Maniatis T (1989) **Molecular Cloning: a laboratory manual**. Cold Spring Harbor Laboratory Press, NY, vol 1, 2, 3

Schlögl PS, Kobarg J, Moreau VH, Leite A, Sabino AA, Eberlin MN, Menossi M (2004) Expression, purification and characterization of a novel bZIP protein from sugarcane. **Plant Science**. 167:583-595

Sealy-Lewis HM & Lockington RA (1984) Regulation of two alcohol dehydrogenases in *A nidulans.* **Current Genetics**. 8:253-259

Sweetman JP, Chu C, Qu N, Greenland AJ, Sonnewald U, Jepson I (2002) Ethanol Vapor is an efficient inducer of the alc gene expression system in model and crop plant species. **Plant Physiology**. 129:943-948

Topfer R, Matzeit V, Gronenborn B, Schell J, Steinbiss HH (1987) A set of plant expression vectors for transcriptional and translational fusions. **Nucleic Acids Research** 15:5890

Ulian EC, Magill JM, Smith RH (1993) Expression and inheritance pattern of two foreign genes in petunia. **Theoretical and Applied Genetics** 88:433-440

Zuo J & Chua NH (2000) Chemical-inducible systems for regulated expression of plant genes. **Current Opinions in Biotechnology** 11:146-151

# Some statistical properties of gene expression clustering for array data

Propriedades estatísticas do agrupamento de genes pela expressão, em dados de arranjos de DNA

Gabriel Coelho Gonçalves de Abreu, Aluísio de Souza Pinheiro, Rodrigo Duarte Drummond, Sandra Rodrigues de Camargo, Marcelo Menossi

# Some Statistical Properties of Gene Expression Clustering for Array Data

Gabriel Coelho Gonçalves de Abreu[1], Aluísio de Souza Pinheiro*[1] , Rodrigo Duarte Drummond[2,3], Sandra Rodrigues Camargo[2] and Marcelo Menossi[2,3]

[1]Department of Statistics - IMECC - UNICAMP - Campinas - SP - Brazil
[2]Laboratory of Functional Genomics - CBMEG - UNICAMP - Campinas - SP - Brazil
[3]Department of Genetics and Evolution - IB - UNICAMP - Campinas - SP - Brazil

Email: Aluísio de Souza Pinheiro*- pinheiro@ime.unicamp.br;

*Corresponding author

## Abstract

**Background:** DNA arrays have been a rich source of data for the study of genomic expression for a wide variety of biological systems. Gene clustering is of the paradigmas quite used to assess the significance of a gene (or group of genes). However, due to the massive input of data concerns should be raised about the apparent statistical significance of some genes. Moreover, most of the gene clustering techniques are applied to cDNA array data without a statistical error measure.

**Results:** We propose an easy-to-implement and simple-to-use technique (via bootstrap resampling) to evaluate the statistical error for the nodes provided by SOM-based clustering. Comparisons between SOM and model-based clustering are presented for simulated as well as for two real data sets. We also implement a bootstrap-based pre-processing procedure for SOM, that improves the false discovery ratio of differentially expressed genes.

**Conclusions:** We propose a simple solution of providing SOM with statistical meaning, via bootstrap resampling. Even for highly structured situations in which model-based procedures should outperform model-free alternatives, SOM will be a very powerful source of exploratory information for gene expression levels in cDNA array data studies.

## Background

The analysis of gene expression by cDNA arrays is a great achievement in genetic technology by which one can study thousands of genes in a single set of experiments at a reasonable cost. However, due to the massive quantity of data and subtle nature of gene expression, the analysis of cDNA arrays data is a challenge to statistical techniques. Besides grouping, since the object of such analysis is the selection of a small number of genes, the automatic grouping of such promising genes would be desirable. Hierarquical clustering [1] and Self Organizing Maps [2] are two of the most popular choices. Since clustering techniques are performed on a sample of gene expressions, groups are subject to an statistical error (see Datta & Datta, 2003, and references therein), whose estimate for clustering methods is usually not an easy task. Kerr & Churchill (2001) provides a group clustering procedure, equiped with error measures, in which the ANOVA model developed in Kerr et al. (2000) is used for grouping genes. Yeung et al. (2001) uses a leave-one-out procedure to select the number of clusters.

We propose a simple error measure estimate for SOM via bootstrap which should help one to understand whether the resulting nodes are statistically significant or simply due to the pre-defined number of nodes of the algorithm. We also study the effects of pre-processing in SOM clustering techniques, filtering genes which are clearly non-induced and will be only a computational and statistical burden to the SOM procedure.

We study the performance of SOM and the model-based method proposed by Kerr & Churchill (2001) in a simulation study and two real data sets. The model-based method has an overall better performance, with much less genes selected, higher sensitivity and higher specificity. The use of bootstrap confidence regions for SOM groups enables a clustering that is statistically meaningful, while the pre-processing improves its sensitivity and specificity.

Section 2 presents the model, the data sets and the simulation procedure. Results are presented in Section 3, where the behavior of each clustering method is discussed. Some final comments are made in Section 4.

## Results and Discussion

Table 1 shows the number of selected genes in DATA1 (Chu et al., 1998) for analysis performed by SOM and the model-based clustering algorithm by Kerr & Churchill (2001). The profiles from Chu et al, 1998 are used as basis for classification. A total of eight clustering procedures were employed. SOM analysis was performed, with or without pre-processing, and three options of nodes: 6, 16 or 30. The model-based method, proposed by Kerr et al. (2001), is employed with pre-processing, and either with a 95% stability

procedure or not. As a general trend, SOM'S results will improve when the number of nodes increase and/or pre-processing is performed, the latter having a slightly stronger effect than the former. Anyway, the model-based method has a much better performance (in selecting less genes) than any SOM's configuration. Notice for instance that when pre-processing and stability are employed on the model-based setup 229 genes are selected while the best SOM's performance happens for 30 nodes and pre-processing, for which 2883 genes are profiled as induced. Performing the SOM's method with more than 7 nodes (number of profiles considered) allows some of those to be associated with non-selected genes, which results in a smaller final number of selected genes.

TABLE 1
TABLE 2

Table 2 summarizes the results for the simulation study. One notices that for low error models (1,2 and 3), the model-based method has a quite good performance, being able to assign all induced genes to its correct profile. Moreover, the model-based selects relatively few non induced genes and has a ratio of approximately 5 false positives to one true positive, which is very important for the more expensive (time consuming and moneywise) analysis to be carried out with every selected gene (e.g. Northern blotting or real-time PCR). The specificity of the procedure decreases for the higher error models (5,6 and 7) while its sensitivity increases. That means that the ratio between false positives and true positives stays reasonable (between 1.88 and 5.43) even in more critical situations. The latter specificity performance should not be considered a problem since, under the error conditions of variances 40, 80 or even 800 times larger than the experiment variance, the very nature of expression profiles should be compromised.

