UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia

Departamento de Genética e Evolução

Paulo Sérgio Schlögl

Caracterização Molecular e Funcional de Proteínas bZIPs de Cana-de-açúcar

Tese apresentada no Instituto de Biologia da Universidade Estadual de Campinas, para obtenção do título de Doutor em Genética e Biologia Molecular – área de concentração Genética Vegetal e Melhoramento.

Orientador: Dr Marcelo Menossi
Co-orientador: Dr Jörg Kobarg

Campinas
2004

Schlogl, Paulo Sérgio

Sch39c Caracterização molecular e funcional de proteínas bZIPs de cana-de-açúcar /Paulo Sérgio Schlogl.--
Campinas, SP: [s.n.], 2004.

Orientador: Marcelo Menossi Teixeira
Co-orientador: Jorg Kobarg

Tese (Doutorado) – Universidade Estadual de Campinas.

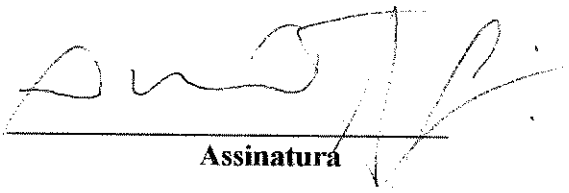
Instituto de Biologia.

1. Cana-de-açúcar. 2. Transcrição genética. 3. DNA. I. Teixeira, Marcelo Menossi. II. Kobarg, Jorg. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título

Campinas, 30 de junho de 2004.

Banca Examinadora

Prof. Dr. Marcelo Menossi Teixeira (Orientador)



Assinatura

Prof. Dr. Ivan de Godoy Maia

Assinatura

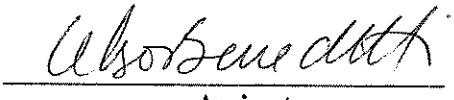
Prof. Dr. Eugênio César Ulian

Assinatura

Prof. Dr. Nilson Ivo Tonin Zanchin

Assinatura

Prof. Dr. Celso Eduardo Benedetti



Assinatura

Agradeço:

Ao saudoso orientador Dr Adilson Leite, pois sem ele grande parte deste trabalho não seria realizado.

Ao Dr Marcelo Menossi pelo empurrão final ao precipício da finalização dos trabalhos.

Ao Dr Jörg Kobarg pela grande ajuda, amizade e apoio nos momentos estressantes da finalização deste trabalho! Valeu...

Aos Professores da pré-banca e banca por todo o auxílio prestado na confecção deste manuscrito.

A todo o pessoal do Lab: Dudu, Dani, Mário, Sílvia.

Ao Fábio, o Vicente, o "cabeça" e o resto do pessoal do Genoma.

Aos "Manóides".

Aos "deuses" do Heavy Metal: Destruction, NEVERMORE, Sodom, Testament, Exodus, Iron Maiden, etc...Que tornaram mais fácil o último ano.

A FAPESP pelo auxílio técnico e pela bolsa.

E a todos que colaboraram de uma forma ou outra neste trabalho.

Índice

1. Resumo

Os fatores de transcrição bZIP ligam-se como dímeros a seqüências de DNA específicas presentes nas regiões promotoras, atuando como moduladores da expressão de vários genes em eucariotos. O conjunto completo e não redundante dos fatores de transcrição bZIP de *Arabidopsis thaliana* foi agrupado em 13 sub-famílias filogenéticas. Essa classificação permitiu organizar 85 clusters de cana-de-açúcar que codificam bZIPs, obtidos no projeto SUCEST. As sub-famílias IV e XIII de *A. thaliana* não foram encontradas em cana de açúcar, provavelmente devido ao baixo nível de expressão destas sub-famílias em cana-de-açúcar.

As análises filogenéticas com as bZIPs da sub-família *Opaco2* suportam um modelo de evolução que envolve a duplicação de dois genes homólogos antes da separação das monocotiledôneas e dicotiledôneas. Expansões posteriores resultaram em três duplicações gênicas nas monocotiledôneas e uma nas plantas dicotiledôneas. Assim, a proteína Opaco2 provavelmente é o resultado de uma duplicação específica das plantas monocotiledoneas e representa uma especialização restrita a essas plantas.

Uma proteína bZIP de cana-de-açúcar, SCbZIP1, foi clonada e caracterizada bioquimicamente. Ensaios de mobilidade eletroforética mostraram que a proteína SCbZIP1 liga-se fortemente a sondas de DNA do tipo G-box, C-box e Hex. A fosforilação por CKII reduz sua afinidade pelo DNA. SCbZIP1 foi capaz de produzir homodímeros e também heterodímeros com duas formas truncadas de Opaco2. O gene é ativo nos estágios iniciais do desenvolvimento de plântulas, sendo expresso nas gemas laterais e também nas flores.

O perfil de expressão dos 85 genes que codificam as bZIPs de cana-de-açúcar foi avaliado com macroarranjos de DNA. Sete genes foram diferencialmente expressos durante o desenvolvimento de plântulas e outros oito foram modulados pelos hormônios ácido abscísico ou metil jasmonato.

1.1. Abstract

bZIP transcriptional factors bind as dimers to specific DNA targets present in the promoter regions of their target genes and act as modulators of gene expression in eukaryotes. The complete set of *Arabidopsis thaliana* bZIPs was grouped in thirteen phylogenetic sub-families. This classification allowed the organization and identification of 85 sugarcane clusters from SUCEST project. Sub-families IV and XIII were not found in sugarcane, probably due to the low level of expression of these sub-families.

Phylogenetic analysis of the Opaque2 subfamily support a model of evolution that involves a duplication of two homologues before the separation of monocot and dicot plants. Subsequent expansions resulted in three gene duplications in monocots and one in dicots. Opaque2 is probably the result of one specific duplication of monocots and represents a restrict specialization in these plants.

A sugarcane bZIP, SCbZIP1 was cloned and biochemically characterized. Electrophoretic mobility shift assays showed that SCbZIP1 binds tightly to G- and C-boxes, Hex motifs, and phosphorylation of SCbZIP1 by CKII decreases its DNA affinity. SCbZIP1 underwent homodimerization and heterodimerization with truncated Opaque 2 from *Coix*. SCbZIP1 mRNA is expressed in early stages of development and in lateral buds and flowers.

The pattern of expression of 85 sugarcane bZIP genes was evaluated by cDNA arrays. Seven genes were differentially expressed during plantlet development and eight were modulated by Abscisic acid or Methyl-jasmonate.

2. Introdução

2.1. Regulação da expressão gênica em eucariotos

O controle da expressão gênica é essencial para a maioria dos fenômenos biológicos, tais como a diferenciação celular, desenvolvimento e respostas a estímulos ambientais. Além disso, a alteração na expressão gênica também está emergindo como causa de diversidade por trás da evolução morfológica dos organismos (Riechmann, 2002).

O controle da expressão gênica acontece basicamente em três níveis distintos:

I - Regulação da transcrição, que compreende a regulação da taxa de transcrição do gene, incluindo seu processamento;

II - Regulação pós transcricional, que regula a edição e tradução do RNA mensageiro já formado;

III - Regulação traducional, que abrange a regulação da tradução do RNAm nas proteínas.

Dentre esses três níveis, o mecanismo predominante no controle da atividade de um gene é a regulação da transcrição ou síntese do RNA mensageiro (RNAm) (Kuhlemeier, 1992; Kornberg, 1999).

2.2. A RNA polimerase II e os promotores dos genes eucarióticos

Organismos eucariontes possuem três enzimas responsáveis pela síntese de moléculas de RNA: as RNA polimerases I, II e III. A RNA polimerase I está envolvida com a transcrição dos genes de RNA ribossomal (RNAr) enquanto que a RNA polimerase III atua na transcrição de RNAs transportadores (RNAt), RNA ribossomal 5S e alguns outros pequenos RNAs nucleares (Kornberg, 1999; Lee e Young, 2000; Riechmann, 2002).

A RNA polimerase II é responsável pela transcrição de genes que codificam proteínas, sintetizando assim os denominados RNAs mensageiros (RNAm). Apesar de ser essencial para o processo de transcrição, essa enzima não é capaz de atuar sozinha. Ela depende de várias proteínas que interagem com o DNA e entre si, os chamados fatores gerais da transcrição, formando grandes complexos protéicos responsáveis pela regulação da síntese do RNAm, além

de outras proteínas que interagem com a própria RNA polimerase II formando um complexo protéico denominado holoenzima. Algumas dessas proteínas, denominadas fatores gerais são necessários para a transcrição de qualquer tipo de gene. Outras, denominadas reguladores (ativadores e repressores) modulam a taxa de transcrição de grupos de genes ou até mesmo de um único gene (Lee e Young, 2000).

Apenas a holoenzima tem capacidade de iniciar a transcrição e responder a ativadores, pois ela contém o complexo mediador da transcrição, formado por várias subunidades protéicas regulatórias (Hengartner et al., 1995; Kornberger, 1999). A composição da holoenzima e do complexo mediador pode ser remodelada de acordo com as necessidades celulares o que permite um controle coordenado de determinado conjunto de genes (Holstege et al., 1998). Os componentes do mediador apresentam funções importantes na regulação da transcrição provendo alvos para os ativadores e sinais regulatórios para a RNA polimerase II e outros componentes do complexo de iniciação (Lee e Young, 2000).

A síntese dos RNAm inicia-se em regiões especificas do genoma, mais precisamente na região flanqueadora 5' dos genes que codificam proteínas. Essas regiões são também conhecidas como promotores. Estas regiões promotoras apresentam pelo menos três sítios alvos de ligação comuns: o sítio de início da transcrição, um sítio de ligação para a proteína TATA Box binding, conhecido como TATA box e seqüências de ligação específicas para outros fatores de transcrição (Lee e Young, 2000).

As seqüências específicas que são identificadas pelos fatores de transcrição são: seqüências ativadoras e ou repressoras "upstream", "enhancers", regiões controladoras de lócus (LCRs, que são similares aos enhancers), seqüências de reconhecimento de TFIIB (BRE), elementos "downstream" e silenciadores (Figura 1) (Lee e Young, 2000). Os fatores de transcrição ligam-se a esses elementos presentes nos promotores estimulando ou reprimindo a transcrição dos genes pela interação com outros fatores presente na holoenzima (Korneberg, 1999; Lee e Young, 2000).

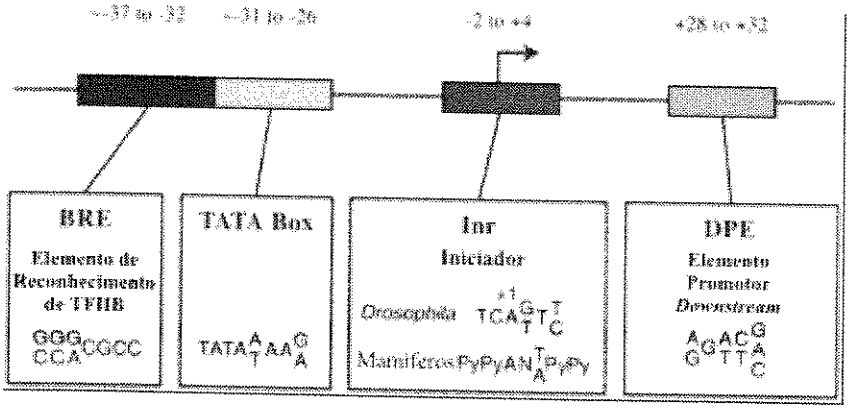


Figura 1 - Composição de um típico promotor eucariótico. O diagrama mostra alguns dos elementos que contribuem para a transcrição basal. Cada um destes motivos de DNA é encontrado em alguns promotores. Um promotor particular pode conter todos ou alguns destes elementos. O TATA box pode funcionar na ausência de BRE, Inr e DPE. Em contraste, DPE requer a presença de Inr. BRE se localiza upstream de TATA. Fonte: Smale and Kadonaga, 2003.

Em muitos promotores da RNA polimerase II foi identificada uma seqüência conservada, localizada a – 25 ou 30 pares de base (pb) do local do início da transcrição. Esta seqüência, formada pelo consenso TATAa/tAa/t, foi chamada TATA box (Figura 1). Apesar de vários promotores de genes apresentarem claramente o elemento TATA-box, a maioria não o faz, sendo que alguns genes possuem sucessões A+T-ricas que parecem funcionar como TATA-box. Embora, a região TATA box seja bem caracterizada, as TPBs (TATA binding proteins ou proteínas ligantes a TATA) ligam-se a uma ampla gama de seqüências derivadas de TATA box, o que pode explicar a variação encontrada neste sítio alvo, embora estas sejam as proteínas mais conservadas entre os eucariotos (Smale e Kadonaga, 2003).

Uma outra seqüência rica em pirimidinas e de consenso YYANt/aYY foi localizada próxima à maioria dos locais de início da transcrição (Figura 1), sendo chamada de elemento iniciador (Inr). Vários fatores de transcrição podem ligar-se a Inr e recrutar a maquinaria de transcrição (Lee e Young, 2000; Smale e Kadonaga, 2003). As seqüências Inr funcionam de maneira similar a TATA box, pois alguns estudos demonstraram que ambos funcionam de forma sinérgica quando separados por 25-30 pb e atuam individualmente se separados por mais de 30 pb (O'Shea-Greenfield e Smale, 1992). Há evidências de que TFIID reconhece e liga-se ao Inr, recrutando a RNA polimerase II durante a formação do complexo de pré-iniciação (Smale e Kadonaga, 2000). A RNA polimerase II, o TAFIIA e o TBP também se ligam ao Inr, formando

complexos estáveis em promotores mesmo sem a presença de TATA box (Weis e Reinberg, 1997).

O elemento promotor "downstream", DPE (Figura 1) é um motivo que é requerido para a ligação do fator geral de transcrição, TFIID, geralmente em promotores que não apresentam TATA box. DPE atua somente em conjunto com Inr mas não atua de forma independente como no caso do TATA box e do Inr (Smale e Kadonaga, 2003). Uma análise de 205 genes de *Drosophila* mostrou uma estimativa sobre a composição dos motivos presentes em promotores, onde 29% contém TATA box sem DPE, 26% possuem DPE sem TATA box, 14% contém ambos e 31% parecia não conter nenhum dos dois motivos (Kutach e Kadonaga, 2000). Esse estudo ainda mostrou que DPE também apresenta ligeiras variações nas suas seqüências funcionais. Tanto DPE como TATA box são reconhecidas e ligadas por TFIID, mas apenas TATA box pode atuar independente de Inr. A principal diferença entre ambos é que TATA box é reprimida pelo complexo protéico NC2/Dr1-Drap1 e o mesmo ativa a transcrição em DPE (Smale e Kadonaga, 2003).

Outra seqüência que está presente em vários promotores de genes é o elemento de reconhecimento de TFIIB (BRE) que está localizada "upstream" de TATA box (Figura 1). Ao contrário de outros elementos encontrados em promotores de eucariotos, BRE não foi caracterizado, pelo menos ainda, em plantas e leveduras. Em humanos BRE interage com TFIIB e estimula a RNA polimerase II durante a transcrição, mas parece reprimir a transcrição em alguns promotores (Smale e Kadonaga, 2003).

Embora os sítios alvos comuns presentes nos promotores sejam conservados, a composição e o contexto funcional dos mesmos nos diferentes promotores variam (Lee e Young, 2000; Riechamann, 2002). Seqüências auxiliares à essas chamadas seqüências "core" são encontradas nos promotores e são alvos para fatores de transcrição específicos que auxiliam na regulação da transcrição. A variação da composição dos promotores eucarióticos leva a hipótese da regulação gênica combinatória, onde o limitado número de fatores de transcrição de um organismo pode regular uma enorme variedade de padrões diferentes de expressão gênica se a regulação da transcrição dos mesmos requerer a ação de múltiplos reguladores (Lee e Young, 2000; Riechmann, 2002).

2.3. A transcrição e a cromatina

Em eucariotos a compactação do DNA na cromatina é importante para a organização do mesmo dentro dos limites do núcleo e na própria regulação gênica (Lee e Young, 2000; Myers e Kornberg, 2000; Riechmann, 2002). A combinação do DNA com as histonas na formação do nucleossomo e a organização destes na cromatina parece restringir o acesso das proteínas que participam da transcrição. Dessa forma, a modificação da estrutura da cromatina e dos nucleossomos facilita a ligação dos ativadores transcricionais e do resto da maquinaria transcricional. Geralmente, a modificação da cromatina ocorre pela acetilação, deacetilação e outras modificações das histonas que desestruturam a cromatina (Kornberg, 1999; Lee e Young, 2000; Myers e Kornberg, 2000). A diminuição da densidade dos nucleossomos altera a expressão de vários genes em levedura (Lee e Young, 2000). O sinergismo entre a interação de múltiplos fatores de transcrição leva a ativação do promotor do vírus de tumor mamário favorecida pelo posicionamento correto nas regiões regulatórias dos nucleossomos (Chaves e Beato, 1997). A ativação do promotor de vitelogenina de *Xenopus* é potencializada pela geração de um "loop" dependente de nucleossomo e o ativador transcricional Hnf3 liga-se estavelmente na sua seqüência alvo somente quando o DNA está empacotado nos nucleossomos (Cirillo e Zaret, 1999). Juntos esses resultados sugerem que o empacotamento do DNA na cromatina fornece um contexto físico para diferentes promotores, aumentando a viabilidade de uma regulação mais específica para os genes dentro das células (Lee e Young, 2000).

A ultra-estruturação da cromatina envolve contato entre os nucleossomos que é mediado pelo menos em parte, pelas regiões N-terminais das histonas. As histonas são alvo de diferentes tipos de modificações tais como, acetilação, fosforilação, metilação e ubiquitinização, o que pode levar a uma alteração estrutural da organização da cromatina. Sendo que existe uma forte correlação entre a acetilação da cauda das histonas, a remodelagem da cromatina e a atividade transcricional (Lee e Young, 2000; Myers e Kornberg, 2000; Riechmann, 2002).

A hiperacetilação das histonas está associada com domínios transcricionalmente ativos e acessíveis da estrutura da cromatina, ocorrendo o oposto em regiões com histonas hipoacetiladas (Hebbes et al., 1994). Provavelmente, isso reflete a remodelagem da cromatina, o que leva a um acesso maior as regiões de ligação da maquinaria transcricional e seus ativadores ou repressores (Garcia-Ramirez et al., 1995; Lee e Young, 2000, Riechmann, 2002).

As primeiras evidências da importância da acetilação na transcrição foram obtidas em *Tetrahymena*, com a proteína p55, que apresentou atividade de histona acetilase (HAT) e que foi relacionada funcionalmente à proteína GCN5 de levedura (Brownell e Allis, 1996). Gcn5 é um fator transcricional e a região promotora onde Gcn5 se liga mostrou alta acetilação das histonas (Lee e Young, 2000). Outros fatores de transcrição, CREB, p53 e receptores hormonais nucleares, ligam-se a um conjunto de co-ativadores, p300/CBP, sendo que este apresenta atividade de acetilar as histonas (Lee e Young, 2000).

A regulação da expressão gênica geralmente envolve um balanço entre a ação de ativadores e repressores que se ligam às regiões promotoras dos genes. Os ativadores estimulam a atividade da maquinaria transcricional e geralmente um ativador pode ser utilizado na ativação de vários genes no genoma, o que promove um mecanismo coordenado no controle dos mesmos. Outros genes podem necessitar a atuação de múltiplos ativadores, o que permite um mecanismo de controle combinatório para a célula (Lee e Young, 2000). Os ativadores ligam-se geralmente de forma cooperativa nas regiões promotoras e mesmo em baixas concentrações podem ativar de uma forma ampla a transcrição. Assim, o arranjo dos ativadores em um complexo único promove a integração de vários *inputs* regulatórios em um único *output* (Lee e Young, 2000).