SOM shows an overall specificity and sensitivity homogeneous performance along the error structures. This however is not a good property since the false positives/true positives ratios will vary between 51.83 and 635.67. Among the SOM procedures the 30 nodes with pre-processing procedure shows the best performance, overall and locally.

Table 3 shows the results for DATA2 (Camargo et al., unpublished manuscript). P-values were computed both as if residuals distribution would be normal (only as a rule of thumb) and, more realistically (see Kerr et al., 2002 and Drummond et al., 2005, for evidences of non-normality of array residuals), via bootstrap resampling. Both sets of p-values resulted in all effects being statistically highly significant. This indicates that all model effects are numerically relevant for the evaluation of the residuals, which will be later used in the bootstrap resamples of the clustering algorithm.

For DATA2, three genes were selected by the model-based method: one for profile $(I_1, R_2, R_2, R_2)$, and two

for profile $(I_1, R_2, M_3, M_4)$. The genes were selected with 95% bootstrap stability, in the sense of Kerr & Churchill (2001), for all four levels 90, 95, 99 and 99.9, with one exception, where a stability of more than 94% but less than 95% was obtained for $(I_1, R_2, M_3, M_4)$ and level 95%.

Analysis by SOM was performed with either 6, 15 or 30 nodes, which results are shown in Table 4. The method selects an average of 15 (9.8 or 4.9) genes to each profile if 6 (16 or 30) nodes, pre-processing and bootstrap confidence region are used. Results worsen without confidence regions; an average of 31.5 (11.8 or 8.2) if 6 (16 or 30) nodes are employed. If neither pre-processing nor bootstrap confidence regions are employed, the average number of selected genes per profile goes to 595.8, 223.4 and 119.1 for 6, 16 and 30 nodes, respectively.

TABLE 3

TABLE 4

FIGURES 1 and 2

Figures 1a, 1b, 2a and 2b show the behavior of the first two principal components of SOM analysis on DATA2 (with 99% bootstrap confidence regions) for 6 nodes with pre-processing, 6 nodes without pre-processing, 16 nodes with pre-processing and 30 nodes with pre-processing, respectively.

One should notice how hard it is to distinguish the behavior of genes (on their first two principal components). One can see, from comparisons between Figures 1a and 1b, that nodes are slightly dislocated when pre-processing is applied and that the confidence regions are larger. This enlargment is due to the smaller effective sample size. However, in either case, five non-overlapping nodes are presented. For 30 nodes (Figure 2b), some stronger differences of nodes location appear, showing that some caution must be applied when choosing the number of nodes. It becomes very hard to distinguish statistically meaningful nodes because most will have some bootstrap overlapping with some of the others. The nodes themselves will be quite different from the ones selected by either 6 or 16 nodes, as illustrated respectively by Figures 1a and 2a.

## Conclusions

DNA arrays can be a powerful tool to understand how genes respond to specific stimulae. However, due to its high dimensionality/low reprodutibility , one must be cautious with its patterns. Computational results can be easily obtained but one should restrain from concluding their statistical significance and generalization. We propose a simple solution of providing SOM with statistical meaning, via bootstrap resampling. Whenever SOM is applied to a data set and bootstrap regions of confidence are built the

analyst is able to clearly separate statistically meaningful nodes from spurious (purely numerical) ones. We think is important to point out that, guven the variability in microarray data, it is not a very googd idea to use SOM with a large number of nodes.

Another issue that we address is the model-based/free approach to clustering. If no prior information is available model-based techniques can be quite inefficient. On the other hand whenever some set of prior profiles can be pursued, model-based techniques will perform much more adequately than purely mode-free techniques, as illustrated in two data sets and by the simulation study. We have seen that prior knowledge used in the model building makes the model-based procedure much more efficient than the model-free alternative. It is also important to have in mind that some supervision can almost always provide superior performance to totally unsupervised procedures.

Even for highly structured situations in which model-based procedures should outperform model-free alternatives, SOM will be a very powerful source of exploratory information for gene expression levels in cDNA array data studies. Its performance and usefulness can be statistically enhanced if a measure of quality for each node (such as the proposed bootstrap confidence region) is used and if several pre-defined number of nodes are employed, preventing distortions due to numerical but not statistically meaningful convergences. This series of runs is not a burden due to the easeness of implementation and computation speed; therefore, we recommend that some runs should be performed, each with a different number of nodes.

## Methods

Two clustering techniques were employed, the SOM and the model-based method by Kerr & Churchill (2001). SOM is implemented in Matlab following Tamayo et al. (1999), Kohonen (2001) and Nikkilä et al. (2002).

Comparisons between SOM and the methods proposed by Kerr & Churchill (2001) were conducted by the analysis of two data sets and a simulation study. Two data sets, DATA1 (Chu et al., 1998), and DATA2 (Camargo et al., unpublished manuscript), are used to exemplify the bootstrap error measure for SOM, as well as to compare its performance to Kerr's model-based method. DATA1 deals with the gene expression during meiosis and sporulation for *Saccaromyces cerevisae*. This data is the result of the study of gene expression during meiosis and spore formation. Analysis were performed in seven time points: 0, 0.5, 2, 5, 7, 9 and 12 hours, each one in a different microarray. Time zero was used as control. DATA2 studies gene expression from *Saccharum officinarum* on the presence of ethanol. Three varieties of sugar-cane were

used, each one studyed in four time points: 0, 1, 4 and 12 hours, being the time point 0 considered as control. Varietie one had three replications in each time, varieties two and three had only on replication in each time.

In SOM's implementation, a bootstrap confidence region was constructed, to infer the statistical significance of each node, i.e, its separation from the other nodes. 500 bootstrap resamples are taken using the same initial values for the nodes. Each resample is taken from the original data and tha SOM's procedure is re-evaluate using the same inicial nodes. The resulting bootstrap nodes are then assigned by distance to the original data nodes. The respective distances from each bootstrap node to its assigned data node are computed and their quantiles are used to construct confidence regions in SOM's principal components.

SOM is a fast and relatively easy method to implement, thence it is a powerful tool for exploratory data analysis. Therefore, it is usually applied to the whole data set and groups are then selected based on the resulting clusters (see Tamayo et al. (1999) for details). We tested whether pre-processing in which clearly non-induced genes are eliminated from the analysis would enable SOM to classify the remaining genes in a more effficient way. The pre-processing (for all model-based and model-free procedures) eliminated genes whose expressions (VG for DATA1 and PVG for DATA2) were statistically equal to the respective control expression for all time points under consideration.

Since the (model-based) method by Kerr works under some pre-selected profiles, it has a simple way of assigning genes (that are considered expressed) to clusters, because profiles are decided prior to testing. For SOM, however, allocation is not as direct. For DATA1 and simulated data, for which one has *known* profiles tags, we choose as the node associated to a particular profile the node to which the majority of the genes from this profile are assigned.

The model for DATA1 is given as in Kerr et al. (2001):

$$ln(y_{ijkg}) = \mu + A_i + D_j + (AD)_{ij} + G_g + (AG)_{ig} + (VG)_{kg} + \varepsilon_{ijkg},$$

where $\mu$ is the overall mean effect, $A_i$ is the effect of the $i$-th array, $D_j$ is th e effect of the $j$-th dye, $(AD)_{ij}$ is the effect of the interaction between the $i$-th array and the $j$-th dye, $G_g$ is the effect of the $g$-th gene, $(AG)_{ig}$ is the effect of the interaction between the $i$-th array and the $g$-th gene, $(VG)_{kg}$ is the effect of the interaction between the $k$-th treatment (each time point) and the $g$-th gene, and $\varepsilon_{ijkg}$ is the error term.

The pre-processing used in both methods was described by Kerr et al. (2001). It is based on 99% bootstrap confidence intervals for the differences in gene expression ($g$) at a treatment ($k$) and control (0):

$(VG)_{kg} - (VG)_{0g}$. All genes that for at least one treatment $k$ (not control) $(\hat{V}G)_{kg} - (\hat{V}G)_{0g} > 0$ and the 99% confidence interval does not contain zero, are selected. For those pre-selected genes the correlation coeficient $r_{gp}$, $p = 1, 2, ..., 7$, between gene $g$ and profile $p$ is calculated. Gene $g$ is assigned to profile $p$ if $r_{gp} > 0.9$ and $r_{gp}$ is the largest of $\{r_{g1}, ..., r_{g7}\}$. The authos define genes "95% stability" as those which are assigned to the same profile in the analysis of original data and at least 95% of bootstrap repeats of the procedure (see Kerr & Churchil (2001) for more details).

DATA1 served as a basis for numerical simulations as follows. For each simulated data set, there were 6118 genes, 6048 of which would be not simulated as a potentially diffrentially expressed gene. From the 6118 simulated genes, 10 would be sampled from each of the seven profiles presented by Chu et al. (1998), with a total of 70 potentially expressed genes per sample. Seven error structures were considered: (1) resampling from the original residuals in Chu's data (variance equals .0146); (2) normal distribution with variance .0146; (3) normal distribution with variance .0300; (4) normal distribution with variance .2500; (5) normal distribution with variance .5000; (6) normal distribution with variance 1.0000; and (7) normal distribution with variance 10.0000; 200 replications were made for each setup.

The second data set was analyzed accordingly to the model:

$$\begin{aligned} ln(y_{ijkg}) &= \mu + P_j + V_k + G_g + (PV)_{jk} + (PG)_{jg} \\ &\quad +(VG)_{kg} + (PVG)_{jkg} + \epsilon_{ijkg}, \end{aligned}$$

where $y_{jkg}$ is the expressed value for the $j$-th plant variety, $k$-th treatment and $g$-th gene, $\mu$ is the expression general mean , $P_j$ the effect due to the $j$-th plant variety, $V_k$ is the effect due to the $k$-th treatment, $G_g$ is the effect due to the $g$-th gene, $(PV)_{jk}$ is the effect due to the interaction between the $j$-th variety and the $k$-th treatment, $(PG)_{jg}$ is the effect due to the interaction of the $j$-th variety and the $g$-th gene, $(VG)_{kg}$ is the effect due to the interaction of the $k$-th treatment and the $g$-th gene, $(PVG)_{jkg}$ is the effect due to the interaction of the $j$-th variety, the $k$-th treatment and the $g$-th gene, and $\varepsilon_{ijkg}$ is the error term for the $i$-th observation of the $(j, k, g)$-th combination of factor levels.

The focus of interest is the combined effect of variety, treatment and gene, i.e. $(PVG)_{jkg}$; $j = 1, 2, 3$; $k = 1, 2, 3, 4$ (where $k = 1$ is the control effect); $g = 1, 2, \ldots, 3575$. We create profiles as follows:

1. Gene expressions were ordered by their values in each time individually.

2. Levels of expression were defined as 90, 95, 99 and 99.9, where the numbers represent the percentage of genic expression considered median.

3. Gene expressions in each time ($k = 1, 2, 3, 4$) were classified in three classes - Induced (I), Repressed (R) or Median (M) - if their expression level was respectively beyond the 95th percentile (97.5th, 99.5th or 99.95th), below the 5th percentile (2.5th, .5th or .05th), or in between those bounds.

4. Profiles were built as $\{(N_1, N_2, N_3, N_4);\ N_k \in \{I_k; M_k; R_k\};\ k = 1, 2, 3, 4\}$.

5. Kerr's procedure was then employed, genes were tentatively classified in one of the $4 \times 3^4 = 324$ classes, and 95% stable genes were selected.

## Authors contributions

GCGC developted all computer programs and analyzed the results. ASP proposed the problem, analyzed the results and drafted the manuscript. SRC performed the biological experiments and array hybridizations. RDD reviewed the literature that motivated this study, assembled and organized the data and discussed the results. MM performed the biological experiments and discussed the results. All authors read and approved the final manuscript.

## Acknowledgements

This work has been partially supported by FAPESP. The authors would like to acknowledge the National Center for High Performance Computing (CENAPAD/UNICAMP) for the use of its resources.

## References

1. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster Analysis and Display of Genome-Wide Expression Patterns**. *Proc Natl Acad Sci USA* 1998, **95**:14863–14868.
2. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting Patterns of Gene Expression with Self-Organizing Maps: Methods and Application to Hematopoietic Differentiation**. *Proc Natl Acad Sci USA* 1999, **96**:2907–2912.
3. Cheng Y: **Convergence and Ordering of Kohonen's Batch Map**. *Neural Computation* 1997, **9**:1667–1676.
4. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I: **The Transcriptional Program on Sporulation in Budding Yeast**. *Science* 1998, **282**:699–705.
5. Datta S, Datta S: **Comparisons and validation of statistical clustering techniques for microarray gene expression data**. *Bioinformatics* 2003, **19(4)**:6459–466.
6. Davison AC, Hinkley DV: **Bootstrap Methods and Their Application** 1997.
7. Drummond RD, Pinheiro A, Rocha CS, Menossi M: **ISER: Selected of differentially expressed genes from DNA array data by nonlinear data transformations and local fitting**. in press.
8. Efron B, Tibshirani R: *An Introduction to the Bootstrap*. New York: Chapman & Hall 1993.
9. Kerr MK, Martin M, Churchill GA: **Analysis of Variance for Gene Expression Microarray Data**. *J Comput Biol* 2000, **7**:819–837.
10. Kerr MK, Churchill GA: **Bootstrapping Cluster Analysis: Assessing the Reliability of Conclusions from Microarray Experiments**. *Proc Natl Acad Sci USA* 2001, **98**:8961–8965.