Uma outra característica importante desses ativadores da transcrição é a capacidade dos mesmos de recrutar os complexos responsáveis pela modificação da cromatina. Evidências dessa função dos ativadores vieram através dos estudos de ativação do gene *HO* em levedura. O fator de transcrição Swi5 recruta o complexo modificador de cromatina Swi/Snf que acetila as histonas antes da associação do segundo ativador, SBF, que então recruta o resto da maquinaria de transcrição para a região promotora (Cosma et al., 1999).

O conjunto de acetilases, p300/CBP é alvo para diferentes ativadores da transcrição incluindo, c-jun, Pit-1 e CREB e é sugerido que p300/CBP sirva de plataforma para o recrutamento de cofatores adicionais (Lee e Young, 2000).

Elp3 é uma acetiltransferase (HAT) que faz parte do complexo de elongamento em levedura que está fortemente associado a RNA polimerase II. Essa ligação física entre uma acetilase e a RNA polimerase II pode prover meios de modificar o estado de acetilação dos nucleossomos nas regiões transcricionais na cromatina (Lee e Young, 2000). Além das histonas, as acetiltransferases podem exercer suas atividades de acetilação em outras proteínas do complexo

de transcrição incluindo os fatores de transcrição tais como, TFIIF, o que produz efeitos positivos e negativos na transcrição dependendo do gene (Imhof et al., 1997). A acetilação de p53 por CBP/p300 e PCAF aumenta a atividade de ligação ao DNA deste fator de transcrição humano. Esse resultado sugere que a ação das acetiltransferases também pode ser crítica na regulação de p53 (Lee e Young, 2000).

A ação antagônica das deacetilases também é importante durante a regulação da transcrição, pois estas enzimas tem sido identificadas como componentes comuns de complexos co-repressores recrutados por diversos dos reguladores da transcrição, indicando que sua função esteja relacionada à repressão gênica. Sin3 é um complexo repressor de levedura, caracterizado pela presença da proteína Sin3 e as deacetilases envolvidas na formação deste complexo são importantes para a função repressora do mesmo (Alland et al., 1997). O complexo Sin3 é capaz de interagir com vários reguladores da transcrição, refletindo os diversos mecanismos empregados para a deacetilação em promotores específicos ou em um conjunto de promotores (Ayer, 1999). Todos esses complexos que remodelam a cromatina contém ATPases essenciais para tal remodelamento e ainda outras subunidades que afetam a regulação, especificidade, e a eficiência da transcrição (Lee e Young, 2000).

2.4. A RNA polimerase II: a holoenzima

A RNA polimerase II e o complexo para iniciação da transcrição são recrutados para ligarem-se ao promotor dos genes pelos fatores de transcrição. O complexo contém a RNA polimerase II, os fatores gerais da transcrição e os complexos chamados de co-ativadores ou mediadores (Kornberg, 1999; Lee e Young, 2000). Todo esse aparato protéico é denominado holoenzima, sendo que a holoenzima melhor caracterizada é a de levedura, contendo a RNA polimerase II, os fatores de transcrição e o complexo mediador Srb (Cramer et al., 2000). O complexo mediador Srb integra sinais entre os fatores de transcrição nos promotores e sua composição pode ser remodelada dependendo das condições encontradas pelas células, o que permite um controle específico para um dado grupo de genes. Vários co-ativadores, homólogos ao complexo Srb, foram isolados em mamíferos capazes de ativar a transcrição de diferentes genes (Lee e Young, 2000).

Uma RNA polimerase II típica apresenta de 10 a 12 subunidades e é capaz de transcrever o RNA *in vitro*, mas incapaz de reconhecer promotores especificos na ausência de fatores de transcrição adicionais. Em levedura todas as 12 subunidades da RNA polimerase II são imprescindíveis para o crescimento e desenvolvimento.

A maior subunidade da RNA polimerase II contém um domínio repetitivo na cauda C–terminal (Tyr-Ser-Pro-Thr-Ser-Pro-Ser), chamado CTD (Figura 2), que é altamente conservado entre eucariotos, variando apenas o número das repetições (Lee e Young, 2000). A função deste domínio parece estar envolvida com o estado de fosforilação. RNA polimerases II sem estarem fosforiladas no domínio CTD são encontradas nos complexos de iniciação (Chesnut et al., 1992), enquanto fosforiladas estão na fase de elongamento (Lee e Young, 2000). A fosforilação do domínio CTD ocorre durante a transição do estado de iniciação para o de elongamento e esta transição leva uma troca de cofatores pela holoenzima. O complexo mediador Srb é associado com as RNA polimerases II com CTD não fosforilado em contraste, o complexo de elongação e vários outros fatores de processamento de RNA se associam com a RNA polimerase II fosforilada (Lee e Young, 2000).

Além da transcrição a RNA polimerase II auxilia na proteção do RNAm, uma vez que o domínio CTD fosforilado interage com a enzima de "capping", além de afetar a atividade de outros componentes do aparato de capping do RNAm (Cho et al., 1997 e 1998). O domínio CTD é essencial para a clivagem do RNAm na região de poliadenilação. Desta forma, a RNA polimerase II com seu domínio CTD parece ser importante também, para a estimulação e recrutamento dos fatores de poliadenilação e assim, produzir uma maquinaria eficiente e coordenada para a formação da extremidade 3' e para a terminação dos transcritos (Bentley, 1999). Algumas evidências sugerem que o spliceossoma pode estar física e funcionalmente ligado ao domínio CTD da RNA polimerase II (Lee e Young, 2000).

Duas quinases que fosforilam o domínio CTD foram caracterizadas e devido a sua forte associação com o aparato de iniciação sugere que tenham funções na regulação deste processo. A Cdk7 uma subunidade do fator geral de transcrição TFIIH tem atividade de quinase do domínio CTD e é critica na transição para que o complexo de elongamento seja estável (Feaver et al., 1991). A quinase Srb10/Cdk8 é um componente do complexo mediador Srb de levedura

que também fosforila o domínio CTD e Srb10 atua como um regulador negativo da transcrição em condições ótimas de crescimento (Holstege et al., 1997).

A estrutura cristalográfica da RNA polimerase II de levedura revelou características interessantes dessa enzima (Figura 2) (ver Cramer et al., 2000 para maiores detalhes). As duas maiores subunidades (Rpb1 e Rpb2), formam uma fenda que contém o sítio ativo e a posição deste em relação à projeção do DNA na fenda sugere que o DNA não segue um caminho direto por entre a enzima. Dois canais foram identificados por onde sai o RNA mensageiro nascente. Foram identificados dois orifícios na base da fenda que podem ser para a entrada ou saída de nucleotídeos, do RNA ou ainda de fatores que afetam a clivagem 3'-5'do RNA durante a correção de erros (Cramer et al., 2000). Enquanto a RNA polimerase II típica é considerada uma unidade funcional, suas subunidades apresentam diversas funções e isso pode sugerir que exista uma multiplicidade de formas de RNA polimerase II nas células (Lee e Young, 2000).

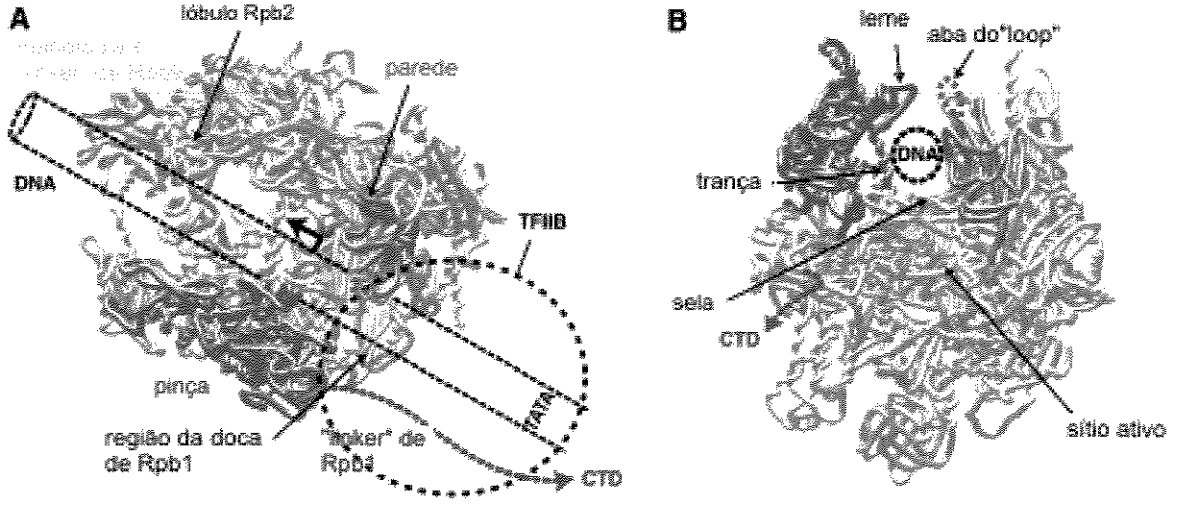


Figure 2 - Estrutura tridimensional da RNA polimerase II. Em **A** é mostrado uma visão superior. O DNA é mostrado como um cilindro. A subunidade Rbp9 envolvida no reconhecimento e seleção do sítio de início de transcrição é mostrada em laranja. Em amarelo são mostrados sítios importantes para a seleção do sítio de iniciação. O círculo ponteado mostra o local do posicionamento do fator geral de transcrição TFIIB durante a formação do complexo de iniciação. Em **B** é mostrado uma visão frontal da RNA polimerase II. Fonte: Cramer et al., 2001.

As atividades mais comuns das subunidades caracterizadas foram: seleção do sítio de iniciação, elongação da transcrição e interação com ativadores (Lee e Young, 2000).

Ppb5, Rpb1 e Rpb9 formam estruturas que posicionam o DNA "downstream". Já Rpb1, Rpb2 e Rpb6 ajudam na estabilidade do complexo de transcrição (Cramer et al., 2000). Rpb4 e Rpb7 formam um complexo dissociável que está implicado na resposta a estresse e na iniciação da transcrição (Edward set al., 1991; Choder e Young, 1993) e outros estudos mostram que Rpb4 e Rpb7 estão envolvidos na interação do DNA e o sítio ativo da RNA polimerase II (Asturias et al., 1997).

2.5. Os fatores basais e a iniciação da transcrição

O conjunto basal ou geral de fatores de transcrição (GTF) requeridos para o reconhecimento específico dos promotores pela RNA polimerase II *in vitro* incluem TFIIA, TFIIB, TFIID, TFIIE, TFIIF e TFIIH (Lee e Young, 2000). Não é clara a função de cada um destes GTFs e possivelmente outros componentes do aparato de transcrição possam ser generalizados como GTFs.

A ordem de ligação para estes GTFs no processo de transcrição *in vitro*, de acordo com vários estudos realizados, começa com a ligação de TFIID ou TBP (TATA binding protein), seguido de TFIIA, TFIIB, um complexo de RNA polimerase II, TFIIF, TFIIE e por fim TFIIH (Lee e Young, 2000; Riechmann, 2000). Após a montagem do aparato de inicio de transcrição, o TFIIH abre o DNA em aproximadamente 12-15 pb e há então, a formação das primeiras ligações fosfodiéster. Normalmente, as RNA polimerases repetem esse ciclo de iniciação da transcrição liberando pequenos RNAs, um processo chamado de iniciação abortiva (Lee e Young, 2000). Eventualmente, a RNA polimerase II escapa das iniciações abortivas e produz um RNA mais longo, mas a RNA polimerase tende a pausar entre 25-30 pares de bases do sítio de início em muitos promotores (Lee e Young, 2000). Estes complexos de elongamento precisam sofrer uma transição crítica para um complexo de elongamento mais hábil para escapar do promotor (Conaway et al., 1998).

O primeiro passo para a formação do complexo de pré-iniciação nos promotores apresentando TATA box é a interação de um pequeno fator basal denominado TBP (TATA *binding protein*) com TATA-box. TBP teve sua estrutura definida por cristalografia e os resultados mostraram que TBP parece uma cela que se apóia no DNA contatando o sulco menor na região de TATA box (Figura 3) e induz uma dobra acompanhada da abertura do mesmo que parece ser importante para os eventos de iniciação da transcrição (Lee e Young, 2000).

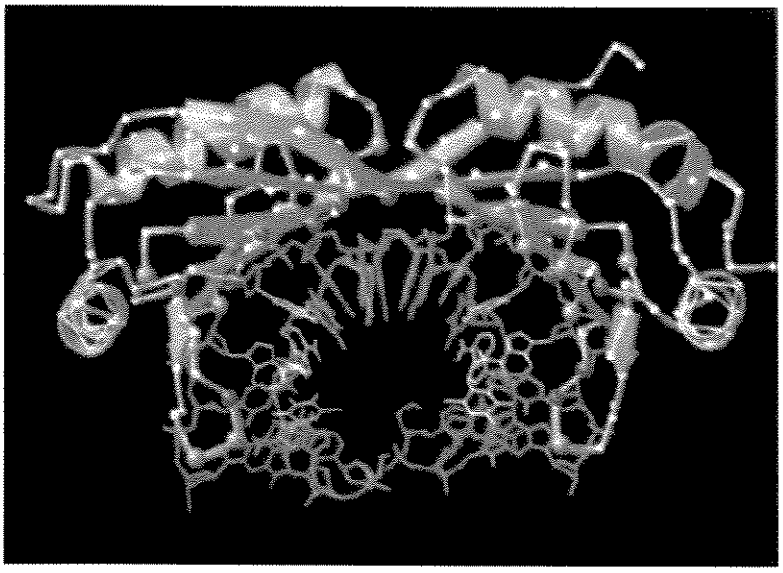


Figura 3 – Estrutura tridimensional da proteína TBP humana ligada a DNA. Os resíduos de aminoácido que interagem com o sulco menor do DNA são mostrados em vermelho e os resíduos que interagem com o esqueleto açúcar/fosfato são mostrados em verde. As α-hélices (e pregas beta são mostradas ao longo da proteína no sentido N-C terminal. Fonte: Näär et al., 2001.

TBP e seus fatores associados (TAFs) formam o complexo TFIID (Figura 4) e o número de TAFs associados a TBP nestes complexos varia entre as diferentes espécies de eucariotos.

Os TAFs são necessários para a ativação da transcrição por uma série de ativadores, mas não para a transcrição basal, sugerindo que os ativadores podem utilizar diferentes TAFs como adaptadores para recrutar TFIID nos diferentes promotores (Lee e Young, 2000). Os TAFs auxiliam na estabilidade da ligação de TFIID ao promotor fazendo interações específicas com as seqüências iniciadoras (INR) e elementos promotores "downstream" (DPE), que estão localizadas perto do sítio de início da transcrição de uma ampla variedade de promotores (Figura 1) (Burke e Kadonaga, 1996 e 1997). A forte afinidade de TFIID pelo promotor conferida pelos TAFs é necessária para permitir altos níveis de atividade transcricional no ambiente restritivo da cromatina (Näär et al., 2001). Vários estudos demonstraram que as subunidades TAF150 e TAF250 (Figura 4) de TFIID ligam-se a Inr e parecem ser importantes para a função deste (Lee e Young, 2000). A TAF250 apresenta uma série de funções que auxiliam TFIID na ativação transcricional entre elas, a de histona acetiltransferase (HAT) e proteína quinase (Imhof et al., 1997; Lee e Young, 2000). Foi demonstrado recentemente que

TAF250 participa da conjugação da ubiquitina, sendo que esta é requerida para a ativação da transcrição em determinados promotores em embriões de *Drosophila* (Pham e Sauer, 2000) (Figura 2). TAF250 apresenta dois Bromo-domínios, que são módulos estruturais encontrados em muitos co-ativadores e fatores remodeladores de cromatina (Näär et al., 2000).

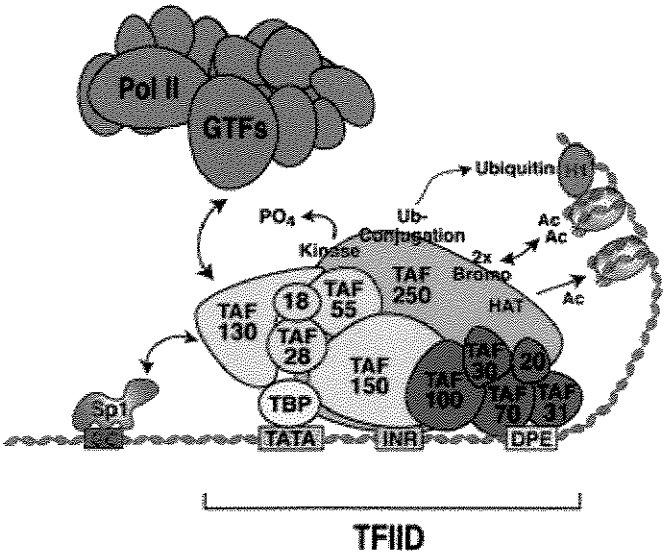


Figura 4 – Modelo das funções associadas com TFIID e suas subunidades. O recrutamento do ativador é iniciado por TAF130 que interage com o transativador Sp1. Várias subunidades de TFIID foram implicadas com ligação ao DNA, incluindo TBP que liga ao TATA box. TAF150 interage com o elemento iniciador (INR). TAF70 e TAF31 interagem com um elemento promotor "downstream" (DPE). O sub-módulo de TAFs mostrados em vermelho interagem com outros complexos ativadores. TAF250 parece exercer várias atividades tais como, de quinase, acetilase de histonas (HAT), de ligação a ubiquitina e na interação com histonas pelo seu Bromodomínio (2x Bromo). Ac: acetil, H1: histona 1, Ub: ubiquitina. TFIID forma o complexo de pré-iniciação, recrutando a RNA polimerase II (RNA pol II) e outros fatores gerais de transcrição (GTFs). Aqui as subunidades representam as proteínas humanas, embora algumas funções tenham sido identificadas em *Drosophila sp*. TAFs: fatores associados a TPB. Fonte: Näär et al., 2001. Ubiquitin: ubiquitina; conjugation: conjugação e Kinase: Quinase.

A próxima etapa de formação do complexo de pré-iniciação é a associação de TFIIA, um fator adicional que interage de maneira estável ao complexo, através de contatos diretos com TBP e com a seqüência de DNA na região 5' do promotor (Lee e Young, 2000). A presença de TFIIA é importante para o processo de transcrição basal quando é necessário estabilizar a interação de TBP com o DNA durante a expressão de alguns genes. TFIIA interage com outros ativadores, além de funcionar como antagonista de repressores da transcrição, impedindo fisicamente a interação destes com o complexo TFIID (Auble et al., 1994; Ge e Roeder, 1994) e esse mecanismo poderia explicar a variação da necessidade de TFAIID para os diferentes promotores (Lee e Young, 2000).

O fator TFIIB interage com o complexo através de interações diretas com TBP e com a seqüência de DNA na região 3' do promotor. Uma das principais funções de TFIIB é a seleção do sítio de início da transcrição, possivelmente, pelo ajuste das distâncias entre os promotores e o sítio de início da transcrição. Estudos estruturais demonstraram que à distância entre TFIIB e o sítio catalítico da RNA polimerase II é de 32 pb, que representa a distância entre TATA box e o sítio de início da transcrição (Lee e Young, 2000). Assim como TFIIA, TFIIB ajuda a estabilizar as interações de TBP-TATA box.

A próxima etapa inclui a interação de TFIIF com a RNA polimerase II e provavelmente interações com TFIIF e TBP. TFIIF tem características semelhantes ao do fator bacteriano sigma, pois, ele se liga a RNA polimerase II suprimindo interações inespecíficas com o DNA e ajuda a estabilizar o complexo de pré-iniciação (Lee e Young, 2000). TFIIF afeta também a topologia do DNA, sendo que TFIIF é crítico para que as interações com o DNA sejam fortes facilitando assim, com uma torção no DNA, a abertura do mesmo na região do promotor. TFIIF também estimula as taxas de elongamento da RNA polimerase II pela supressão das pausas transientes durante a transcrição, talvez interagindo com fatores de elongamento ou devido a sua atividade de quinase (Lee e Young, 2000).