11. Kerr MK, Afshari CA, Bennett L, Bushel P, Martinez J, Walker NJ, Churchill GA: **Statistical analysis of a gene expression microarray experiment with replication**. *Statistica Sinica* 2002, **12**:203–217.

12. Kohonen T: *Self-Organizing Maps*. Berlin, Germany: Springer, 3rd edition 2001.

13. Nikkilä J, Törönen P, Kaski S, Venna J, Castrén E, Wong G: **Analysis and visualization of gene expression data using Self-Organizing Maps**. *Neural Networks* 2002, **15**:953–966.

14. Nogueira FT, De Rosa VEJ, Menossi M, Ulian EC, Arruda P: **Rna Expression Profiles and Data Mining of Sugarcane Response to Low Temperature**. *Plant Physiol* 2003, **132**:1811–1824.

15. Resson H, Wang D, Natarajan P: **Adaptive double self-organizing maps for clustering gene expression profiles**. *Neural Networks* 2003, **16**:633–640.

16. Törönen P, Kolehmainen M, Wong G, Castrén E: **Analysis of gene expression data using self-organizing maps**. *Federation of European Biochemical Societies* 1999, **451**:142–146.

17. Yeung KY, Haynor DR, Ruzzo WL: **Validating clustering for gene expression data**. *Bioinformatics* 2001, **17(4)**:309.

## Figures

### Figure 1 - 99% Confidence Region for SOM's first two PC's with 6 Nodes - DATA2

**(a)** Pre-processed **(b)** Whithout Pre-processing

### Figure 2 - 99% Confidence Region for SOM's first two PC's - DATA2

**(a)** with 16 Nodes Pre-processed **(b)** with 30 Nodes - Pre-processed

## Tables

### Table 1 - Number of Selected Genes for data from [4] - SOM and Model-based method proposed by Kerr et al. (2001)

NPP - SOM without pre-processing

PP - Pre-processed SOM

NO - Without 95% Bootstrap stability for Model-based

YES - 95% Bootstrap stability for Model-based

| Profile | SOM | | | | | | Model-based | |
|---|---|---|---|---|---|---|---|---|
| | 6 nodes | | 16 nodes | | 30 nodes | | | |
| | NPP | PP | NPP | PP | NPP | PP | NO | YES |
| 1 | 789 | 977 | 1,678 | 1,690 | 1,148 | 878 | 82 | 12 |
| 2 | 2,194 | 692 | 1,730 | 571 | 982 | 371 | 52 | 11 |
| 3 | 777 | 288 | 755 | 656 | 367 | 449 | 88 | 15 |
| 4 | 2,044 | 923 | 735 | 0 | 594 | 597 | 163 | 28 |
| 5 | 314 | 0 | 0 | 155 | 463 | 385 | 254 | 118 |
| 6 | 0 | 0 | 498 | 174 | 82 | 0 | 168 | 39 |
| 7 | 0 | 791 | 0 | 210 | 568 | 203 | 20 | 6 |
| Total | 6,118 | 3,681 | 5,396 | 3,456 | 4,204 | 2,883 | 827 | 229 |

### Table 2 - Simulation Results for SOM and Model-based Technique - Performance by Clustering Methods. Correctly and Incorrectly Classified Percentages. False Discovery Ratio.

(A) Percentage of Potentially Differentially Expressed Genes that are Correctly Classified as Expressed

| Case | | Error Structure | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| (A) | (1) | 100 | 100 | 100 | 57.44 | 7.80 | 0.36 | 0.00 |
| | (2) | 30.27 | 26.72 | 28.73 | 22.97 | 15.40 | 11.20 | 7.86 |
| | (3) | 29.33 | 30.61 | 30.57 | 17.48 | 17.43 | 10.41 | 6.51 |
| | (4) | 38.24 | 42.86 | 38.37 | 43.39 | 39.84 | 28.58 | 21.50 |
| | (5) | 38.09 | 43.01 | 36.84 | 44.40 | 32.84 | 27.79 | 11.19 |
| (B) | (1) | 5.78 | 6.37 | 5.73 | 1.25 | 0.31 | 0.02 | 0.002 |
| | (2) | 74.87 | 73.96 | 71.52 | 66.26 | 66.01 | 58.86 | 57.86 |
| | (3) | 47.37 | 50.85 | 45.34 | 29.51 | 32.78 | 25.81 | 21.25 |
| | (4) | 46.93 | 55.61 | 53.94 | 64.86 | 48.76 | 49.60 | 50.82 |
| | (5) | 31.67 | 33.19 | 48.61 | 26.63 | 26.02 | 25.63 | 16.35 |
| (C) | (1) | 4.99 | 5.50 | 4.95 | 1.88 | 3.43 | 5.59 | - |
| | (2) | 213.68 | 239.12 | 215.10 | 249.21 | 370.34 | 454.10 | 635.67 |
| | (3) | 133.96 | 143.55 | 128.15 | 145.89 | 162.51 | 214.24 | 281.78 |
| | (4) | 106.04 | 112.12 | 121.46 | 129.16 | 105.73 | 149.96 | 204.24 |
| | (5) | 68.96 | 66.68 | 114.01 | 51.83 | 68.46 | 79.67 | 126.24 |

(B) Percentage of Not Differentially Expressed Genes that are Wrongly Classified as Expressed

(A) False Discovery Ratio

(1) Model-based

(2) SOM6 NPP - SOM with 6 nodes, without pre-processing

(3) SOM6 PP - SOM with 6 nodes, with pre-processing

(4) SOM30 NPP - SOM with 30 nodes, without pre-processing

(5) SOM30 PP - SOM with 30 nodes, with pre-processing

**Table 3 - Analysis of Variance for Etanol Data**

SS - sum of squares; df - degrees of freedom; MS - mean squares; Ratio = MS/0.0638

* both from a F-distribution table and bootstrap quantile

| Source | SS | df | MS | Ratio | p-value* |
|---|---|---|---|---|---|
| P | 298.10 | 2 | 149.05 | 2,336.52 | 0.000 |
| V | 25.39 | 3 | 8.46 | 132.68 | 0.000 |
| G | 37,665.72 | 3,574 | 10.54 | 165.21 | 0.000 |
| PV | 58.94 | 6 | 9.82 | 153.98 | 0.000 |
| PG | 6,192.42 | 7,148 | 0.87 | 13.58 | 0.000 |
| VG | 867.58 | 10,722 | 0.08 | 1.27 | 0.000 |
| PVG | 1,560.83 | 21,444 | 0.07 | 1.14 | 0.000 |
| Error | 1,824.46 | 28,589 | 0.06 | | |
| Total | 48,493.43 | 71,499 | | | |

**Table 4 - Number of Genes Assigned by SOM to Each Node - Etanol Data**

Sample Statistics: Min - minimum; $Q_1$ - First Quartile; $Q_2$ - Median; $Q_3$ - Third Quartile;

Max - Maximum; IR - Interquartile Range; $\bar{x}$ - Mean; SD - Standard Deviation

PP (or NPP) - with Pre-processing (Without)

IB (or NIB) - with Bootstrap Confidence Region (Without)

| Number of Nodes | | Number of Genes | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Min | $Q_1$ | $Q_2$ | $Q_3$ | Max | IR | $\bar{x}$ | SD |
| 6 | NPP+NIB | 107 | 137 | 593 | 1058 | 1110 | 921 | 595.8 | 464.8 |
| | PP+NIB | 3 | 8 | 13 | 71 | 77 | 63 | 31.5 | 31.2 |
| | PP+IB | 1 | 1 | 5 | 43 | 44 | 42 | 15 | 18.7 |
| 16 | NPP+NIB | 16 | 50 | 188 | 289 | 1037 | 239 | 223.4 | 245.6 |
| | PP+NIB | 1 | 7 | 12 | 16 | 33 | 9 | 11.8 | 7.8 |
| | PP+IB | 1 | 3 | 7 | 13 | 31 | 10 | 9.8 | 7.7 |
| 30 | NPP+NIB | 4 | 13 | 63 | 124 | 819 | 111 | 119.1 | 178.4 |
| | PP+NIB | 0 | 2 | 5 | 11 | 21 | 9 | 8.2 | 6.6 |
| | PP+IB | 1 | 3 | 4 | 6 | 17 | 3 | 4.9 | 3.4 |

## Additional Files

### Additional file 1 — figure1a.eps

This is Figure 1a.

### Additional file 2 — figure1b.eps

This is figure 1b. The description under figures 1 a and b is: 99% Confidence Region for SOM's first two PC's with 6 Nodes - DATA2 **(a)** Pre-processed **(b)** Whithout Pre-processing.

### Additional file 3 — figure2a.eps

This is Figure 2a.

### Additional file 4 — figure2b.eps

This is figure 2b. The description under figures 1 a and b is: 99% Confidence Region for SOM's first two PC's - DATA2 **(a)** with 16 Nodes Pre-processed **(b)** with 30 Nodes - Pre-processed.

**FIGURE 1A**

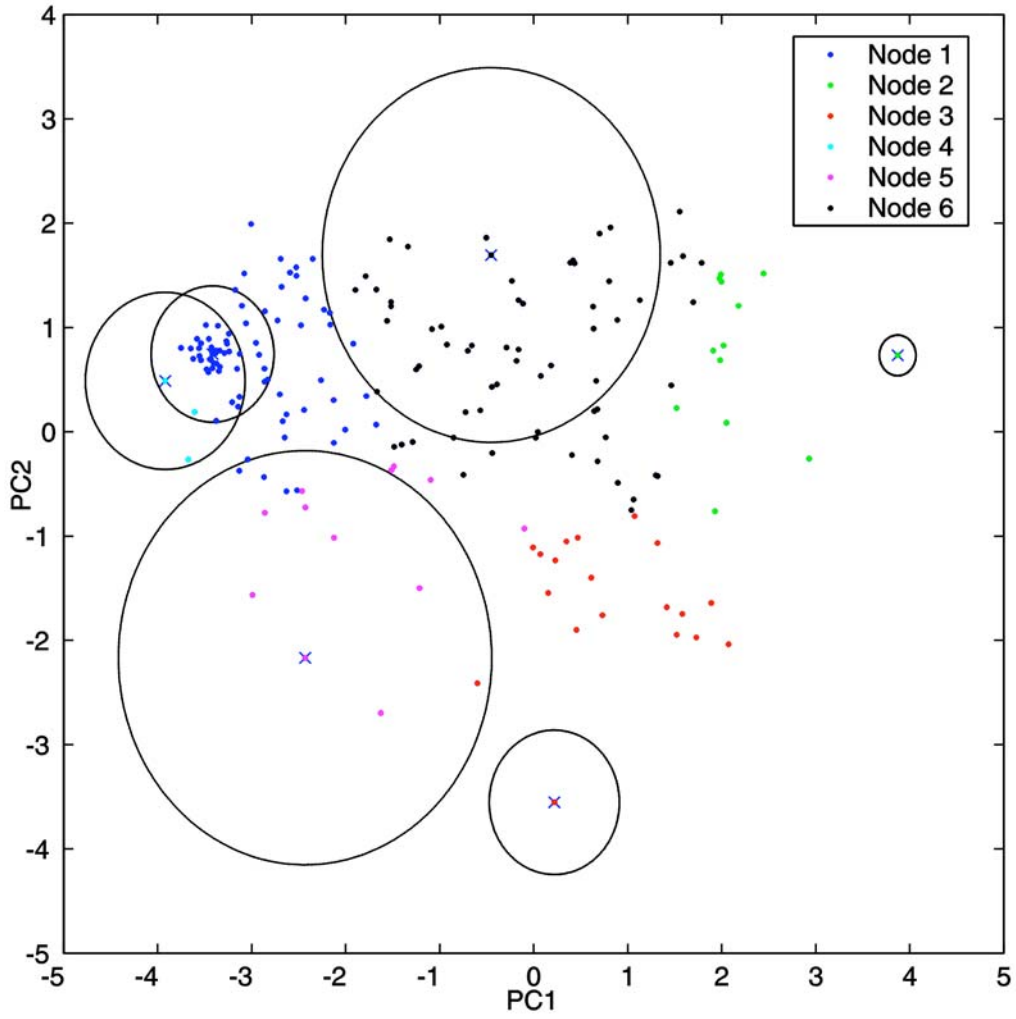

Node 1
Node 2
Node 3
Node 4
Node 5
Node 6
PC1
PC2
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