Em seguida ocorre a interação de TFIIE com o complexo e sua função parece estar envolvida na manutenção da abertura do DNA no promotor, pois TFIIE liga a DNA simples fita. É sugerido que a exigência de TFIIE no complexo de transcrição seja promotor específica, pois há genes que não necessitam a presença do mesmo, tanto *in vivo* como *in vitro* (Holstege et al., 1998).

A formação do complexo de pré-iniciação é completada pela interação de TFIIH. As subunidades que compõe esse fator geral apresentam três atividades, a de ATPase dependente de DNA, helicase ATP dependente e CTD quinase (Lee e Young, 2000). O Mutações nas subunidades que apresentam a atividade de helicase são responsáveis por vários tipos de doenças genéticas em humanos, incluindo *Xeroderma pigmentosum*, síndrome de Cocakayne e tricodiodistrofia (Coin e Egly, 1998). TFIIH apresenta atividade de quinase que fosforila a RNA polimerase II na região CTD que permite o processo de transição para a fase de elongamento do RNA.

Para que a holoenzima RNA polimerase II produza os seus transcritos deve haver a troca e interação de cofatores associados ao processo de elongamento e como discutido anteriormente, a

fosforilação do motivo CTD facilita todo este processo (Lee e Young, 2000). Além disso, o CTD quando fosforilado recruta a enzima que faz o "cap" do RNAm nascente e auxilia na liberação do promotor (Cho et al., 1997).

Vários fatores estão envolvidos nos processos de liberação do promotor, no escape e na eficiência do elongamento do RNA. O escape do promotor é um passo importante dentro da regulação do início da transcrição, havendo um balanço entre fatores regulatórios que favorecem ou desfavorecem o mesmo (Lee e Young, 2000). Os fatores DRB e DSIF e NELF em conjunto inibem da síntese de RNAm e a fosforilação do domínio CTD da RNA polimerase II (Dubois et al., 1994). Já alguns fatores de transcrição contribuem para o processo de elongamento são eles, TFIIF, TFIIH e P-TEFb. Este último antagoniza as ações de NELF e DSIF, pois foi isolado em *Drosophila* como um fator de elongamento e que fosforila CTD (Lee e Young, 2000). Um fator humano similar foi descrito e caracterizado como uma quinase que estimula a transcrição da proteína Tat de HIV (Gold et al., 1998). TFIIF diminui a freqüência das iniciações abortivas e previne paradas do complexo de transcrição durante os passos iniciais do elongamento. O mecanismo de ação de TFIIH não é claro, mas é sabido que ele auxilia no enrolamento do DNA a RNA polimerase II, além de associar-se com fatores de elongamento e helicases (Lee e Young, 2000).

Outro fator de transcrição isolado por sua propriedade de auxiliar a RNA polimerase II a sintetizar os longos RNAs, foi TFIIS. TFIIS auxilia a RNA polimerase II a suplantar vários impedimentos durante a transcrição tais como, sítios de pausa intrínsecos. TFIIS estimula a RNA polimerase fazendo com que esta ative a sua função de endoribonuclease, fazendo com que a RNA polimerase clive o RNAm nascente, o que auxilia o escape para a fase de elongamento (Lee e Young, 2000).

2.6. Regulação da expressão gênica em plantas

As plantas durante os processos de desenvolvimento e diferenciação, resposta a mudanças ambientais, estresse e defesa precisam integrar uma ampla variedade de sinais para regular seus complexos padrões de expressão gênica.

Em *Arabidopsis thaliana* cerca de 3.000 genes são envolvidos na regulação da transcrição gênica, sendo que muitos deles são ativadores ou repressores transcricionais, outros são

moduladores da estrutura da cromatina ou contribuem para o aparato transcricional (Finnegan, 2001).

Como demonstrado em outros eucariotos, as plantas também utilizam fatores transcricionais que recrutam os complexos de remodelamento de cromatina para as regiões promotoras dos genes codificadores de proteínas. Um exemplo seria a regulação do gene da β-phaseolina (*phs*) em feijão (Li et al., 2001a). Esse gene é silenciado nos tecidos vegetativos devido ao posicionamento de um nucleossomo sobre o motivo TATA-box do seu promotor, tornando inacessível o TATA box para a proteína TBP. Quando ocorre a modificação da cromatina e o nucleossomo é deslocado o gene é altamente expresso durante o desenvolvimento da semente (Li et al., 1998). Essa modificação é resultante da presença do fator de transcrição semente específico, PvALF, que é membro da família ABI3/VP1 (Li et al., 1999).

Foram compilados no genoma de *Arabidopsis* (http://chromdb.biosci.arizona.edu/) 220 diferentes genes envolvidos na modelagem ou envolvidos com a cromatina, incluindo 22 homólogos do complexo remodelador de cromatina SWI2/SNF2, 12 histonas acetilases (HATs) entre outros, assim fica evidente que o remodelamento da cromatina também é importante no controle da expressão gênica em plantas (Riechmann, 2002).

Existe elevada conservação entre genes ortólogos nos eucariotos, além da conservação dos mecanismo moleculares de remodelagem da cromatina e da regulação gênica. A histona acetilase de milho do tipo RPD3 foi utilizada na complementação de um mutante homólogo em levedura (Rossi et al., 1998). O gene *BUSHI* (*BSH*) de *Arabidopsis* complementa parcialmente um mutante do gene *snf5* componente do complexo modelador de cromatina SWI/SNF de levedura (Brzeski et al., 1999). Homólogos em *Arabidopsis* do complexo CBP/p300 foram capazes de ativar a transcrição em cultura de células de mamíferos (Bordoli et al., 2001), além de outros genes tipo acetilases (Eshed et al., 1999; Ogas et al., 1999), remodeladores de cromatina dependentes de ATP (Ahringer, 2000).

Além destas similaridades, parece que novas características na regulação da expressão gênica mediada pela cromatina foram desenvolvidas nas plantas. Algumas histonas acetilases (classe HD2) parecem ser encontradas apenas em plantas (Dangl et al., 2001). Outros genes envolvidos na modelagem de cromatina também não foram detectados em genomas de outros eucariotos.

Este é o caso de MOM1, uma proteína relacionada ao complexo SWI/SNF (Amedeo et al., 2000).

As proteínas envolvidas na regulação da expressão gênica podem apresentar diferenças estruturais nos diferentes eucariotos, sendo algumas destas específicas de plantas, sendo este um exemplo de inovação evolucionária originada por exon "shuffling", deleções e inserções (Patthy, 2003; International Human Genome Sequencing Consortium, 2001). Em *Arabidopsis* por exemplo, proteínas tipo CBP/p300 não apresentam o bromodomínio e a região de ligação a CREB altamente conservadas em animais (Bordoli et al., 2001).

Outros genes de *Arabidopsis* envolvidos com cromatina têm sido funcional ou geneticamente caracterizados e mostraram toda importância do controle transcricional que a cromatina exerce nas plantas. A redução da atividade do gene *AtHD1* (uma histona deacetilase de *Arabidopsis*, HDAC) por RNA anti-senso causou alterações pleitrópicas no desenvolvimento, sugerindo um papel regulador global para este gene durante o desenvolvimento (Wu et al., 2000a). Outro gene de *Arabidopsis* tipo SWI/SNF2, *Decrease in DNA Methylation1* (*DDM1*), é requerido para manter os níveis de metilação do DNA e estabilizar o comportamento dos transposons (Singer et al., 2001).

Embora plantas e outros eucariotos apresentem um conteúdo similar de genes de fatores de transcrição, a organização das seqüências de DNA sobre as quais esses atuam podem ser diferentes. Geralmente, em animais as seqüências que regulam a expressão temporal e espacial se extendem por vários Kilobases (kb), já em plantas estas são curtas e se espalham por 1 kb em média. As compactas seqüências 5' promotoras de plantas, principalmente em Arabidopsis, repetem o mesmo padrão de expressão gênica nativa quando ensaiadas em plantas transgênicas. Mas nem sempre isto ocorre, pois muitas seqüências regulatórias podem ser encontradas em regiões "downstream" do sítio de iniciação da transcrição (Deyholos e Sieburth, 2000; Yu et al., 2001). Por exemplo, o segundo intron do gene *AGAMOUS* (*AG*), um MAD-box envolvido no desenvolvimento das flores, contém dois sítios de ligação para reguladores da sua transcrição, LFY e WUSCHEL (Deyholos e Sieburth, 2000; Lohmann et al., 2001).

Apesar de algumas diferenças estruturais nas regiões promotoras de eucariotos, a regulação da expressão gênica nos mesmos é decorrente de múltiplos *inputs*, tendo como vantagem a natureza combinatorial dos mecanismos eucarióticos de transcrição (Riechmann, 2002). Os estímulos

convergem através de diferentes elementos no promotor e que regula de forma coordenada à expressão do gene. Geralmente, os elementos regulatórios dos promotores são organizados de forma modular que podem ser particionadas nos diferentes e discretos sub-elementos que contém os diferentes sitos de ligação dos fatores transcricionais (Riechmann, 2001). Nas plantas a regulação da expressão gênica pelos sistemas de módulos e a ação sinérgica dos mesmos foi caracterizada com o promotor CaMV 35S (promotor do vírus mosaico de couve-flor). Esse promotor viral é ativo em praticamente todos os tecidos em plantas transgênicas, e pode ser dissecado em sub-domínios que conferem expressão tecido-específica (Benfey et al., 1990a e 1990b).

Essa função combinatória e sinérgica dos elementos do promotor eucariótico é acompanhada pelo modo combinatório de ação dos fatores de transcrição que se ligam aos diferentes elementos do promotor e permitem um mecanismo regulatório diversificado mesmo com um número limitado de fatores de transcrição e sítios alvos nos promotores (Riechmann, 2002). A presença dos diferentes elementos do promotor de forma adjacente é uma necessidade devido às interações entre os diferentes fatores de transcrição entre si e também, com o próprio aparato transcricional (Riechmann, 2002). Essas interações permitem o aumento no repertório regulatório e o aumento da especificidade dos mesmos. Desta forma um dado fator de transcrição poderia, ter um papel importante nas respostas a diferentes sinais. A proteína EmBP1 parece fazer parte de um complexo multiprotéico que inclui VP1 e GF14 que participa na resposta ao hormônio ABA (Shultz et al., 1998). VP1 ilustra um bom exemplo do controle combinado onde esta proteína pode apresentar funções de ativador ou repressor da transcrição dependendo do contexto apresentado pelo promotor. VP1 atua como um ativador da transcrição do gene *Em* em trigo (Shultz et al., 1998), enquanto que atua como repressor da expressão da α-amilase num promotor com contexto diferente (Hoecker et al., 1995).

2.7. Os reguladores de transcrição em eucariotos

Um regulador de transcrição típico é constituído por três módulos: um domínio de multimerização, um domínio de ligação ao DNA e um domínio de ativação. Isso permite que um conjunto relativamente pequeno de módulos gere reguladores de transcrição altamente especializados (Tjian e Maniatis, 1994).

O domínio de multimerização é responsável pela formação de dímeros, tetrâmeros ou outros multímeros. Muitos reguladores são ativos somente na forma multimérica, como, por exemplo, alguns tipos de bZIP, como AP-1, C-Jun e C-Fos (Wagner e Green, 1994).

Os domínios de interação com DNA e de ativação permitem aos reguladores de transcrição reconhecer seqüências alvo no DNA e interagir com os fatores basais do complexo de pré-iniciação, respectivamente (Riechmann, 2002).

A interação dos reguladores com fatores basais de transcrição pode resultar na estabilização ou desestabilização do complexo de pré-iniciação ou ainda alterar sua taxa de formação (Latchman, 1997).

Diferentes fatores, tais como o tipo celular, o meio ambiente e a fase do desenvolvimento, podem levar à síntese de um regulador ou gerar modificações pós-traducionais em um regulador pré-existente de modo a alterar sua atividade. Dessa maneira, os reguladores coordenam a expressão de diferentes genes em reposta aos diversos estímulos impostos (Gilmartin et al., 1990).

A fosforilação é uma das principais formas de modificação de um regulador de transcrição (Hunter e Karin, 1992; Hill e Treisman, 1995; Hunter, 1995). Ela pode modular a atividade de um regulador, alterando sua capacidade de interagir com o sítio alvo no DNA ou com o complexo de pré-iniciação, ou ainda regular sua localização nos diferentes compartimentos celulares (Gille et al., 1992; Gonzales e Montminy, 1989; Shua et al., 1993).

Um regulador de transcrição específico pode apresentar afinidades diferentes para diversos sítios no DNA, interagir com o DNA de maneira cooperativa com outros fatores ou ainda competir com outros fatores pelo sítio de ligação. Essa competição ocorre quando diferentes reguladores reconhecem o mesmo sítio, ou porque seus sítios se sobrepõem ou estão suficientemente próximos, impedindo a interação simultânea dos fatores com o DNA.

Os reguladores de transcrição podem ser agrupados em diferentes classes ou famílias cujos membros apresentam estruturas relacionadas. As classes estruturais de reguladores de transcrição melhor caracterizadas em plantas são: bZIP (*basic-leucine zipper*), bHLH (*basic helix-loop-helix*), MYB, proteínas contendo homeodomínio (HDs), MADS box e *Zinc Finger* (Kerstetter *et al.*, 1994; Ramachandran *et al.*, 1994; Takatsuji *et al.*, 1994; Meshi e Iwabushi, 1995; Purugganan *et al.*, 1995).

2.8. Os reguladores do tipo bZIP

Os reguladores do tipo bZIP são caracterizados pela presença da região bZIP. Tal região apresenta usualmente 60 a 80 resíduos, sendo subdividida em dois domínios. O primeiro, denominado zíper de leucinas, é caracterizado por repetições de resíduos de leucina a cada sete aminoácidos numa extensão de 30 a 40 resíduos, sendo responsável por interações entre bZIPs, promovendo assim dimerizações. O segundo, adjacente ao zíper de leucinas é denominado domínio básico, sendo constituído por aproximadamente 30 resíduos dos quais a maioria deles apresenta caráter básico. Tal domínio é responsável pela interação com o sítio-alvo do DNA (Figura 5).

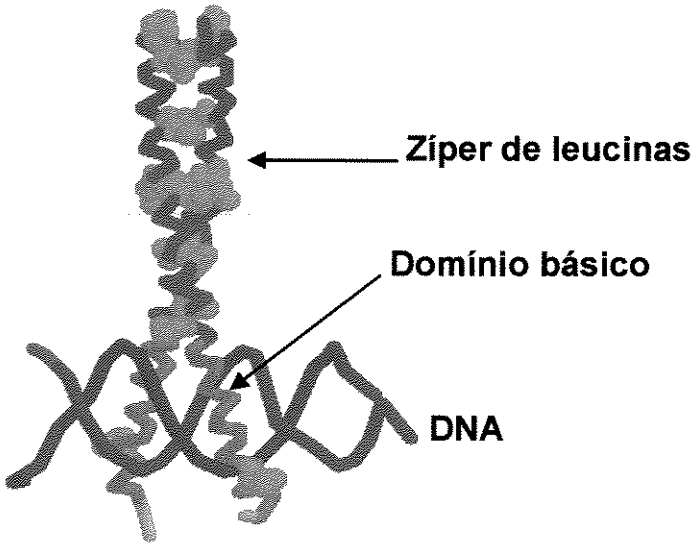


Figura 5 - Estrutura do dímero de GCN4 ligado ao DNA. A dimerização ocorre através do zíper de leucinas e a ligação ao sulco maior do DNA pelo domínio básico. Em vermelho está representado o zíper de leucinas, seguido pelo domínio básico em laranja. A posição amino-terminal está representada em amarelo e as posições das leucinas em verde. O DNA está representado em azul (Ellenberger *et al.* 1992).

As propriedades estruturais das proteínas do tipo bZIP foram verificadas através estudos cristalográficos realizados inicialmente com GCN4, um ativador transcricional de fungos que pertence a essa classe de proteínas (O'Shea et al., 1991; Keller et al., 1995). Esses estudos demonstraram que os monômeros protéicos formam α-hélices contínuas (Figura 5). No domínio

zíper de leucinas, as hélices são torcidas, produzindo uma estrutura do tipo *"coiled-coil"* ou superespiraladas.

As estruturas primárias das regiões protéicas que formam *"coiled-coil"* são caracterizadas pela repetição de seqüências de sete resíduos de aminoácidos correspondendo a duas voltas na α-hélice. Tais resíduos são descritos de "a" a "g" de acordo com sua posição na hélice (Figura 6).

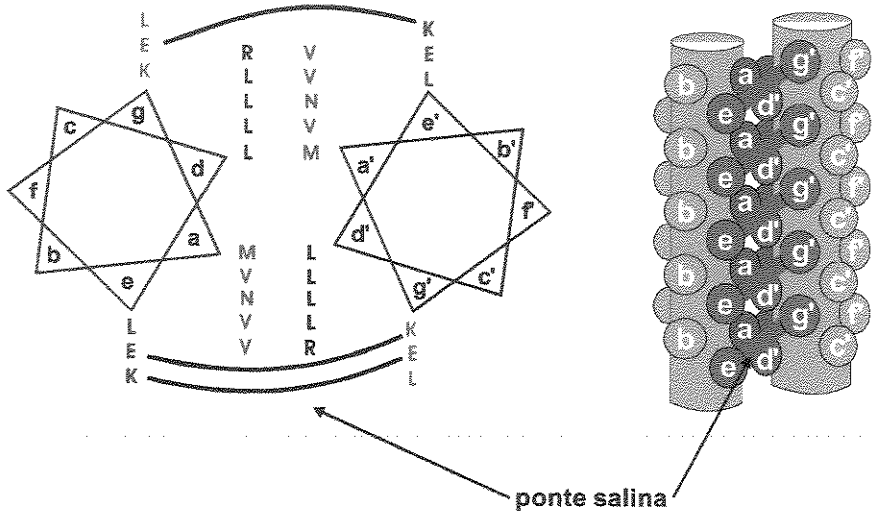


Figura 6 - Representação esquemática das α-hélices referentes ao zíper de leucinas do regulador GCN4. Os resíduos das posições "a" estão representados em verde, "d" em azul, "e" em vermelho e "g" em rosa. Em **A** as α-hélices estão representadas na forma de estrelas de sete pontas, ressaltando a posição de cada resíduo. As atrações eletrostáticas do tipo ponte salina estão representadas pelas linhas em azul. Em **B** as α-hélices estão representadas por barras e as posições das cadeias laterais de cada resíduo estão representadas por esferas.

Os resíduos presentes nas posições "e" e "g" são geralmente formados por aminoácidos carregados capazes de formar interações do tipo ponte salina entre duas subunidades de α-hélice. Os resíduos das posições "a" e "d" por sua vez são quase sempre aminoácidos apolares que formam interações hidrofóbicas entre as α-hélices. (Sodek *et al.*, 1972).

As interações do tipo ponte salina entre os resíduos "e" e "g" possuem grande raio de ação, porém não promovem o correto direcionamento das moléculas. As interações hidrofóbicas entre os resíduos "a" e "d" possuem raio de ação bem menor que as pontes salinas, porém são capazes

de orientar corretamente as α-hélices, determinando também a especificidade da associação de subunidades iguais ou diferentes (Lupas, 1996).

Assim, a dimerização entre bZIPs pode ocorrer entre subunidades idênticas, formando homodímeros, ou subunidades diferentes, formando heterodímeros (O'Shea *et al.*, 1992). Algumas b-ZIPs podem formar apenas homodímeros (GCN4), enquanto outras formam apenas heterodímeros (FOS). Existem, porém b-ZIPs que apresentam capacidade de formar tanto homodímeros quanto heterodímeros (JUN) (Smeal *et al.*, 1989; Kouzarides & Ziff, 1989).

A formação de homo ou heterodímeros pode constituir-se em um mecanismo de controle de expressão gênica (Chiu *et al.*, 1989; Schutte *et al.*, 1989).

Apesar dos reguladores do tipo bZIP apresentarem um domínio relativamente simples de ligação com o DNA, eles são capazes de reconhecer um grande número de seqüências de DNA discriminando-as suficientemente a fim de regular a transcrição de diversos genes com diferentes promotores (Riechmann, 2002).

2.9. bZIPs em plantas

Em *Arabidopsis* as bZIPs representam uma das três maiores famílias de fatores de transcrição, juntamente com os fatores de transcrição MYBs e MADS (Riechamann e Ratcliff, 2000).
As bZIPs de plantas participam de uma ampla gama de processos, tais como desenvolvimento de semente (Ciceri et al., 2000), fotomorfogênese (Osterlund et al., 2000), defesa contra patógenos (Niggeweg et al., 2000), desenvolvimento de órgãos (Chuang et al., 1999), sinalização de sacarose (Rook et al., 1998), resposta a hormônios (Choi et al., 2000; Filkelstein and Lynch, 2000; Uno et al., 2000; Niggeweg et al., 2000) e alongamento celular (Yin et al., 1997).

Até 1998 cerca de 50 bZIPs de plantas haviam sido identificadas e classificadas de acordo com similaridade da região bZIP em 5 diferentes famílias (Vettore *et al.*, 1998). Posteriormente o seqüenciamento do genoma de *Arabidopsis* permitiu o acesso ao conjunto completo de genes que codificam bZIPs nessa planta, com o qual elaboramos uma classificação mais abrangente destes fatores através de análise filogenética (Vincentz *et al.*, 2001).

As bZIPs de plantas mostram afinidade para seqüências de DNA com um core ACGT (Foster et al., 1994; Izawa et al., 1994). Três tipos principais destes elementos foram caracterizados, os G -,

C - e A - boxes (Izawa et al., 1994). Os G-box (ACGTG) foram mais caracterizados porque fazem parte do contexto de promotores de vários genes de plantas (Williams et al., 1992). Vários estudos mostraram que as bZIPs apresentam preferências em relação as bases que flanqueiam o core ACGT e estas definem a afinidade e especificidade da interação das bZIPs e o DNA (Williams et al., 1992; Foster et al., 1994; Izawa et al., 1994).

Existem muitas evidências de que os fatores de transcrição bZIPs participam dos processos de regulação da expressão gênica em plantas interagindo com fatores basais da expressão e outros tipos de fatores de transcrição (Singh, 1998; Fan e Dong, 2002). Zhu e colaboradores, (2002), demonstraram a interação entre TBP (TATA Binding Protein) e a bZIP RF2a em arroz. Essa interação levou a um aumento na transcrição promovida pelo promotor do vírus do arroz, *Tungro bacilliform*. Outro exemplo de regulação combinada é a interação dos fatores de transcrição Dof com bZIPs que estimula a ligação das bZIPs nos promotores (Chen et al., 1996). No promotor do gene da glutationa-S-transferase 6 (*GST6*) existem vários sítios de ligação das proteínas Dof que estão perto dos elementos ocs, que são sítios de ligação de bZIPs. Há evidências de que bZIPs e Dof participam da regulação da expressão dos genes das proteínas de reserva das sementes de milho (Vicente-Carbajosa et al., 1997). A bZIP TGA2 participa da resposta mediada por ácido salicílico e interage com o fator NPR1 (nonexpresser PR gene) na regulação da expressão dos genes PR (*pathogenesis related genes*) (Fan e Dong, 2002).

Esses estudos demonstram que as bZIPs, uma das maiores famílias de fatores de transcrição presentes nos genomas das plantas tem papel fundamental nos mais variados processos fisiológicos.

3. Apresentação dos artigos

Esta tese é constituída de dois artigos publicados, um artigo aceito para publicação e outro em preparação. No primeiro artigo publicado propõem-se uma classificação filogenética para os fatores de transcrição bZIPs de *Arabidopsis* e discute-se como esta classificação ajudou na identificação e organização dos cDNAs de cana-de-açúcar codificando bZIPs que foram seqüenciados no projeto SUCEST.

O segundo artigo publicado versa a evolução da família de proteínas homólogas a Opaco 2 de milho, bem como discute a conservação de motivos estruturais importantes dos membros desta família, e a formação dos grupos de ortólogos entre as plantas dicotiledôneas e monocotiledôneas.

O artigo 3 aceito para publicação na revista Plant Science relata sobre a clonagem e caracterização de um cDNA de cana-de-açúcar codificando uma bZIP.

O manuscrito em preparação mostra como as bZIPs de cana-de-açúcar são expressas nos diferentes estágios de desenvolvimento e quando as plantas de cana-de-açúcar foram tratadas com os hormônios ácido abscísico e metil-jasmonato.

3.1. Artigos

1- Phylogenetic relationships between *Arabidopsis* and sugarcane bZIP transcriptional regulatory factors.

Michel Vincentz, Paulo S. Schlögl, Luis Gustavo G. Corrêa, Fabiana Kühne and Adilson Leite. (2001). **Genetic and Molecular Biology 24**, 55-60.

2- Evolutionary pattern of angiosperm bZIP factor homologous to the maize Opaque 2 regulatory protein.

Michel Vincentz, Cláudia Bandeira Kobarg, Luciane Gauer, Paulo S. Schlögl and Adilson Leite. (2003). **J Mol Evol. 56**(1), 105-16.

3- Expression, purification and characterization of a novel bZIP protein from sugarcane.

Paulo Sérgio Schlögl, Jörg Kobarg , Victor Hugo Moreau, Adilson Leite, Adão A. Sabino, Marcos N. Eberlin and Marcelo Menossi.

No prelo na revista Plant Science.

4- Expression profiles of bZIP in sugarcane.

Paulo Sérgio Schlögl , Fábio Tebaldi Nogueira , Rodrigo Drummond, Juliana Felix, Vicente E. De Rosa Jr., Adilson Leite , Eugênio C. Ulian, and Marcelo Menossi.

Manuscrito a ser submetido.

Genetics and Molecular Biology, 24 (1-4), 55-60 (2001)

Phylogenetic relationships between *Arabidopsis* and sugarcane bZIP transcriptional regulatory factors

Michel Vincentz [1,2], Paulo S. Schlögl[1], Luis Gustavo G. Corrêa[1], Fabiana Kühne[1] and Adilson Leite[1]*

Abstract

We built a complete and non-redundant database of bZIP transcriptional regulatory factors from the *Arabidopsis* reference genome. These *Arabidopsis* bZIP factors were ordered into thirteen families of evolutionary related proteins and this classification was used to identify and organize sugarcane cDNAs encoding bZIP proteins. We also show how this classification should help in defining putative clusters of orthologous groups of higher plant bZIP regulators and briefly discuss the expected benefits of this procedure to efficiently characterize sugarcane bZIP transcriptional regulators.

INTRODUCTION

Growth and development of all organisms largely relies on appropriate regulation of gene expression. Differential gene expression mainly occurs through the control of transcription initiation rates by transcriptional regulatory factors. These factors are usually defined as sequence-specific DNA binding proteins that recognize regulatory sequences in the promoter of a gene and are capable of modulating transcription (Holstege and Young, 1999; Kornberg, 1999 and Singh, 1998). Transcriptional regulators can be grouped into families (or super families) of related proteins according to the structural or primary sequence similarities of their DNA binding domain (Riechmann *et al.*, 2000; Wingender *et al.*, 2000).

The basic leucine zipper (bZIP) transcriptional regulatory factors have been described in all eukaryotes. Their DNA binding domain consists of a region rich in basic amino acids that binds to DNA and a so-called leucine zipper that consists of several heptad repeats of hydrophobic residues and which causes dimerization. The X-ray structure of the yeast GCN4 bZIP domain complexed to DNA target sites has shown that the bZIP is completely α-helical in structure. The two leucine zippers are packed in a coiled-coil structure for dimerization, while the basic regions of the dimer fits into the major groove of the half-sites of the target DNA (Hurst, 1995).

Genetic, molecular and biochemical studies indicate that the bZIP factors of higher plants are important regulators of plant specific processes such as fotomorphogenesis (Osterlund *et al.*, 2000); organ development (Walsh *et al.*, 1997; Chuang *et al.*, 1999); cell elongation and morphogenesis (Yin *et al.*, 1997; Fukazawa *et al.*, 2000); control of nitrogen to carbon balance during seed development (Ciceri *et al.*, 1999); defense mechanisms (Niggeweg *et al.*, 2000; Zhang *et al.*, 1999); sucrose signalling (Rook *et al.*, 1998) and the response to hormones (Choi *et al.*, 2000; Finkelstein et Lynch, 2000; Uno *et al.*, 2000; Niggeweg *et al.*, 2000) and light (Schindler *et al.*, 1992; Wellmer *et al.*, 1999).

With the sequencing of the *Arabidopsis thaliana* (*Arabidopsis*) genome, a possible complete higher plant gene index was described (The *Arabidopsis* Genome Initiative, 2000). This repertoire of genes is likely to be representative of all higher plant genes that carry out essential functions and it therefore constitutes a invaluable reference data set which will help to better understand the evolution of cellular and developmental processes of higher plants.

Within this context, we initiated a comprehensive characterization of higher plant bZIP factors and we describe here, the generation of a probable complete and non redundant set of 72 bZIP factors encoded by the reference *Arabidopsis* genome (see also Riechmann *et al.*, 2000). A phylogenetic classification of this set of factors was established using conditions that were used previously to assess the phylogenetic relationships of 50 higher plant bZIP factors (Vettore *et al.*, 1998). We show how this classification has allowed us to efficiently characterize sugarcane expressed sequence tags (ESTs) encoding bZIP proteins and illustrate how this classification can be used to identify putative clusters of orthologous groups of higher plant bZIP factors including sugarcane bZIP genes. It is expected that defining such clusters should be useful in rationalizing the systematic characterization of higher plant bZIP proteins and more specifically sugarcane bZIPs.

[1]*Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, 13081-970 Campinas, SP, Brazil.*
[2]*Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, 13081-970 Campinas, SP, Brazil.*
Send correspondence to Michel Vincentz. E-mail: mgavince@obelix.unicamp.br.

RESULTS AND DISCUSSION

Phylogenetic classification of *Arabidopsis* bZIP transcriptional regulatory factors

A complete and non-redundant set of *Arabidopsis* bZIP factors was built from the NCBI GenBank and protein databases and MIPS MATDB accessions. The amino acid sequences of the bZIP domain of four accessions were further edited based on amino acids sequences alignments (BAB02051; AAD23721; T06089 and AAF67360) and one new putative bZIP protein not yet annotated at MATDB or GeneBank was identified (At2gBZN). Three proteins with a truncated basic region or leucine zipper were not included in our database, the total number of proteins in our database being 72.

The evolutionary relationships between the members of our *Arabidopsis* bZIP proteins collection was evaluated by phylogenetic analysis of the aligned amino acids sequences of their bZIP domain (Figure 1). The unrooted tree inferred from neighbor-joining analysis of the bZIP domain data set is shown in Figure 2. Based on the branching pattern, the tree was resolved into thirteen families. Most of the families show moderate to strong bootstrap support. Concerning families VI and VII, which are poorly resolved, we noticed that all members of these two families, as well as the genes of families IV and V, form a group of bZIP genes without introns. We also noticed that all members of several families share partially identical exon-intron gene organization (data not shown), supporting the pattern of clustering defined here. Finally, the bZIP protein AAG51519 does not fit into any of the Familie, although we included it into Family X based on its blastp best hit with proteins of Family X.

Index of sugarcane bZIP factors

The ordered set of *Arabidopsis* bZIP regulators was used to efficiently detect and classify sugarcane contigs encoding bZIP transcriptional regulators. In a first step, one or two query sequences consisting of full-length protein sequence of each of the 13 *Arabidopsis* bZIP families (Figure 2) were utilized to screen the SUCEST database, candidate sugarcane contigs being selected based on the presence of at least one conserved protein motif among several members of each *Arabidopsis* bZIP family. In a second step, selected sugarcane contigs were included into one of the *Arabidopsis* families according to their blastp best hit. Our strategy allowed us to identify 121 sugarcane contigs encoding candidate bZIP transcription factors. The pattern of distribution of the sugarcane contigs among the 13 *Arabidopsis* families is shown in Figure 3. No sugarcane contig related to Families IV and XIII were detected. The interpretation of this pattern is not straightforward but we suggest that it may reflect the number of genes included in each *Arabidopsis* family and/or the expression level of sugarcane genes related to each of these families.

Basic Region — Leucine Zipper

KRQKRKQSNRESARRSRLRKQAECEQLQQRVESLSNENQSLRDE

Figure 1 - Amino acid sequence of the bZIP domain used in the phylogenetic analysis of *Arabidopsis* bZIP factors. The sequence is from the *Arabidopsis* GBF1 regulator. Leucines of the leucine zipper are underlined.

Putative clusters of orthologous groups of monocot and dicot bZIP factors

To further characterize the sugarcane bZIP factors we initiated a comparative analysis to identify Putative Clusters of Orthologous Groups (PCOG) of higher plants bZIP factors. A Cluster of Orthologous Group (COG) consists of individual orthologous genes or orthologous groups of paralogs from several completely sequenced genomes (Tatusov *et al.*, 1997). The term ortholog refers to homologous genes that have been created by a speciation event, *i.e.* are versions of the same gene in different organisms, and paralogs are homologous genes that result from a duplication event within a genome (Tatusov *et al.*,1997 and Thornton and DeSalle, 2000). Orthologs usually retain the same function, whereas paralogs can explore new functions. An important consequence of defining COGs is that it allows to predict with some confidence the structure and function of uncharacterized members of the COG.

To detect PCOGs of bZIP factors of higher plants, we built a data set consisting of all monocot and dicot bZIP protein sequences avalaible in GenBank plus the reference database formed by the 13 *Arabidopsis* bZIP families (Figure 2). The neighbor-joining distance method (Saitou and Nei, 1987) was used to identify the PCOGs. Several of the situations we encountered are illustrated in Figure 4. A simple PCOG consisting of individual putative orthologs which includes the maize regulator *Liguleless2* is shown in Figure 4A. The simplest interpretation of this PCOG is that the *Arabidopsis* AAF22906 and the sugarcane II.8 proteins are functionally related to the maize regulator *Liguleless2* involved in maize leaf development (Walsh *et al.*, 1997).

Several PCOGs with more complex relationships between members are shown in Figures 4B and 4C. For instance, PCOG 1 of family XII (Figure 4C) can be described as an orthologous group of two *Arabidopsis* and two sugarcane paralogs. Cluster 1 of family VIII (Figure 4B) is even more complex. It consists of a putative group of *Arabidopsis* /monocot orthologs (*Arabidopsis* AAF67360, maize OHP1, rice REB, barley BLZ1 and the sugarcane VIII.4 proteins) and one group of monocot orthologs (maize, *Coix* and *Sorghum* Opaque2 regulators).

We noticed that some *Arabidopsis* bZIP factors are encoded by genes that are part of two co-linear genomic sequences formed by several highly similar genes. Such proteins are therefore likely to be paralogs that originated with the large-scale chromosomal duplications that formed the

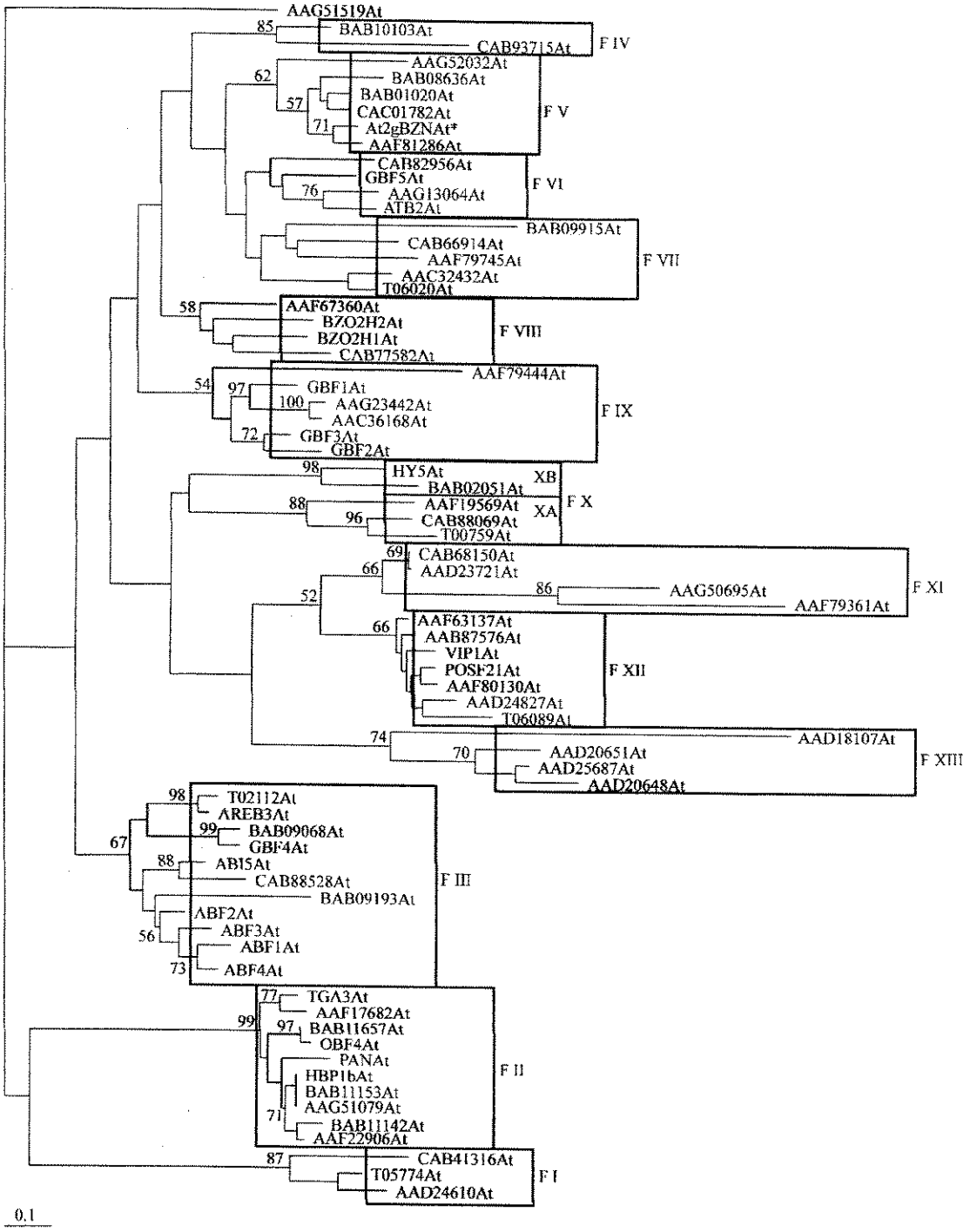


Figure 2 - Phylogenetic relationships of the complete set of *Arabidopsis* bZIP transcription factors. Unrooted tree of the *Arabidopsis* bZIP factors inferred from a neighbor-joining analysis of bZIP domain amino acid sequences data set (Figure 1). The tree was organized into thirteen families (F I to F XIII). Bootstrap of 1000 replicates is indicated as percentages along the branches when higher than 50%. In most case the proteins are identified by the accession number. Accession numbers of proteins with a name are given in Materials and Methods. At2gBZNAt* is a bZIP protein not yet annotated. The scale bar corresponds to 0.1 estimated amino acid substitution per site.