**FIGURE 1B**

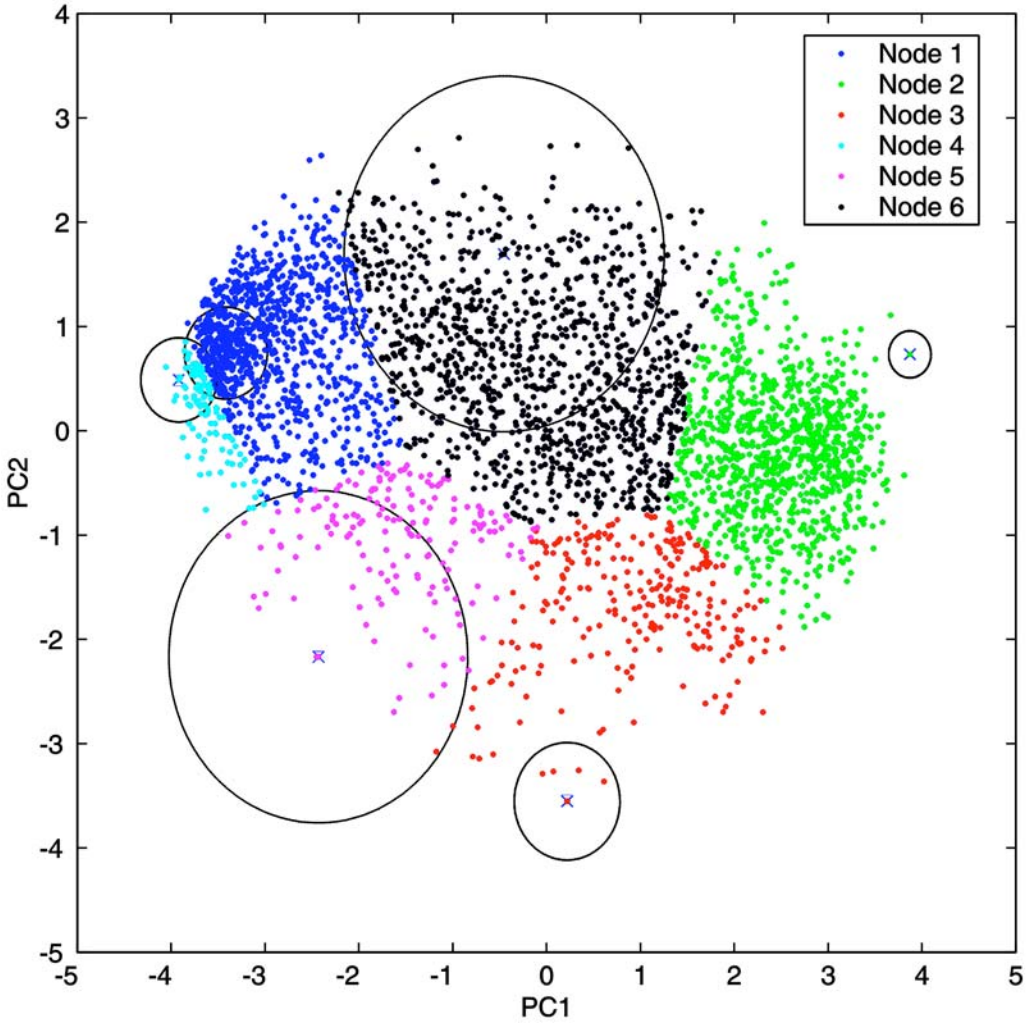

Node 1
Node 2
Node 3
Node 4
Node 5
Node 6
PC1
PC2
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5