Arabidopsis genome (The *Arabidopsis* Genome Initiative, 2000; Vision *et al.*, 2001). For instance, POSF21 and AAF80130 in PCOG 3 of Family XII (Figure 4C), are encoded by genes that are part of two co-linear segments of at least six genes on chromosome II and I, respectively (Result not shown). These two Arabidopsis bZIPs paralogs are

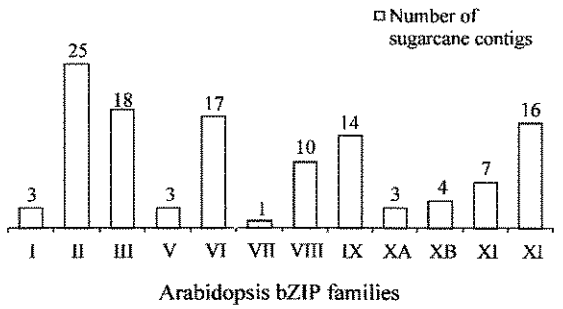


Figure 3 - Distribution of sugarcane contigs encoding bZIP factors among the thirteen Arabidopsis Families defined in Figure 2.

closely related to the rice RF2A that seems to be important for differentiation of leaf cells (Yin *et al.*, 1997). It remains to be shown whether or not that they are functionally related to RF2A and also to what extent they are redundant.

The polyploid origin of the sugarcane genome (Daniels and Roach, 1987) may prevent us distinguishing sugarcane paralogs from allelic forms of the same locus. However, this complexity should not hamper our ability to reach reasonable conclusions about the clustering pattern and functional inference. For example, it is difficult to infer whether or not the two sugarcane contigs XII.4 and XII.6 in PCOG 2 (Figure 4C) are two alleles of the same gene or not, while contig XII.7 could be a corresponding paralog (Figure 4C). However, a clear orthologous relationships between these three sugarcane bZIP proteins and the *Arabidopsis* protein VIP1 can be proposed (Figure 4C).

Based on the strategy described in this paper, we are now organizing all higher plant bZIP factors into PCOGs and hope to use this information to further characterize sugarcane bZIP transcriptional regulators.

MATERIALS AND METHODS

The non redundant data set of *Arabidopsis* bZIP factors was obtained through iterated searches of the GenBank and protein database at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) and the Munich information center for protein sequences (MIPS) *Arabidopsis thaliana* database (MATDB, http://www.mips.biochem.mpg.de/proj/thal/) using different known bZIP query sequences and the blastp and tblastn programs (Altschul *et al.*, 1990) at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and the MIPS (http://mips.gsf.de/ proj/thal/db/search/search_frame.html) servers. Additionally, with the recent publication of the *Arabidopsis* genome (The *Arabidopsis* Genome Initiative, 2000), a key word search was also performed at MATDB (v211200).

Sugarcane contigs (a contig or cluster is a consensus sequence derived from several overlapping and highly similar ESTs sequences) coding for bZIP proteins were detected by using *Arabidopsis* full-length bZIP protein sequence as query sequences to screen the SUCEST (sugar-

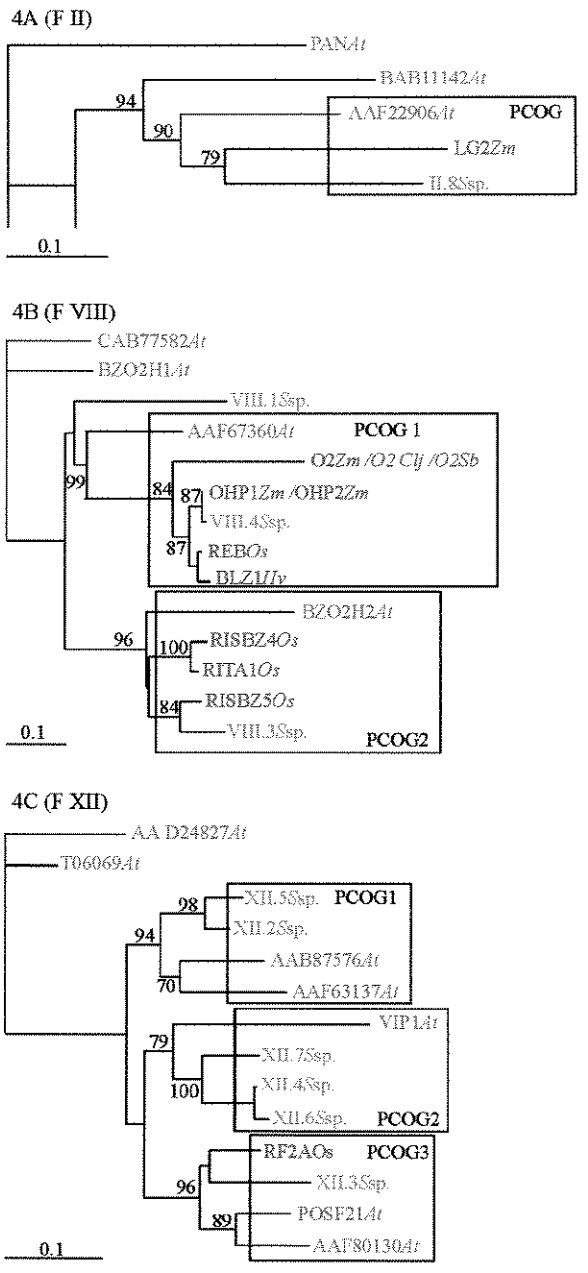


Figure 4 - Identification of Putative Clusters of Orthologous Groups of monocot and dicot bZIP factors (PCOG). Aligned amino acid sequence data set of monocot and dicot bZIP proteins belonging to the reference Arabidopsis Family II (F II; 4A), Family VIII (F VIII; 4B) and XII (F XII; 4C) were built and the corresponding unrooted trees inferred from a neighbor-joining analysis of these data sets are shown. Only part of the tree corresponding to Family II is shown (4A). Arabidopsis proteins are shown in blue, monocot proteins are shown in red and sugarcane proteins are shown in green. Accession numbers are given in Materials and Methods. Species abbreviations are: *At*, *Arabidopsis thaliana*; *Clj*, *Coix lacryma-jobi*; *Hv*, *Hordeum vulgare*; *Os*, *Oryza sativa*; *Sb*, *Sorghum bicolor*; *Ssp*, *Saccharum* sp. and *Zm*, *Zea mais*. Bootstrap support is indicated as percentage of 1000 replicates. The scale bar corresponds to 0.1 estimated amino acid substitution per site.

cane EST genome project) database (http://www.sucest.lbi.dcc) with the locally available tblastn program. Sugarcane contigs consists of one to several overlapping and highly similar EST reads assembled with the PHRAP program (P.Green, htpp://bozeman.mbt.washington.edu/phrap.docs/phrap.html; -penalty -15 -bandwidth 14 -minscore 100 -shatter_greedy, Meidanis personnal communication).

Protein sequences were aligned with the CLUSTALX program (Thompson *et al.*, 1997). Amino acid sequence data was analyzed by the neighbor-joining method (Saitou and Nei, 1987) using the NEIGHBOR program (PHYLIP, Phylogeny Inference Package version 3.57c; Felsenstein, 1993) and PAM distances (Dayhoff *et al.*, 1978), obtained with the PRODIST program (PHYLIP). Bootstrap assessment of tree topology in neighbor-joining analysis was performed with the SEQBOOT program (PHYLIP). Trees were displayed with the TREEVIEW program (Page, 1996). DNA sequence analysis was carried out with the DNASIS program (Pharmacia). Motifs conserved among members of each *Arabidopsis* bZIP family (Figure 2) were detected with the help of the MEME program (Bailey and Elkan, 1994; http://meme.sdsc.edu/meme/website/).

The accession numbers of the *Arabidopsis* bZIP proteins shown in Figure 2 are: Family II: PAN, AAD49979; OBF4, CAA49524; HBP1b, BAB11154; TGA3, S46523. Family III: ABF4 (AREB2), AAF27182; ABF3, AAF27181; ABF1, AAF27179; ABF2 (AREB1), AAF27180; AREB3, BAB12406; GBF4, P42777; ABI5, AAD21438. Family VI: GBF5, AAG17474; ATB2,T05279. Family VIII: BZO2H1, AAG25727; BZO2H2, AAG25728. Family IX: GBF1, P42774; GBF2, P42775; GBF3, P42776. Family X: HY5, BAA21116. Family XII: VIP1, AAF37279; POSF21, AAD26486. The accession numbers of bZIP factors shown in Figure 4 are: LG2*Zm* (*liguleless2*), AAC39351; RF2A*Os*, AAC49832; O2*Clj*, S42493; O2*Sb*, CAA50642; O2*Zm*, A34800; OHP1*Zm*, JQ2147; OHP2*Zm*, JQ2148; REB*Os*, BAA36492; BLZ1*Hv*, T04477; RITA1*Os*, T03990. The accession numbers for RISBZ4*Os* and RISBZ5*Os* sequences are not yet available but can be found in Onodera *et al.* (2001). The accession numbers at SUCEST for the Sugarcane contigs (clusters) shown in Figure 4 are: II.8, SCCCCL3001H09.g; VIII.1, SCVPRZ3030A04.g; VIII.3, SCCCCL4005C09.g; VIII.4, SCQSRT1036D12.g; XII.2, SCCCRT2002D06.g; XII.3, SCJFRZ2034D03.g; XII.4, SCAGHR1016H08.g; XII.5, SCSFLR2016F09.g; XII.6, SCACRZ3034F03.g; XII.7, SCJLAM1064H01.g.

ACKNOWLEDGMENTS

This work was supported by grant from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) / Auxílio à Pesquisa Nº 1999/02839-9. PSS and LGGC are supported by grants from FAPESP. We thank an anonymous referee for helpful comments.

NOTE ADDED IN PROOF

Since we submitted this article for publication the protein At2gBZN in Figure 1 was annotated as At2g04038 at MATDB.

RESUMO

Construímos um banco de referência não redundante de fatores de regulação da transcrição do tipo bZIP a partir de dados do genôma de *Arabidopsis thaliana*. Os fatores bZIP de *Arabidopsis* foram ordenados em treze famílias de proteínas evolutivamente relacionadas e essa classificação foi usada para organizar os cDNAs de cana de açúcar que codificam proteínas bZIP. Além disso, mostramos que essa classificação poderá ser útil para definir "Putative Clusters of Orthologous Groups" de reguladores bZIP de plantas superiores.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and **Lipman, D.** (1990). Basic Local Alignment Tool. *J. Mol. Biol. 215*: 403-410.

Bailey, T.L. and **Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28-36, AAAI Press, Menlo Park, California.

Choi, H., Hong, J., Há, J., Kang, J. and **Kim, S.Y.** (2000). ABFs, a family of ABA-responsive element binding factors. *The Journal Biol. Chem. 275*: 1723-1730.

Chuang, C-F., Running, M.P., Williams, R.W. and **Meyerowitz, E.M.** (1999). The Perianthia gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. *Genes and Development 13*: 334-344.

Ciceri, P., Locatelli, F., Genga, A., Viotti, A. and **Schmidt, R.J.** (1999). The activity of the maize Opaque-2 transcriptional activator is regulated diurnally. *Plant Physiol. 121*: 1321-1327.

Daniels, J. and **Roach, B.T.** (1987). Taxonomy and evolution. In *Sugarcane improvement through breeding* (Heinz, D.J., ed.). Elsevier Press, Amsterdam, pp. 7-87.

Dayhoff, M.O., Schwartz, R.M. and **Orcutt, B.C.** (1978). In *Atlas of protein sequence and structure*, v. 5, Suppl. 3 (Dayhoff, M.O., ed.). National Biochemical Research Foundation, Silver Spring, MD: 345-352.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA.

Finkelstein, R.R. and **Lynch, T.J.** (2000). The *Arabidopsis* Abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *The Plant Cell 12*: 599-609.

Fukazawa, J., Sakai, T., Ishida, S., Yamaguchi, I., Kamiya, Y. and **Takahashi, Y.** (2000). Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *The Plant Cell 12*: 901-915.

Holstege, F.C.P. and **Young, R.A.** (1999). Transcriptional regulation: contending with complexity. *Proc. Natl. Acad. Sci. USA 96*: 2-4.

Hurst, H. (1995). Transcription factor 1: bZIP proteins. *Protein Profile 2*: 105-168.

Kornberg, R.D. (1999). Eukayotic transcriptional control. *Trends in Genetics 15*: 46-49.

Niggeweg, R., Thurow, C., Kegler, C. and **Gatz, C.** (2000). Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid-and auxin-inducible expression of as-1-containing target promoters. *The Journal Biol. Chem. 275*: 19897-19905.

Osterlund, M.T., Wei, N. and **Deng, X.W.** (2000). The roles of photoreceptor system and the COP1-targeted destabilization of HY5 in light control of *Arabidopsis* seedling development. *Plant Physiol. 124*: 1520-1524.

Page, R.D.M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences 12*: 357-358.

Riechmann, J.L., Heard, J., Martin, J., Reuber, F., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghanderhari, D., Sherman, B.K. and **Yu, G.-L.** (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science 290*: 2105-2110.

Rook, F., Gerrits, N., Kortstee, A., van Kampen, M., Borrias, M., Weisbeek, P and **Smeekens, S.** (1998). Sucrose-specific signalling represses translation of the *Arabidopsis* ATB2 bZIP transcription factor gene. *The Plant Journal 15*: 253-263.

Saitou, N. and **Nei, M.** (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*: 406-425.

Schindler, U., Menkens, A.E., Beckmann, H., Ecker, J.R. and **Cashmore, A.R.** (1992). Heterodimerization between light-regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins. *EMBO J. 11*: 1261-1273.

Singh, K.B. (1998). transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol. 118*: 1111-1120.

Tatusov, R.L., Koonin, E.V. and **Lipman, D.J.** (1997). A genomic perspective on protein families. *Science 278*: 631-637.

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

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and **Higgins, D.G.** (1997). The CLUSTALX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. *Nucl. Acids Res. 25*: 4876-4882.

Thornton, J. and **DeSalle, R.** (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet. 1*: 41-73.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and **Yamaguchi-Shinozaki, K.** (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. USA 97*: 11632-11637.

Vettore, A.L., Yunes, J.A., Cord Neto, G., da Silva, M.J., Arruda, P. and **Leite, A.** (1998). The molecular and functional characterization of an Opaque2 homologue gene from *Coix* and a new classification of plant bZIP proteins. *Plant Mol. Biol. 36*: 249-263.

Vision, T.J., Brown, D.G. and **Tanksley, S.D.** (2001). The origin of genomic duplications in *Arabidopsis*. *Sience 290*: 2114-2117

Walsh, J., Waters, C.A. and **Freeling, M.** (1997). The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. *Genes and Development 11*: 208-218.

Wellmer, F., Kircher, S., Rügner, A., Frohmeyer, H., Schäfer, E. and **Harter, K.** (1999). Phosphorylation of the parsley bZIP transcription factor CPRF2 is regulated by light. *The Journal of Biol. Chem. 274*: 29476-29482.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I. and **Matys, V.** (2000). TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res. 28*: 316-319.

Yin, Y., Zhu, Q., Da, S., Lamb, C. and **Beachy, R.** (1997). RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, function in vascular development. *EMBO J. 16:* 5247-5259.

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

J Mol Evol (2003) 56:105–116
DOI: 10.1007/s00239-002-2386-1

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary Pattern of Angiosperm bZIP Factors Homologous to the Maize Opaque2 Regulatory Protein

Michel Vincentz,[1,2] Claudia Bandeira-Kobarg,[1] Luciane Gauer,[1] Paulo Schlögl,[1] Adilson Leite[1]

[1] Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Cidade Universitaria "Zeferino Vaz," Distrito Barão Geraldo, 13081-970, Campinas, SP, Brazil
[2] Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, Cidade Universitaria "Zeferino Vaz," Distrito Barão Geraldo, 13081-970, Campinas, SP, Brazil

Received: 3 July 2002 / Accepted: 14 August 2002

Abstract. Opaque2 (O2) is a bZIP transcriptional regulatory factor involved in the control of seed storage proteins synthesis as well as carbon and nitrogen metabolism during maize seed development. Phylogenetic analysis of a possible complete and nonredundant collection of angiosperm bZIP factors resulted in the identification of 20 angiosperm O2-homologues that defined what we call the O2 gene family. Members of the family share a highly conserved bZIP DNA binding domain and several other motifs which define important functional features. The O2 family was enriched by the identification of 25 new putative angiosperm *O2* homologous genes in EST databases and in the rice genome. Based on parsimony analysis, the collection of *O2* homologues was organized into one eudicot–monocot and three monocot groups of orthologous genes and two groups of eudicot genes. These results support a model of the evolution of the O2 family that involves two *O2* homologous gene duplications before the separation of monocots and eudicots. Further expansion of *O2* homologues resulted in at least three and one gene duplications in the monocot and eudicot lineages, respectively. *O2* appears to have been the result of a monocot-specific gene duplication event, and the possibility that *O2* represents a functional specialization restricted to monocots is suggested.

Key words: Angiosperm — *Arabidopsis thaliana* — bZIP factors — Opaque2 — Phylogeny

Introduction

Transcriptional regulatory factors of the basic leucine zipper (bZIP) class have been described in all eukaryotes (Wingender et al. 2000). These factors bind DNA as dimers through a conserved DNA binding domain. This domain is formed by a region rich in basic amino acids that interact with the DNA target site and by a zipper of leucines that consists of several heptad repeats of hydrophobic residues that promote dimerization (Hurst 1995). In angiosperms (flowering plants), Opaque2 (O2) is one of the best-characterized bZIP transcriptional regulators. *O2* is an important regulatory locus of seed endosperm development of the monocot species *Zea mays* (maize) (Schmidt 1993). O2 specifically accumulates in the developing endosperm, where it activates the expression of α- and β-prolamin storage protein genes (Schmidt et al. 1992; Vicente-Carbajosa et al. 1997; Cord Neto et al. 1995), the b-32 albumin genes (Lohmer et al. 1991), and the gene for the cytoplasmic pyruvate orthophosphate dikinase (Gallusci et al. 1996). *O2* is also involved in the control of lysine accumulation (Kemper et al. 1999) and threonine metabolism

Correspondence to: Michel Vincentz; *email:* mgavince@obelix.unicamp.br

(Damerval and Le Guilloux 1998). Overall, these results point to a role for *O2* in the coordinated regulation of storage protein synthesis as well as carbon and nitrogen metabolism during seed development. The recent suggestion that the diurnal modulation of O2 DNA binding activity by phosphorylation/dephosphorylation is regulated by diurnal metabolic fluxes is in line with the latter view (Ciceri et al. 1999).

Identifying O2-related bZIP factors among angiosperms and understanding their evolution are of interest for two reasons: first, because O2 plays an important role in seed development and, second, because O2 has been extensively studied. A phylogenetic analysis using the amino acid sequences of the highly conserved bZIP domain of 50 angiosperm bZIP proteins identified a cluster of 8 monocot and eudicot O2-related proteins, which may form a gene family (Vettore et al. 1998). Although the probable *O2* orthologues from the close relatives of maize, sorghum and *Coix*, are known, the orthologous/paralogous relationships among all the O2-related proteins which are included in this putative family are unclear. More precisely, the existence of *O2* orthologues in eudicotyledonous plants has not been established.

The purpose of this work is to get a detailed picture of the evolution of O2-related proteins in angiosperms, with the additional objective of defining the conditions for a broader analysis of the evolutionary history of angiosperm bZIP factors. To this end, we characterized four cDNAs that represent the complete set of *O2* homologues in the model eudicot plant *Arabidopsis thaliana* (*Arabidopsis*). Additionally, we identified a set of 41 higher-plant *O2* homologous genes, 24 of which were detected in EST databases. This set of genes defines what we call the O2 gene family. Phylogenetic analysis revealed that the evolution of the O2 family could be explained by monocot- and eudicot-specific gene duplication events from three ancestral genes. The possibility that *O2* is restricted to monocot species is also discussed.

Materials and Methods

Cloning of bZIP cDNAs by 3′ Amplification of cDNA Ends, cDNA Library Screening, and DNA Sequencing

The 3′ ends of bZIP cDNAs were amplified in two steps using two pairs of nested degenerated primers and the M13 reverse primer as a 3′-anchor in the cloning vector λ ZAP II. The first pair of nested primers is based on the sequence SNRESARRS, which is conserved in the basic domain of several plant bZIP proteins. The primers are BC5 (5′-TCHAAYMGDGARTCWGC-3′), which corresponds to the peptide SNRESA, and BC6 (5′-aaggaattcGARTCWGCHAGRAGGTC-3′), which corresponds to the peptide ESARRS. The second pair of nested primers is based on the sequence (V/A)KVKM (A/G)E(D/E), which is conserved in the leucine zipper of O2-related proteins. The degenerated primers are ZC3.1 (5′-GYNAAGGTRAAGATGG-3′), which corresponds to the peptide (V/A)KVKM (A/G), and ZC3.2 (5′-aaggaattcGTRAAGATGGSNGARG-3′), which corresponds to the peptide KVKM (A/G)E(D/E). *Eco*RI sites were included in primers BC6 and ZC3.2 to facilitate subsequent cloning.

A cDNA library from *Arabidopsis thaliana* (ecotype Columbia) green siliques, which was constructed in λ ZAP II (Giraudat et al. 1992), was used as the template for the first round of amplification. Approximately $3 \cdot 10^8$ phages were amplified with 10 pmol of M13 reverse primer and 150 pmol of either BC5 or ZC3.1 in a 100-µl reaction mixture containing 1.5 m*M* $MgCl_2$, 200 µ*M* deoxynucleotide triphosphate, and 3 U of *Taq* DNA polymerase in the buffer supplied by the manufacturer (BRL). Amplification conditions were 94°C for 4 min followed by 35 cycles (94°C 1 min, 42°C for 2 min, and 72°C for 2 min) and an extension step of 10 min at 72°C. The PCR products were then purified on a QIAquick PCR purification kit (Quiagen) to eliminate unincorporated primers and nucleotides. An aliquot of 3 µl of the purified product was then used in the second round of amplification, which also included 10 pmol of M13 reverse primer and 75 pmol of either primer BC6 or primer ZC3.2. The total reaction volume was 50 µl containing 1.5 m*M* $MgCl_2$, 200 µ*M* deoxynucleotide triphosphate, and 3 U of *Taq* DNA polymerase (BRL). An initial denaturation step at 94°C for 4 min was followed by 36 cycles: 94°C for 1 min, 50°C (first three cycles) and 56°C (the remaining 33 cycles) for 2 min, 72°C for 2 min, and an extension step of 10 min at 72°C. PCR products larger than 200 bp were gel purified, digested with *Eco*RI, and cloned into pBluescript SK^+ (Stratagene) for sequencing, which was done by the dideoxy dye terminator method (Perkin Elmer).

Amplification products encoding polypeptides similar to known bZIP proteins were used as probe to isolate the corresponding full-length cDNAs from the seed cDNA library using standard protocols (Sambrook et al. 1989) and starting with $8 \cdot 10^5$ recombinant phages. After four plaque purification steps, positive phages were converted to pBluescript SK^- phagemid clones by in vivo excision using the R408 helper phage according to the manufacturer's instructions (Stratagene) except that the last growth step was at 42°C. The complete sequence of the full-length *BZO2H1* and *BZO2H4* (EST T22560; Table 1) cDNAs required the subcloning of an internal *Hin*dIII/*Bam*HI fragment of 700 bp (in pBluescript SK^+) and an *Eco*RI fragment of 900 bp (in pUC18), respectively. Standard protocols were used for cloning (Sambrook et al. 1989).

Construction of a Nonredundant Data Set of Angiosperm bZIP Factors and Identification of Tentative Unique Genes from EST Collections

Construction of a nonredundant set of *Arabidopsis* bZIP (Vincentz et al. 2001) and of other angiosperm bZIP proteins was achieved through iterated searches of GenBank protein databases at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) and the MAtDB (Munich Information Center for Protein sequences *Arabidopsis thaliana* Database; http://www.mips.biochem.mpg.de/proj/thal/; v211200) using different known and distantly related bZIP as query sequences and the blastp and tblastn programs (Altschul et al. 1990) at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and MAtDB servers (http://mips.gsf.de/proj/thal/db/search/search_frame.html). Additionally, key word searches were also performed at the NCBI and MAtDB. Editing of exon sequences encoding some of the bZIP domains involved modifications of exon/introns junctions that were guided by amino acid sequence alignments and by the presence of donor (GT) and acceptor (AG) intron splice sites. ESTs related to O2 and its homologous sequences were selected from the GenBank dbEST (http://www.ncbi.nlm.nih.gov/) using tblastn program and protein sequences of O2 homologues (Table 1) as query sequences at the NCBI blast server. Selected ESTs were

Table 1. Opaque2 homologous proteins

Protein	Accession No.		
	Protein	Genomic	cDNA
Eudicot			
BZO2H1 *At*	(N) AAC78255	CRM IV	AF3 10222
	(M) At4g02640	AC002330	
BZO2H2 *At*	(M) At5g24800	CRM V	AF3 10223
		AF069716	
BZO2H3 *At*	(N) AAF67360	CRM V	AF310224
	(M) At5g28770	AF262041	AA042606
BZO2H4 *At*	(N) CAB77582	CRM III	AY057509
	(M) At3g54620	AL138656	T22560
BZI-1 *Nt*	AAL27150	—	AY061648
CPRF2 *Pc*	Q99090	—	X58577
Monocot			
O2 *Clj*	S42493	X78287	X78286
BLZ1 *Hv*	T04477	X80068	—
BLZ2 *Hv*	CAA71795	—	Y10834
REB *Os*	BAA36492	ABO21736	—
RISBZ1 *Os*	BAB39173	ABO53475	ABO53472
RISBZ4 *Os*	BAB39174	AAAA10058 68[a]	ABO53473
RISBZ5 *Os*	BAB39175	OSJNBb0065 C04[b]	ABO53474
RITA1 *Os*	T03990	AJ001267 AAAA01000 425[a]	L34551
02 *Sb*	CAA50642	X71636	—
SPA *Ta*	T06767	—	Y09013
02 *Zm*	AAA33489	X15544	M29411
OHP1 *Zm*	JQ2147	—	L00623
OHP1b *Zm*	AAC49533	—	U35063
OHP2 *Zm*	JQ2148	—	L06478

Note. (M) MAtDB Arabidopsis database (http://www.mips.biochem.mpg.de/proj/thal/), (N) NCBI (http://www.ncbi.nlm.nih.gov/). The underlined accession numbers are from the ESTs we refer to in the text. BZO2H1, -H2, and -H3 and the cDNA accession numbers are our own submissions. Species abbreviations: *At, Arabidopsis thaliana; Clj, Coix lacryma-jobi; Hv, Hordeum vulgare; Nt, Nicotiana tabacum; Os, Oryza sativa; Pc, Petroselinum crispum; Sb, Sorghum bicolor; Ta, Triticum aestivum; Zm, Zea mays.*

[a] *Oryza sativa* ssp. *indica* (NCBI).

[b] *Oryza sativa* ssp. *japonica* (MATDB).

compared pairwise with the blastn program using default parameters, and overlapping sequences that share at least 98% identity over 150 nucleotides or 100% identity over 100 nucleotides were assembled into a consensus sequence (CS) that defines tentative unique genes (TUGs). Care was taken to avoid probable unspliced variants. A second dbEST search was performed with the TUGs CS to complete our analysis. Finally, an additional assembly step was performed using the PHRAP program (Green 1994). To build a final TUG CS, the CS derived from PHRAP was compared with Clustal X (Thompson et al. 1997) to the TUGs CS obtained with blastn. Rice genomic sequences (ssp. *indica* and *japonicum*) were searched for O2 homologues at the NCBI and MATDB blast servers.

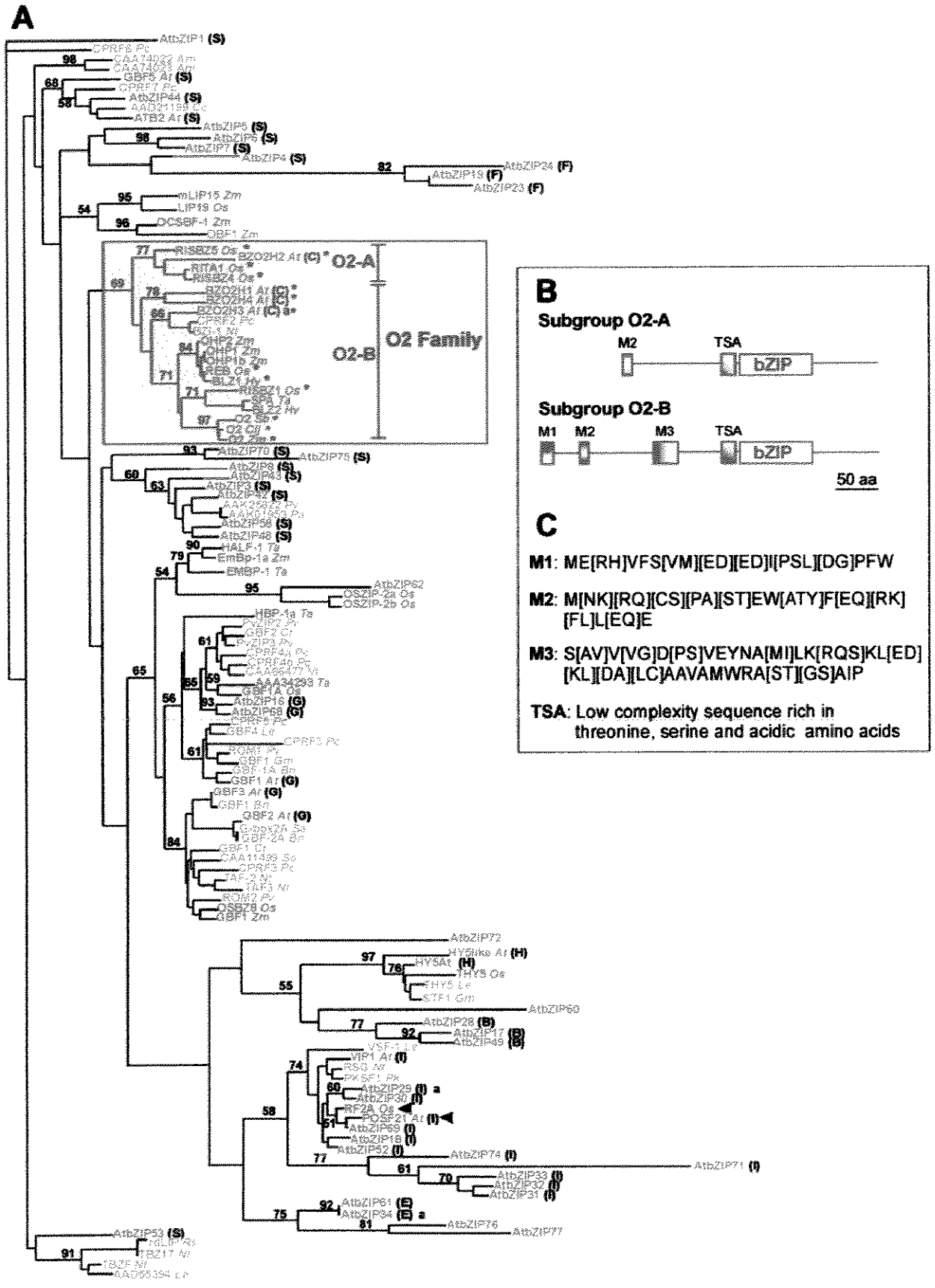
Phylogentic Analysis, Computer Programs, and Web Servers

Comparison of protein sequences with the blastp program (Altschul et al. 1990) was done using default parameters but without filtering. Standard unweighted maximum parsimony analysis was performed on nucleotide sequences by heuristic search using the branch swapping nearest-neighbor interchanges algorithm (search level, 3; initial tree search by random addition option with 100 replications) implemented in the Molecular Evolutionary Genetic Analysis (MEGA) package v 2.1 (Kumar et al. 2000). Neighbor-joining analyses of nucleotide sequence data were performed using different methods of distance estimation that are provided in the program MEGA package v 2.1 (Kumar et al. 2000). Maximum parsimony analysis of amino acid sequences was conducted with the PROTPAR program [PHYLIP, Phylogeny Inference Package version 3.57c (Felsenstein 1993)] and distance method analyses of amino acid sequences data were done with the NEIGHBOR program [PHYLIP, Phylogeny Inference Package version 3.57c (Felsenstein 1993)] using PAM distances (Dayhoff et al. 1978), which were obtained with the PROTDIST program (PHYLIP). The number of synonymous differences per synonymous site (d_s) and the number of nonsynonymous differences per nonsynonymous site (d_n) were calculated based on the method of Pamilo–Biancho–Li (Pamilo and Bianchi 1993; Li 1993) as implemented in MEGA v 2.1. Relative rate test analysis was conducted by the method of Tajima (1993) as implemented in MEGA v 2.1 and using *Arabidopsis BZO2H3* as an outgroup. DNA sequence analysis was carried out with the DNASIS program (Pharmacia). Sequences were aligned with Clustal X (Thompson et al. 1997). Prediction of protein secondary structure was obtained through the predictprotein server (http://www.emblheidelberg.de/predictprotein/) (Rost 1996). Identification of conserved protein motifs was performed with the motif discovering tool MEME (Bailey and Elkan 1994; http://meme.sdsc.edu/meme/website/).

Results

Definition of the Angiosperm O2 Family and of the Modular Protein Structure of Its Members

We identified four *Arabidopsis* cDNAs whose predicted polypeptides were found to be highly similar to O2 based on blastp comparisons. These *Arabidopsis* O2-related proteins were named BZO2H1, BZO2H2, BZO2H3, and BZO2H4 (*b*asic leucine zipper *O2 ho*mologous) (Table 1). *BZO2H1* and *BZO2H2* full-length cDNAs were isolated from a seed cDNA library (see Materials and methods). The full-length cDNA sequence of BZO2H3 was reconstructed by joining a partial cDNA that we had isolated from a seed cDNA library and covers the 5′-end mRNA sequence with the overlapping sequence of an EST (accession No. AA042666) that covers the 3′ end of the mRNA sequence. The full-length cDNA sequence of BZO2H4 was obtained from an EST (accession No. T22560). To define the group of monocot and eudicot O2 homologues we performed a phylogenetic analysis of a possibly complete and nonredundant repertoire of 168 *a*ngiosperm *bZIP* factors (ABZ data set) that we constructed from sequences available in public databases. From the complete sequence of the *Arabidopsis* genome (The Arabidopsis Genome Initiative 2000), it was possible to include in the ABZ data set a possible complete and nonredundant set of 76 *Arabidopsis* bZIP proteins. This set of *Arabidopsis*


A
AtbZIP1 (S)
98
68
58
AtbZIP5 (S)
98
AtbZIP7 (S)
AtbZIP4 (S)
82
AtbZIP19 (F)
AtbZIP24 (F)
AtbZIP23 (F)
95
54
96
O2-A
O2-B
O2 Family
99
78
84
71
71
97
93
AtbZIP70 (S)
AtbZIP75 (S)
60
63
AtbZIP8 (S)
AtbZIP43 (S)
AtbZIP3 (S)
AtbZIP42 (S)
AtbZIP58 (S)
AtbZIP48 (S)
90
79
54
95
AtbZIP52 (S)
91
61
65
55
59
56
93
AtbZIP16 (G)
AtbZIP68 (G)
61
GBF1 At (G)
GBF3 At (G)
GBF2 At (G)
84
77
AtbZIP72
97
76
55
AtbZIP60
77
92
AtbZIP28 (B)
AtbZIP17 (B)
AtbZIP49 (B)
74
60
58
61
77
61
70
92
75
81
AtbZIP61 (E)
AtbZIP34 (E) a
AtbZIP76
AtbZIP77
AtbZIP71 (I)
0.1
B
Subgroup O2-A
M2
TSA
bZIP
Subgroup O2-B
M1
M2
M3
TSA
bZIP
50 aa
C
M1: ME[RH]VFS[VM][ED][ED]I[PSL][DG]PFW
M2: M[NK][RQ][CS][PA][ST]EW[ATY]F[EQ][RK][FL]L[EQ]E
M3: S[AV]V[VG]D[PS]VEYNA[MI]LK[RQS]KL[ED][KL][DA][LC]AAVAMWRA[ST][GS]AIP
TSA: Low complexity sequence rich in threonine, serine and acidic amino acids

bZIP factors was obtained by the integration of our own data (Vincentz et al. 2001) and those of The bZIP Research Group (Jakoby et al. 2002) and is slightly smaller than the set of 81 bZIP proteins described earlier (Riechmann et al. 2000).

In agreement with previous results (Vettore et al. 1998), significant amino acid sequence similarity among all these angiosperm bZIP proteins was found to be restricted to the minimum bZIP DNA binding domain, which consists of the basic motif and three leucine repeats (of 44 amino acids, which correspond to positions 228 to 271 in the maize O2 sequence: Fig. 2). A neighbor-joining analysis of the minimum bZIP domain (amino acid sequences) of the ABZ data set revealed the existence of two large clusters of proteins with significant bootstrap support (results not shown). These two clusters include all *Arabidopsis* bZIP factors of family III/group A (51% bootstrap support) and family II/group D (98% bootstrap support) of Vincentz et al. (2001) and Jakoby et al. (2002), respectively. As these two clusters did not contain O2 and were found to be responsible for the phylogenetic analysis of the ABZ data set being restricted to the minimal bZIP domain, their members were excluded from the ABZ collection to create a subset of the ABZ collection (SABZ data set) that includes 122 bZIP proteins. The length of the sequences that could be aligned from the SABZ data set was increased by two leucine repeats (16 amino acids) compared to the minimal bZIP domain. This, in turn, allowed an improvement in the resolution of the evolutionary relationships among the SABZ set of proteins. The unrooted tree inferred from a neighbor-joining analysis of the bZIP domain amino acid sequences of the SABZ data set is shown in Fig. 1A. This tree identifies a well-supported group of proteins, which most likely are O2 homologues, and was defined as the O2 family (Fig. 1A). This family is formed by 6 eudicot (including the complete set of *Arabidopsis* O2 homologues BZO2H1, BZO2H2, BZO2H3, and BZO2H4) and 14 monocot proteins (Table 1). The gene structure of 13 members of the O2 family is available and was compared. For all of them, 78% of the bZIP domain is encoded by exons 4 and 5, whose size and positions are conserved (Fig. 2). Furthermore, this feature appears to be specific to members of the O2 family as judged by the analysis of the complete set of *Arabidopsis* genes encoding bZIP factors (data not shown). These results indicate that this set of 13 genes of the O2 family are homologous and support the notion that the set of 20 angiosperm bZIP factors that are included in the O2 family are indeed O2 homologous proteins.