**FIGURE 2A**

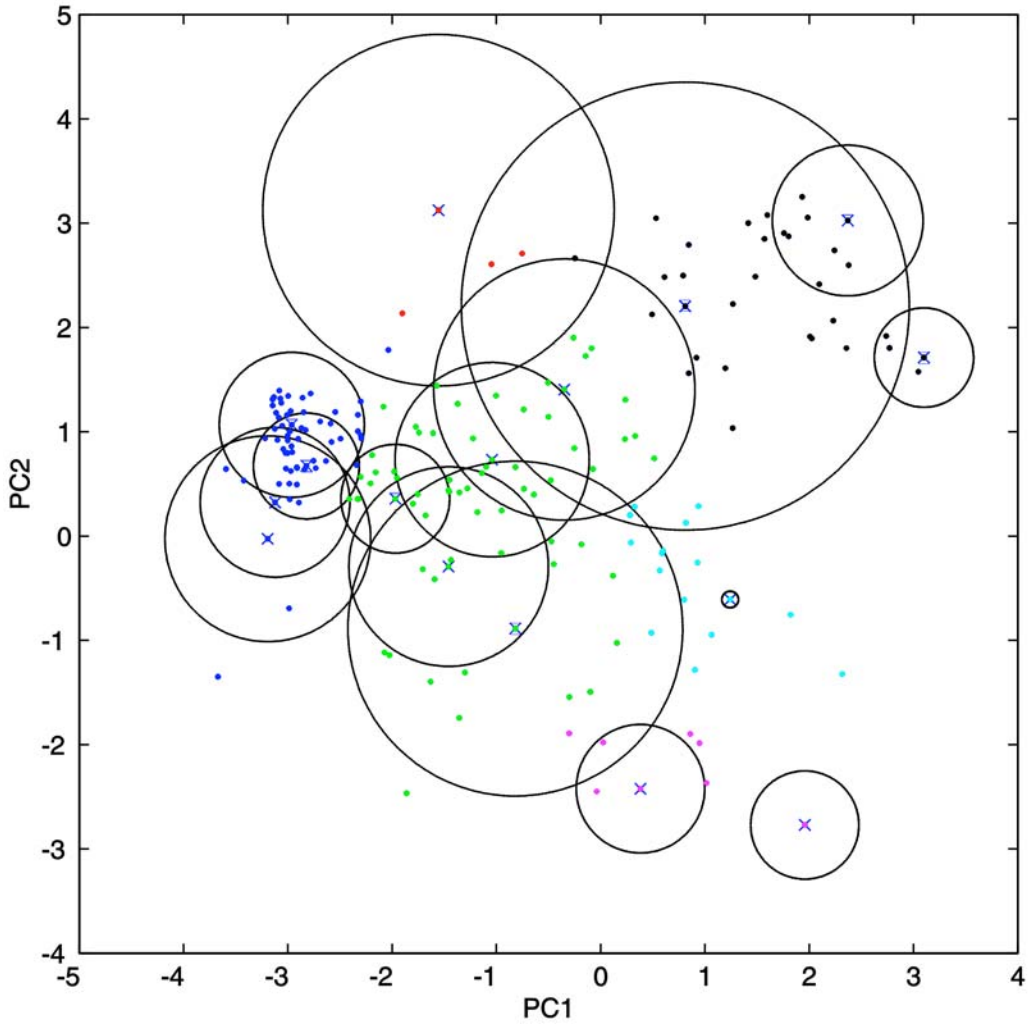
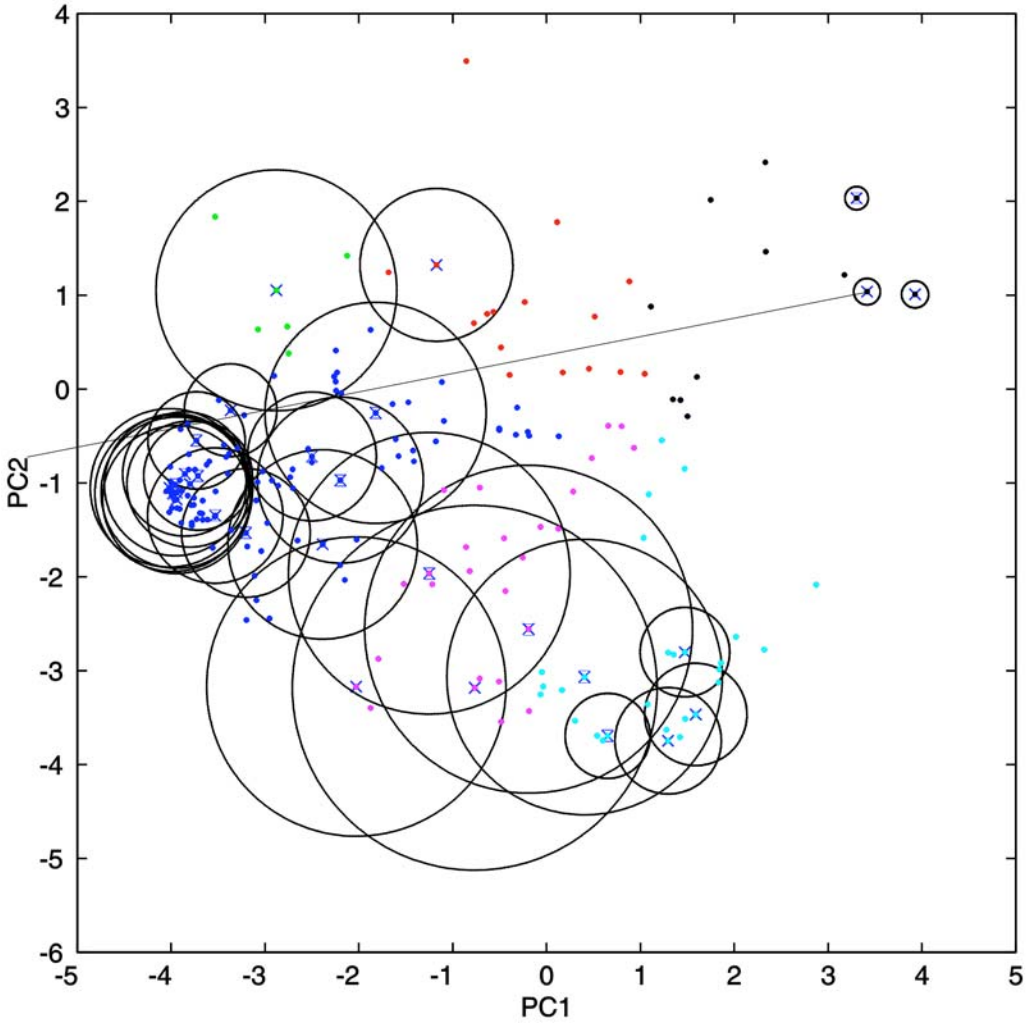


**FIGURE 2B**



# *Conclusões*

O principal objetivo deste trabalho foi a identificação de genes expressos em folhas de cana-de-açúcar que fossem induzidos pela aplicação foliar de uma solução diluída de etanol, visando desta forma, a obtenção de um sistema de indução química para controle da expressão gênica. Os resultados obtidos no presente trabalho, nos permitem concluir que:

- A técnica de macroarranjos de DNA foi eficiente e reprodutível durante a análise e identificação em larga escala, de genes com padrão alterado pelo tratamento com etanol em folhas de cana-de-açúcar;
- A aplicação externa de etanol em folhas de cana-de-açúcar foi responsável pela rápida indução da expressão de vários genes, tais como ERD4 e NPRI;
- O gene *SsNAC23*, já descrito como induzido por estresse de frio, estresse hídrico e herbivoria, também teve sua expressão induzida pela aplicação de etanol em folhas de cana-de-açúcar.
- A viabilidade do sistema *alc* de expressão gênica não pode ser avaliada de forma conclusiva até o momento em plantas de cana-de-açúcar, potencialmente transformadas com o cassete contendo o sistema *alc* controlando a expressão do gene *GUS*.

Andjelkovic V, Thompson R (2006) Changes in gene expression in maize kernel in response to water and salt stress. Plant Cell Reports 25: 71-79

Andrews J, Bouffard GG, Cheadle C, Lu J, Becker KG, Oliver B (2000) Gene discovery using computacional and microarray analysis of transcription in the *Drosophila melanogaster* testis. Genome Research 10:2030-2043

Aoyama T, Chua NH (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. The Plant Journal 11 (3): 605-612

Burnquist WL (2000) Os arquitetos da nova cana. Revista Pesquisa FAPESP 59: 3-4

Caddick MX, Greenland AJ, Jepson I, Krause KP, Qu N, Riddell KV, Salter MG, Schuch W, Sonnewald U, Tomsett AB (1988) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. Nature Biotechnology 16: 177-180

Cançado GMA, De Rosa Jr VE, Fernandez JH, Maron LG, Jorge RA, Menossi M (2005) Glutathione S-transferase and aluminum toxicity in maize. Functional Plant Biology 32: 1045-1055

Carvalho LCC (1993) Perspectivas da cultura da cana-de-açúcar para a década de noventa. In Camara GMS, Oliveira EAM (ed) Produção de cana-de-açúcar. Piracicaba FEALQ p. 1-17

Daniel J, Roach BT (1987) Taxonomy and evolution. In Heiz, DJ (Ed). Sugarcane improvement throught breeding. Elsevier, Amesterdam 7-84

De Araujo PG, Rossi M, de Jesus EM, Saccaro NLJr, Kajihara D, Massa D, Felix JM, Drummond RD, Falco MC, Chabregas SM, Ulian EC, Menossi M, Van Sluys MA (2005) Transcriptionally active transposable elements in recent hybrid sugarcane. Plant Journal 44: 707-717