The proteins of the O2 family share a highly conserved bZIP domain which has the double function of DNA binding and nuclear localization (Varagona and Raikel 1994). Prediction of coiled-coil structures by the Coil program (Lupas 1996) indicated that up to nine leucine (hydrophobic residues) heptad repeats may be involved in the leucine zipper dimerization domain of the O2-related proteins (Fig. 2). Additionally, a set of conserved motifs which may define important functional sequences was identified (Fig. 1B). Motifs M1, M2, and TSA participate in

Fig. 1. A Phylogenetic tree defining the O2 family. The unrooted tree was inferred by the neighbor-joining method using PAM distances of bZIP-domain amino acid sequences (basic motif plus five leucines of the leucine zipper; positions 228 to 287 of the maize O2 protein in Fig. 2) of approximately two-thirds of the complete and nonredundant set of angiosperm bZIP factors that are available in databases. Bootstrap values of 500 replicates are indicated as percentages along the branches. The O2 family is *boxed in gray*. Known gene structures of members of the O2 family are marked with an *asterisk*. *Arabidopsis*, other eudicot, and monocot proteins are shown in *blue*, *green*, and *red*, respectively. The bZIP-domain amino acid sequences of proteins marked with an *a* were edited (see Materials and Methods). AtbZIP76 (accession No. AAG50695) and AtbZIP77 (accession No. NP_564460) are putative new *Arabidopsis* proteins. AtbZIP73 (MATDB accession No. At2g13130) presents a stop codon in the leucine zipper coding part and was therefore not included in our data set. The classification of *Arabidopsis* bZIPs established by The bZIP Research Group (Jakoby et al. 2002) is indicated in *parentheses* (B, C, E, F, G, H, I, and S). The two *arrows* point to the outgroup used in the parsimony analysis in Fig. 3. Species abbreviations and accession numbers for the O2 family members are as in Table 1. Accession numbers of *Arabidopsis* proteins are those of The bZIP Research Group (Jakoby et al. 2002). Other accession numbers for eudicot proteins are as follows: rdLIP *Rs*, BAA34938; TBZF *Nt*, BAB13719; CPRF6 *Pc*, CAC00657; CPRF7 *Pc*, CAC00658; TBZ17 *Nt*, BAA22204; GBF1 *Cr*, AAD42937; GBF2 *Cr*, AAD42938; PvZIP2 *Pv*, AAK39130; PvZIp3 *Pv*, AAK39131; CPRF5 *Pc*, CAC00656; GBF1 *Gm*, AAB00096; G-box2A *Sa*, T10472; GBF-1A *Bn*, CAA58774; GBF1 *Bn*, AAB03379; GBF-2A *Bn*, CAA58772; ROM2 *Pv*, AAC49474; ROM1 *Pv*, AAB36514; CPRF3 *Pc*, CAA41452; GBF4 *Le*, CAA52896; CPRF4a *Pc*, CAA71768; CPRF4b *Pc*, CAA71770; TAF-2 *Nt*, CAA88492; TAF-3 *Nt*, CAA88493; THY5 *Le*, CAB57979; STF1 *Gm*, AAC05017; RSG *Nt*, BAA97100; PKSF1 *pk*, AAC04862; VSF-1 *Le*, CAA05898. For monocot proteins accession numbers are as follows: mLIP15 *Zm*, BAA05117; Lip19 *Os*, CAA40596; OCSBF-1 *Zm*, CAA44607; OBF1 *Zm*, JQ0984; HBP-1a *Ta*, BAA02304; HALF-1 *Ta*, BAA10928; EMBP-1 *Ta*, AAA68428; GBF1A *Os*, T03241; OSBZ8 *Os*, AAB40291; OSZIP2a *Os*, AAC49557; OSZIP2b *Os*, AAC49558; GBF1 *Zm*, AAA80169; EMBP-1a *Zm*, CAB62402; THY5 *Os*, BAB62558; RF2a *Os*, AAC49832. Species abbreviations not listed in Table 1 are as follows: *Am*, *Antirrhinum majus*; *Bn*, *Brassica napus*; *Cc*, *Capsicum chinense*; *Cr*, *Catharanthus roseus*; *Gm*, *Glycine max*; *Le*, *Lycopersicon esculentum*; *Pa*, *Phaseolus acutifolius*; *Pk*, *Paulownia kawakamii*; *Pv*, *Phaseolus vulgaris*; *Rs*, *Raphanus sativus*; *Sa*, *Sinapsis alba*; *So*, *Spinacia oleracea*; *Vf*, *Vicia faba*. The scale bar corresponds to 0.1 estimated amino acid substitution per site. **B** Modular organization of O2 homologous proteins. The distribution of the conserved motifs M1, M2, M3, and TSA along the protein sequence is schematized. These motifs were defined with the help of the MEME tool (Bailey and Elkan 1994). bZIP, basic leucine zipper DNA binding domain shown in Fig. 2. **C** Multilevel consensus sequences (as defined by MEME) of the conserved motifs shown in B.

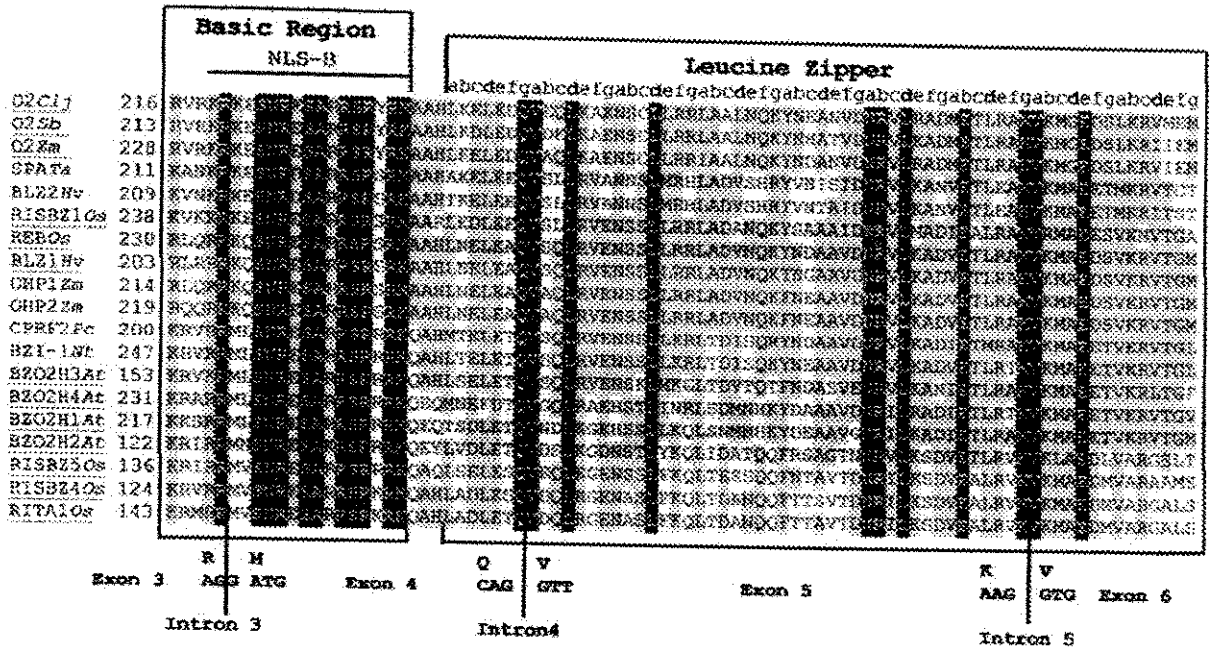


Fig. 2. Alignment of bZIP amino acid sequences of angiosperm O2 homologous proteins. The basic region includes a nuclear localization signal (NLS). The position of the amino acids in each heptad repeat of the leucine zipper is indicated (a to g) and the position (d) of leucines (hydrophobic residues) is shown in *bold face*. Identical and conserved amino acids are *boxed in gray* when present in at least 50% of the proteins. Strictly conserved residues are highlighted in *black*. Proteins whose gene structures are known are *underlined* and the positions of introns 3, 4, and 5 are shown. Species abbreviations and accession numbers are as in Table 1. The sequence of OHP1b is 97% identical to OHP1 and was therefore not included.

transcriptional activation (Schmitz et al. 1997; Vincente-Carbajosa et al. 1998; Onodera et al. 2001), and the control of nucleocytoplasmic distribution partly involves the TSA and M3 motifs (Varagona and Raikhel 1994; Kircher et al. 1999).

These motifs can be considered as informative shared derived characters that were used to classify members of the O2 family further into the two subgroups (clades) O2-A and O2-B, which are distinguished by the presence/absence of motifs M2 and M3 (Fig. 1B). This classification is consistent with the phylogenetic analysis where members of subgroup O2-A form a well-supported cluster of proteins (Fig. 1A).

Evolution of the O2 Gene Family

Our main objective in initiating a detailed analysis of the evolution of the O2 family was to establish orthologous relationships among members of the family through the identification of groups of orthologous genes (GO). A GO consists of individual orthologous genes or orthologous groups of paralogues from several lineages (Tatusov et al. 1997). Orthologous and paralogous genes are homologous genes that result from a speciation event and from a duplication event within a lineage, respectively (Tatusov et al. 1997; Fitch 2000; Thornton and DeSalle 2000). An important aspect of defining GOs is that it should facilitate the identification of ancestral genes and should be useful to rationalize the systematic analysis of yet uncharacterized proteins (Thornton and DeSalle 2000).

In the first step, to improve the phylogenetic analysis of the angiosperm O2 family, we searched for new O2 homologues that would be represented in plant EST databases. ESTs selected through iterated searches in the GenBank plant dbEST were assembled to form a consensus sequence (CS) that represents tentative unique genes (TUGs). The polypeptides corresponding to each TUG were confirmed to belong to the O2 family if they had the best blastp match with one of the O2 homologues defined here (Table 1). Based on the 96.5% amino acid identity that was observed between the coding sequences of the two maize recent paralogues, OHP1 and OHP1b (Pysh and Schmidt, 1996), only those CSs whose deduced amino acid sequence showed less than 97% identity over their full-length sequence with one of the O2 homologues were considered to represent new genes. Our scheme resulted in the identification of 11 monocot and 13 eudicot putative new *O2* homologous genes. We noticed, however, that in some cases, nonoverlapping CSs whose deduced polypeptides show a higher similarity to the same O2 homologue might actually identify the same gene. Additionally, we identified in the recently published rice genome (Yu et al. 2002) a new *O2* homologous gene (*RBZO2H*). This new *O2* homologous rice gene and nine of the TUGs which

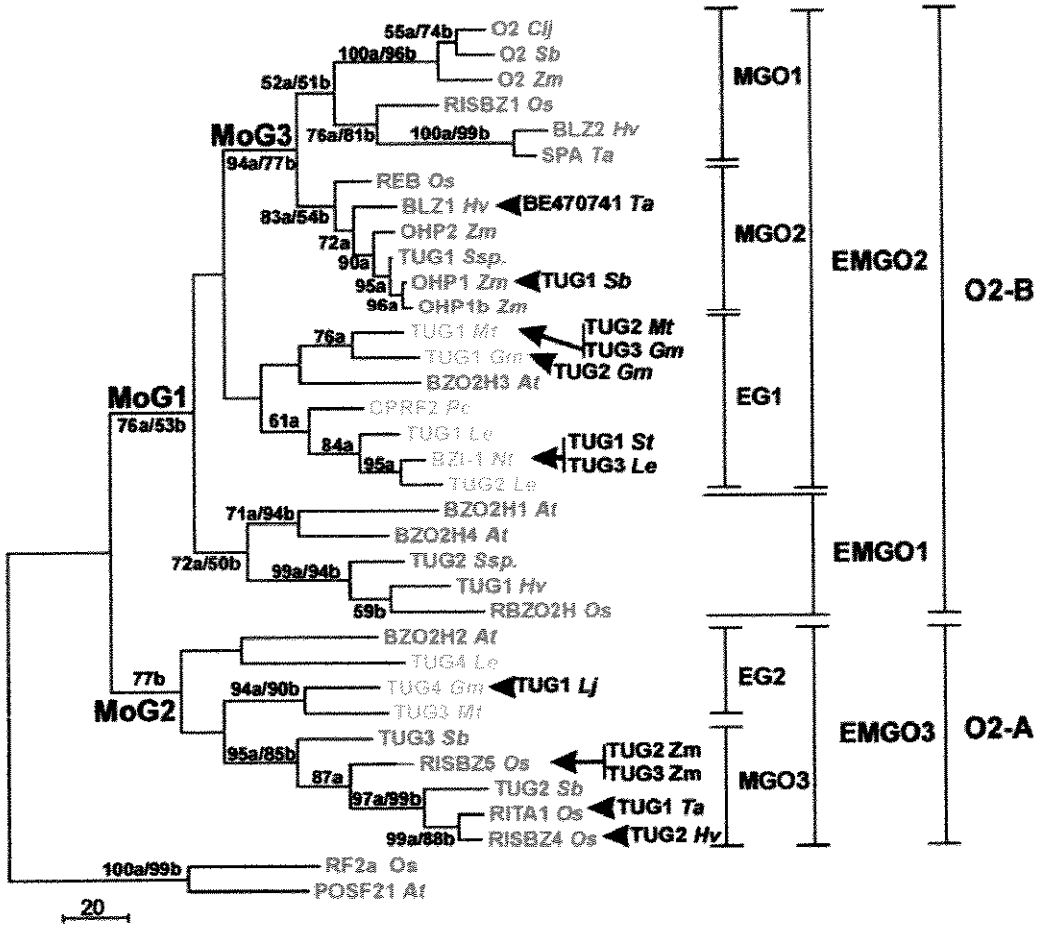


Fig. 3. Phylogeny of the *O2* homologous genes. A representative maximum parsimony rooted tree is shown. Parsimony analysis of bZIP-domain (positions 228 to 313 of maize O2; Fig. 2) DNA sequences of *O2* homologues was obtained by the nearest-neighbor interchange (branch swapping) algorithm included in the MEGA 2 package v 2.1 (Kumar et al. 2000). Bootstrap support over 50% obtained from an analysis including all three sites of each codon (a) or obtained from an analysis including the first two nucleotides of each codon (b) are shown along the branches. For the analysis with three sites, the number of informative sites was 204/258; the tree length, 1241; the informative site consistency index, 0.36; the informative site retention index, 0.6; and the informative site rescaled consistency index, 0.22. For the analysis with two sites, the number of informative sites was 119/172; the tree length, 477; the informative site consistency index, 0.49; the informative site retention index, 0.72; and the informative site rescaled consistency index, 0.35. *Arabidopsis*, other eudicot, and monocot proteins are shown in *blue*, *green*, and *red*, respectively. The GenBank accession number of the genomic sequence (*Oryza sativa* ssp. *indica*) encoding RBZO2H is AAAA01005225. Tentative unique genes (TUGs) that were not included in the phylogenetic analysis point to their closest related O2 homologue in the tree (*arrow*). Highest similarity of the TUGs that were not included in the parsimony analysis to one of the O2 homologues was defined by protein distances that were measured with the PRODIST program (PHYLIP) on aligned amino acid sequences that were obtained with Clustal X using default parameters. MoG, *m*onophyletic group of proteins; MGO, monocot group of orthologues; EG, eudicot group of genes; EMGO, eudicot–monocot group of orthologues. O2-A and O2-B refer to the two subgroups defined in Fig. 1. Species abbreviations are as in the legend to Fig. 1, and *Ssp.* stands for *Saccharum* sp. The scale bar represents 20 substitutions (obtained from an analysis including the three nucleotides of codons).

encoded a complete bZIP domain were included in the phylogenetic analysis of the O2 family presented hereafter.

The evolutionary history of the O2 family was estimated from the bZIP DNA binding domain, which is the unique well-defined structural and functional domain present in all O2 homologues. The phylogeny of the O2 family was evaluated by a maximum parsimony analysis of nucleotide sequences of the bZIP domain shown in Fig. 2. This analysis included a pair of outgroup sequences formed by the monocot RF2a gene from rice and by the POSF21 gene from *Arabidopsis* (Fig. 1). These sequences were chosen as outgroups based on the shared position of an intron in the bZIP domain with *O2* homologues, indicating that these outgroups and the O2 family most likely derive from a common ancestor (result not shown). The rooted tree inferred from the parsimony analysis is shown in Fig. 3. Parsimony analysis of bZIP domain amino acid sequences as well as distance methods (neighbor joining) applied to nucleotide and amino acids sequences

of the bZIP domain data set gave essentially the same results (data not shown).

To interpret the tree shown in Fig. 3, we considered the following three criteria. First, bootstrap support over 50% was retained for the branching pattern. Second, assuming that the complete set of *Arabidopsis* and rice *O2* homologues was identified and no selective gene loss occurred, each group of eudicot orthologues should include at least one *Arabidopsis* gene, each group of monocot orthologues should include at least one rice gene, and each group of eudicot–monocot orthologues should include at least one *Arabidopsis* and one rice gene. Third, the inferred gene phylogeny should be consistent with the known species phylogeny. Accordingly, the tree in Fig. 3 was organized into two main *mo*nophyletic groups, MoG1 and MoG2. The consistency between the composition of these two monophyletic groups and the composition of the two subgroups O2-A and O2-B, which were defined by different means (Fig. 1), provides additional support for the definition of MoGO1 and MoG2. MoG1 was further divided into MoG3, which is formed by the two *m*onocot *g*roups of *o*rthologous genes MGO1 (includes O2) and MGO2, the *e*udicot *g*roup of proteins EG1, and the *e*udicot–*m*onocot *g*roup of *o*rthologues EMGO1 (Fig. 3). MoG2 was resolved into the eudicot group of genes EG2 and the monocot group of orthologues MGO3 (Fig. 3). This general organization was supported by the observation that the polypeptides encoded by the six monocot and six eudicot TUGs, which were not included in our phylogenetic analysis because they do not cover a complete bZIP domain, are more closely related to the monocot and eudicot O2 homologues, respectively (Fig. 3). The relationships among the eudicot genes in EG1 and EG2 (Fig. 3) could not be resolved clearly, essentially because the two *Arabidopsis* genes *BZO2H3* (EG1) and *BZO2H2* (EG2) do not cluster significantly with any of the other eudicot genes of these groups. However, considering that EG1 and EG2 fulfill the second and third criteria used earlier to interpret our tree (see above), they are likely to represent eudicot groups of orthologues. We also noticed that the two *Arabidopsis* genes in EMGO1, *BZO2H1* on chromosome IV and *BZO2H4* on chromosome III, are part of two conserved and collinear segments formed by two matching genes (result not shown). These data indicate that *BZO2H1* and *BZO2H4* are two paralogues that probably arose with one of the large-scale duplications that formed the *Arabidopsis* genome (Vision et al. 2001). Finally, the orthologous relationship among members of MGO1, which includes O2 (Fig. 3), was further supported by their shared seed-specific expression, which is a characteristic restricted to this group of genes (Schmidt et al. 1992; Albani et al. 1997; Yunes et al. 1998; Oñate et al. 1999; Onodera et al. 2001; our unpublished data). Together these data suggest that the O2 family can be organized into three eudicot–monocot groups of orthologous genes (EMGO1, -2, and -3; Fig. 3).

Table 2. d_s and d_n among members of the monocot group of orthologues MGO1 and monocot group of orthologues MGO2

	d_s (±SE)	d_n (±SE)
MGO1: *O2 Zm–O2 Sb*	0.408 (0.108)	0.049 (0.017)
MGO2: OHP1 *Zm*–TUG2 *Sb*	0.086 (0.039)	0

Note. d_s and d_n were estimated over 234 nucleotides that encode 91% of the bZIP domain (positions 236 to 313 for the maize O2 factor: Fig. 2). Species abbreviations are as in Table 1, Note.