De Rosa Jr VE, Nogueira FTS, Menossi M, Ulian EC, Arruda P (2005) Identification of methyl jasmonate responsive genes in sugarcane using cDNA arrays. Brazilian Journal of Plant Physiology 17: 173-180

Dong Q, Roy L, Freeling M, Walbot V, Brendel V (2003) ZmDB, an integrated database for maize genome research. Nucleic Acids Research 31: 244-247

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Research 9: 950-959

Falco MC, Silva-Filho MC (2003) Expression of soybean proteinase inhibitors in transgenic sugarcane plants: effects on natural defense agains *Diatraea saccharalis* (Lepdoptera: Crambidae). Plant Physiology and Biochemistry 41: 761-766

Felix JM, Drummond RD, Nogueira FTS, De Rosa Jr VE, Jorge RA, Arruda P, Menossi M (2002) Genoma Funcional: Uso de arranjos de DNA em náilon para a análise de expressão gênica em larga escala. Biotecnologia Ciência e Desenvolvimento 24: 60-67

Gatz C (1997) Chemical control of gene expression. Annual Review of Plant Physiology and Plant Molecular Biology 48: 89-108

Gatz C, Quail PH (1988) Tn10-encoded Tet repressor can regulated an operator-containing plant promoter. Proceedings of the National Academy of Sciences, USA 85:1394-1397

Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. Current Opinion in Plant Biology 5: 122-127

Heinz DJ, Tew TL (1987) Hybridization procedures. In: Heinz DJ ed. Sugarcane Improvement through breeding. Elsevier, Amesterdam, 313-342

Ingelbrecht IL, Irvine JE, Mirkov (1999) Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. Plant Physiology 119: 1187-1197

Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano HY., (2000) Molecular analysis of the NAC gene family in rice. Molecular General Genetics 262:1047-1054

Kuramae EE, Fenille RC, Rosa VE, Rosa DD, Ulian EC (2002) Minig the enzymes involved in the detoxification of reactive oxygen species (ROS) in sugarcane. Molecular Plant Pathology 4: 251-259

Ma H, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A. (2003) An EST survey of the sugarcane transcriptome. Theoretical and Applied Genetics 108: 851-863

Macedo IC, Nogueira LAH (2005) Biocombustíveis. Núcleo de Assuntos Estratégicos da Presidência da República: Secretaria de Comunicação de Governo e Gestão Estratégica. Disponível em http://www. Planalto.gov.br/secom/nae

Matsuoka S, Garcia AAF, Arizono H (1999a) Melhoramento de cana-de-açúcar. In: Borém A. ed. Melhoramento de espécies cultivadas. 2 ed. Viçosa, UFV, p. 205-251

Melo MO, Tanaka AS, Silva-Filho MC (2003) Molecular evolution of Bowman-Birk type proteinase inhibitors in flowering plants. Molecular Phylogenetic Evolution 27:103-112

Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH (1998) QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane. Genome Research 11:2075-2084

Moore PH (2005) Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum. Field Crops Research 92:119-135

Murkherjee, SK (1957) Origin and distribuition of *Saccharun*. Botanic Gas 119:55-61

Nogueira FTS, De Rosa VE, Menossi M, Ulian EC, Arruda P (2003) RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiology 132: 1811-1824

Nogueira FTS, Schlogl PS, Camargo SR, Fernandez JH, De Rosa Jr VE, Pompermayer P, Arruda P (2005) SsNAC23, a member of NAC domain protein family, is associated with cold, herbivory, and water stress in sugarcane. Plant Science 169:93-106

Papini-Terzi FS, Rocha FR, Vencio RZN, Oliveira KC, Felix JM, Vicentini R, Rocha CS, Simoes ACQ, Ulian EC, di Mauro SMZ, Da Silva AM, Pereira CAB, Menossi M (2005) Transcription prolling of signal transduction-related genes in sugarcane tissues. DNA Research 12: 27-38

Raizer AJ (1998) Interações genótipos x ambientes e estabilidade fenotípica em cana-de-açúcar no Estado de São Paulo. Dissertação de Mestrado – Escola Superior de Agricultura "Luis de Queirós", Universidade de São Paulo.

Ryan PR, Di Tomaso JM, Kochian LV. (1993) Aluminum toxicity in roots: an investigation of spatial sensitivity and the role of the root cap. Journal of Experimental Botany 44: 437-446

Saciloto RFZ (2003) Inserção do gene PR5K em cana-de-açúcar visando induzir resistência ao fungo da ferrugem *Puccinia melanocephala*. Dissertação de Mestrado, 74p., Escola Superior de Agricultura "Luiz de Queiróz", Universidade de São Paulo, Piracicaba.

Salter MG, Paine JA, Riddell KV, Jepson I, Greenland AJ, Caddick MX, Tomsett AB (1998) Characterization of the ethanol-inducible alc gene expression system for transgenic plants. The Plant Journal 16: 127-132

Schuler G. (1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes. Journal Molecular Medicine 75: 694-698

Soares-Costa A, Beltramini LM, Thiemann OH, Henrique-Silva F. (2002) A sugarcane cystatin:recombinant expression, purification and antifungal activity. Biochemical and Biophysical Research Communications 296: 1194-1199

Souza GM, Simoes ACR, Oliveira KC, Garay HM, Fiorini LC, Gomes FD, Nishiyama-Junior MY, Da Silva AM (2001) The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. Genetics and Molecular Biology 24: 25-34

Stevenson GC (1965) Genetics and breeding of sugarcane. London, Longmans. 284p

Tai PYP, Lentini RS (1998) Freeze damage of Florida sugarcane. In Anderson DI, eds, Sugarcane Handbook Ed 1, Florida Cooperative Extension Service, University of Florida, Gainesville, pp1-3

Tomashow MF (2001) So what's new in the field of plant cold acclimatation? Lots! Plant Physiology 125: 89-93

Van Loon LC, Antoniw JF. (1982) Comparison of the effects of salicylic acid and ethephon with virus-induced hypersensitivity and acquired resistance in tobacco. Netherlands Journal of Plant Pathology 88:237-256

Vettore AL, Da Silva FR, Kemper EL, Arruda P. (2001) The libraries that made SUCEST. General Molecular Biology 24: 1-7

White JA, Todd J, Newman T, Focks N, Girke T, De Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C. (2000) The metabolic pathway from carbohydrates to seed oil. Plant Physiology 124: 1582-1594

Wu Y, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fuji F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T. (2002) Plant Cell 14: 525-535

Xin Z, Browse J. (2000) Cold comfort farm: the acclimatation of plants to freezing temperatures. Plant Cell Enviromental 23: 893-902