Our phylogenetic analysis also indicates that the monocot MGO1 and MGO2 emerged as the result of a gene duplication that must have occurred after the separation of monocots and eudicots and some time before the radiation of grasses (Fig. 3). To characterize this duplication event further, we decided to define d_s and d_n for pairs of orthologues in each of these two MGOs. To obtain reliable estimations for d_s ($d_s < 0.5$), we used the maize and sorghum *O2* sequences in MGO1 and their corresponding paralogues, the maize *OHP1* and the sorghum *TUG1* sequences in MGO2 (Fig. 3). We also verified the homogeneity of substitution rates among these two groups of orthologues by applying Tajima's (1993) relative rate test. No rate heterogeneity was observed among the maize and the sorghum *O2* (MGO1) or among the maize *OHP1* and the sorghum *TUG2* orthologues (MGO2). However, these two pairs of orthologues evolved at a significantly different rate (result not shown). Estimations of d_s and d_n were then calculated for 91% of the bZIP sequence, which is the limit imposed by the sorghum TUG2. As shown in Table 2, d_s and d_n were higher in MGO1 than in MGO2. The same trend was observed for the rice *RISBZ1* and the barley *BLZ2* genes (MGO1) and the corresponding paralogues in MGO2, the rice *REB* and the barley *BLZ1* genes (result not shown). Our data suggest that genes in MGO1 (*O2* orthologues) evolved more rapidly than their corresponding paralogues in MGO2.

Discussion

As the first step toward a broad analysis of bZIP factor evolution in angiosperms, we choose to focus our analysis on the origin of the maize *O2* regulatory locus, which is one of the best-characterized bZIP proteins (see Introduction). Additionally, our interest was to identify *O2* orthologues in eudicot species. To this end, we developed a two-step phylogenetic approach. First, from the analysis of the possible complete and nonredundant set of known angio-

sperm bZIP proteins (ABZ data set), we defined a subset of bZIP factors (SABZ data set) that allowed us, in the second step, to identify a group of 20 O2 homologous proteins (the O2 family; Fig. 1). Our analysis also allowed us to identify several other clusters of proteins with significant bootstrap support (Fig. 1) and this grouping was found to be consistent with classification schemes described earlier for the *Arabidopsis* bZIP proteins (Vincentz et al. 2001; Jakoby et al. 2002). Taken together, our results indicate that the ABZ and SABZ data sets should be useful to improve our knowledge about the evolution of angiosperm bZIP factors.

We further characterized the *O2* homologous genes by determining the modular structure of the corresponding proteins. The functionally essential bZIP DNA binding domain is highly conserved and includes a leucine zipper that is possibly formed by nine leucines (hydrophobic residues). Conserved motifs involved in the control of nuclear translocation or transcriptional activation were also identified (Fig. 1B). Such functional motifs can be considered as shared derived characters and allowed us to divide the O2 family into the two evolutionary distinct subgroups O2-A and O2-B (Fig. 1). Finally, we notice that more than half of the protein sequence of all O2 homologues is poorly conserved, which could reflect either weak functional constraints or functional diversification as has been suggested in the case of the *R* family of basic helix–loop–helix regulatory genes (Purugganan and Wessler 1994). These observations indicate that an accurate phylogenetic analysis of the O2 family should rely mainly on the sequence of the bZIP domain, and this seems to be true for the majority of the others bZIP families (Fig. 1 and unpublished results).

The interpretation of the parsimony analysis of the bZIP domain of O2 homologues relied on (1) bootstrap support over 50%, (2) the assumption that the complete set of *Arabidopsis* and rice O2 homologues was identified, and (3) congruence between the inferred gene phylogeny and the known species phylogeny. Following these criteria, the examination of the tree inferred from parsimony analysis led to the identification of three eudicot/monocot groups of orthologues (EMGO1, -2, and -3 in Fig. 3). An effort was made to incorporate information extracted from EST databases into the phylogenetic analysis. This approach was shown here to improve the reliability of the organization of homologous genes into groups of orthologues.

The monocot species in our analysis are represented exclusively by members of the grass family, which diverged approximately 60 million years ago (MYA) (Kellog 2001). Identifying groups of orthologues among grasses should therefore be facilitated by the fact that they form a compact group of species.

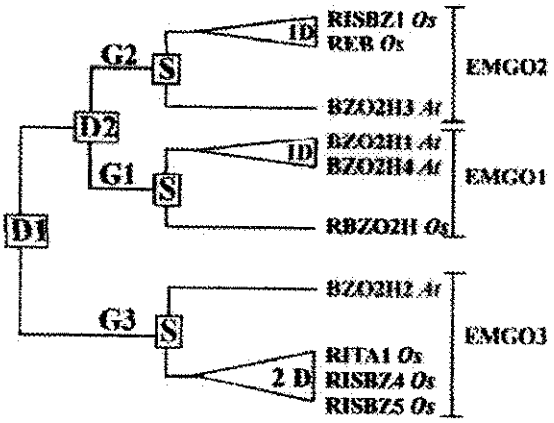


Fig. 4. Models of O2 family evolution. In the ancestral lineage of angiosperms, two duplications, D1 and D2, produced the three *O2* homologues G1, G2, and G3 (ancestral functions). Following the separation (S) of monocots and eudicots, lineage-specific duplications (D) occurred. Representative eudicot genes were from *Arabidopsis* (*At*) and are *boxed in gray*, and representative monocot genes are from rice (*Os*).

However, whenever groups of orthologues evolve at different rates, such as in the case of the MGO1 and MGO2 genes (Fig. 3), additional information such as expression pattern, functional hints, or map position may be required to support orthology. On the other hand, the difficulty of establishing the relationships among eudicot genes, as in the case of the genes included in EG1 and EG2 (Fig. 3), is due partly to the early radiation of eudicot species, such as tomato (*Lycopersicon esculentum*) and *Arabidopsis*, which diverged about 112–156 MYA (Yang et al. 1999). The general strategy presented here sets the conditions for a broader analysis of the evolutionary relationship among angiosperm bZIP factors and the recent publication of the rice genome (Goff et al. 2002; Yu et al. 2002) should significantly improve such an analysis.

The phylogenetic analysis of the O2 family (Fig. 3) suggests a single model to explain the evolutionary history of the angiosperm O2 homologues (Fig. 4). In this model, three O2 homologous genes were present in the angiosperm ancestral lineage. After the separation of eudicots and monocots, lineage-specific gene duplications further shaped the O2 family. This model underscores the potential for functional diversification or innovation from the lineage-specific gene duplication events (Ohno 1970; Walsh 1995; Hughes 1994; Nadeau and Sankoff 1997; Zhang et al. 1998; Gu 1999; Lynch and Conery 2000; Lynch and Force 2000; Ohta 2000; Wendel 2000). For instance, the monocot–dicot EMGO3 (Fig. 3) is formed by one eudicot gene, which is likely to represent an ancestral function, and up to three monocot genes, which have diverged from each other and consequently provided the opportunity to acquire new functional specificity.

Not much is known about the possible functions of the factors included in EMGO3 (but see Onodera et al. 2001). Reverse genetic approaches aimed at inactivating the unique *Arabidopsis* gene *BZO2H2* that is included in EMGO3 is an obvious approach to learn more about EMGO3 genes.

A more informative example is provided by MGO1 (O2 orthologues) and MGO2. These two MGOs were produced by a gene duplication that is restricted to monocots and that must have happened some time before the diversification of grasses (Fig. 3). The large differences in d_s and d_n between MGO1 and MGO2 genes (Table 2) most likely reflects altered functional constraints between these two MGOs and supports functional change between them (Zhang et al. 1998; Gu 1999; Graur and Li 2000). As this conclusion is based on the analysis of the bZIP domain, functional divergence between MGO1 and MGO2 genes may concern DNA binding and/or dimerization specificity. Functional divergence between MGO1 and MGO2 genes is further supported by molecular data gained from the maize *O2* gene (MGO1) and its paralogue *OHP1* (MGO2). For instance, the *in vitro* interaction of O2 with a member of the Dof class of plant Cys_2–Cys_2 zinc-finger DNA binding protein is not shared by OHP1 (Vicente-Carbajosa et al. 1997). Additionally, the expression patterns of *O2* and *OHP1* are different. O2 is specifically expressed in the endosperm and is under the control of a circadian clock, while OHP1 is expressed in the endosperm, root, shoots, leaves, and embryos and is not regulated by an endogenous clock (Ciceri et al. 1999; Pysh et al. 1993). Changes in the coding and regulatory sequences seem, therefore, to have contributed to the functional divergence between MGO1 and MGO2 genes (Wendel 2000). *O2* orthologues (MGO1) may have evolved toward the function of integrating the synthesis of prolamine seed storage proteins with carbon and nitrogen metabolism during endosperm development. The fact that those prolamines are considered to have appeared specifically in grasses (Shewry et al. 1995) raises the interesting possibility of a coordinated evolution of prolamines and *O2*. An implicit consequence of this model is that *O2* is a function that has been recently acquired, while MGO2 genes, which seem to be under stronger selective constraints (Table 2), represent a conserved ancestral function possibly involved in the control of some aspect of carbon and nitrogen metabolism. One predominant adaptive event in this model may have been the acquisition of endosperm-specific expression by *O2*. It remains to be determined if neofunctionalization (Walsh 1995) or subfunctionalization (Lynch and Force 2000) is responsible for the preservation of MGO1 and MGO2 genes.

O2 controls lysine degradation during maize endosperm development by regulating the expression of the gene encoding lysine–ketoglutarate reductase/saccharopine deshydrogenase (*LKR/SDH*) (Kemper et al. 1999), and it was recently shown that knockout *Arabidopsis LKR/SDH* mutants accumulate lysine during seed development (Zhu et al. 2001). It appears, therefore, that control of lysine catabolism as a means of regulating its accumulation in seeds is conserved among angiosperms. It will be interesting to see if the *Arabidopsis* bZIP factor BZO2H3, which is more closely related to the monocot MGO1 and MGO2 genes (Fig. 3), is involved in this regulatory process.

Acknowledgments. The authors thanks Dr. J. Giraudat (Institut des Sciences Végétales, Gyf sur Yvette) for the *Arabidopsis thaliana* cDNA library and the Arabidopsis DNA Stock Center for ESTs. We also acknowledge the support of Dr. P. Arruda and L. Grivet, F.R. Silva, and L.B. Klaczko for critically reading the manuscript. S.L. Martins Silva is acknowledged for technical assistance. This work was supported by a grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)/Projeto Temático de Equipe (PTE)-99/02198-3.

References

Albani D, Hammond-Kosack MCU, Smith C, Conlan S, Colot V, Holdsworth M, Bevan M (1997) The wheat transcriptional activator SPA: A seed specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. Plant Cell 9:171–184

Altschul SF, Gish W, Miller W, Myers EW, Lipman D (1990) Basic local alignment tool. J Mol Biol 215:403–410

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

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, CA, pp 28–36

Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ (1999) The activity of the maize Opaque2 transcriptional activator is regulated diurnally. Plant Physiol 121:1321–1327

Cord Neto G, Yunes JA, Vettore AL, da Silva MJ, Arruda P, Leite A (1995) The involvement of Opaque2 in β-prolamine gene regulation in maize and Coix suggests a more general role for this transcriptional activator. Plant Mol Biol 27:1015–1029

Damerval C, le Guilloux M (1998) Characterization of novel proteins affected by the *o2* mutation and expressed during maize endosperm development. Mol Gen Genet 257:354–361

Dayhoff MO, Schwartz RM, Orcutt BC (1978) In: Dayhoff MO (ed) Atlas of protein sequence and structure, Vol 5, Suppl 3. National Biochemical Research Foundation, Silver Spring, MD, pp 345–352

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA

Fitch WM (2000) Homology. A personal view on some of the problems. Trends Genet 16:227–231

Gallusci P, Varrot S, Matsuoko M, Maddaloni M, Thompson RD (1996) Regulation of cytosolic pyruvate, orthophosphate dikinase expression in developing maize endosperm. Plant Mol Biol 31:45–55

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. Plant Cell 4:1251–1261

Goff AS, Ricke D, Lan T-H, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonicum*). Science 296:92–100

Graur D, Li W-H (2000) Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA

Green P (1994) *phrap* (http://www.genome.washington.edu/)

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16:1664–1674

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B 256:119–124

Hurst H (1995) Transcription factor 1: bZIP proteins. Protein Profile 2:105–168

Jakoby M, Weisshaar B, ge-laser W, Carbajosa JV, Tiedemann J, Kroj T, Parcy F (The bZIP Research Group) (2002) bZIP transcription factors in Arabidopsis. Trends Plant Sci 7:106–111

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125:1198–1205

Kemper EL, Cord Neto G, Papes F, Martinez Moraes KC, Leite A, Arruda P (1999) The role of Opaque2 in the control of lysine-degrading activities in developing endosperm. Plant Cell 11:1981–1993

Kircher S, Wellmer F, Nick P, Rügner A, Schäfer E, Harter K (1999) Nuclear import of the parsley bZIP transcription factor CPRF2 is regulated by phytochrome photoreceptors. J Cell Biol 144:201–211

Kumar S, Tamura K, Jakobsen IB, Nei M (2000) Molecular evolutionary genetic analysis (MEGA) v 2.1. http://www.megasoftware.net/

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. J Mol Evol 36:96–99

Lohmer S, Maddaloni M, Motto M, Di Fonzo N, Hartings H, Salamini F, Thompson RD (1991) The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. EMBO J 10:617–624

Lupas A (1996) Prediction and analysis of coiled-coil structures. Methods Enzymol 266:513–525

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Force A (2000) The probability of gene preservation by subfunctionalization. Genetics 154:459–473

Nadeau JH, Sankoff FD (1997) Comparable rates of gene loss and functional divergence after genome duplication early in vertebrate evolution. Genetics 147:1259–1266

Ohno S (1970) Evolution by gene duplication. Spring Verlag, Berlin–Heidelberg–New York

Ohta T (2000) Evolution of gene families. Gene 259:45–52

Oñate L, Vicente-Carbajosa J, Lara P, Díaz I, Carbonero P (1999) Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. J Biol Chem 274:9175–9182

Onodera Y, Suzuki A, Wu C-Y, Washida H, Takaiwa F (2001) A rice functional transcription activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem 276:14139–14152

Pamilo P, Bianvhi NO (1993) Evolution of the Zfx and Zfy genes: Rates and interdependence between the genes. Mol Biol Evol 19:271–281

Purugganan MD, Wessler SR (1994) Molecular evolution of the plant R regulatory gene family. Genetics 138:849–854

Pysh LD, Schmidt RJ (1996) Characterization of the maize OHP1 gene: Evidence of gene copy variability among inbreds. Gene 177:203–208

Pysh LO, Aukerman MJ, Schmidt RJ (1993) OHP1: A maize basic domain/leucine zipper protein that interacts with Opaque2. Plant Cell 5:227–236

Riechmann JL, Heard J, Martin J, Reuber F, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghanderhari D, Sherman BK, Yu GL (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Rost B (1996) PHD: Predicting one-dimensional protein structure by profile based neural networks. Methods Enzymol 266:525–539

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schmidt RJ (1993) Opaque-2 and zein genes expression. In: Verna DPS (ed) Control of plant gene expression. CRC Press, Boca Raton, FI, pp 337–355

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G (1992) Opaque-2 is a transcriptional activator that rePCOGnizes a specific target site in 22-kD zein genes. Plant Cell 4:689–700

Schmitz D, Lohmer S, Salamini F, Thompson RD (1997) The activation domain of the maize transcription factor Opaque-2 resides in a single acidic region. Nucleic Acids Res 25:756–763

Shewry PR, Napier JA, Tatham AS (1995) Seed storage proteins: Structures and biosynthesis. Plant Cell 7:945–956

Tajima F (1993) Simple methods for testing molecular clock hypothesis. Genetics 135:599–607

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Thornton JW, DeSalle R (2000) Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet 1:41–73

Varagona MJ, Raikhel NV (1994) The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. Plant J 5:207–214

Vettore AL, Yunes JA, Cord Neto G, da Silva MJ, Arruda P, Leite A (1998) The molecular and functional characterization of an Opaque2 homologue gene from *Coix* and a new classification of plant bZIP proteins. Plant Mol Biol 36:249–263

Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamine box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. Proc Natl Acad Sci USA 94:7685–7690

Vicente-Carbajosa J, Oñate L, Lara P, Diaz I, Carbonero P (1998) Barley BLZ1: A bZIP transcriptional activator that interacts with endosperm-specific gene promoters. Plant J 13:629–640

Vincentz M, Schlögl P, Corrêa LG, Kühne F, Leite A (2001) Phylogenetic relationships between Arabidopsis and sugarcane bZIP transcriptional regulatory factors. Gen Mol Genet 24:55–60

Vision TJ, Brown DG, Tanksley SD (2001) The origin of genomic duplications in Arabidopsis. Science 290:2114–2117

Walsh JB (1995) How often do duplicated genes evolve new functions. Genetics 139:421–428

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F (2000) TRANSFAC: An integrated system for gene expression regulation. Nucleic Acids Res 28:316–319

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and

dates of divergence between Brassica and other angiosperm lineages. J Mol Evol 48:597–604

Yu J, Hu S, Wang J, et al. (2002) A draft sequence of the rice genome (*Oyza sataiva* L. ssp. *Indica*). Science 296:79–91

Yunes JA, Vettore AL, da Silva MJ, Leite A, Arruda P (1998) Cooperative DNA binding and sequence discrimination by the Opaque2 bZIP fector. Plant Cell 10:1941–1955

Zhang J, Rosenberg HF, Nei M (1998) Positive darwinian selection after gene duplication in primate ribonuclease gene. Proc Natl Acad Sci USA 95:3708–3713

Zhu X, Tang G, Granier F, Bouchez D, Galili G (2001) A T-DNA insertion knockout of the bifunctional lysine-ketoglutarate reductase/saccharopine dehydrogenase gene elevates lysine levels in Arabidopsis seeds. Plant Physiol 126:1539–1545

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Plant Science xxx (2004) xxx–xxx
PLANT SCIENCE
www.elsevier.com/locate/plantsci

Expression, purification and characterization of a novel bZIP protein from sugarcane

Paulo Sérgio Schlögl [a], Jörg Kobarg [b], Vitor Hugo Moreau [c], Adilson Leite [a,✠], Adão A. Sabino [d], Marcos N. Eberlin [d], Marcelo Menossi [a,*]

[a] *Departamento de Genética e Evolução, Instituto de Biologia, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, CP 6010, 13083-875 Campinas, SP, Brazil*
[b] *Laboratório Nacional de Luz Síncrotron, Centro de Biologia Molecular Estrutural, Rua Giuseppe Máximo Scolfaro 10.000, CP 6192, 13084-971 Campinas, SP, Brazil*
[c] *Núcleo de Biotecnologica (NuBioTec), Faculdade de Tecnologia e Ciências (FTC), Av. Luis Vianna Filho, CP 8812, Pituaçu, 41820-785, Salvador, Brazil*
[d] *Laboratório de Espectrometria de Massa Thompson, Instituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas, SP, Brazil*

Received 12 January 2004; received in revised form 30 April 2004; accepted 2 May 2004

Abstract

The basic leucine zipper (bZIP) proteins form a large family of transcriptional factors in plants and other eukaryotes. Plant bZIP transcriptional factors are divided into subfamilies and are involved in regulating a large range of physiological processes, from plant development to responses to biotic and abiotic stimuli. In this work, we cloned a novel bZIP of sugarcane into the pET3c vector and expressed the recombinant SCbZIP1 (66-170) protein in *Escherichia coli* BL21 (DE3) plysS. The recombinant protein was purified by heat-treatment and reversed phase chromatography. Northern blot analysis showed that *SCbZIP1* was expressed early in development on day 4, but was not induced by abscisic acid (ABA) or exposure to cold. The identity of the recombinant protein was confirmed by mass spectrometry and CD spectroscopy showed an alpha-helical content of 33%. Electrophoretic mobility assays showed that SCbZIP1 (66-170) bound strongly to G-box, Hex and C-box DNA motifs. SCbZIP1 (66-170) was phosphorylated in vitro by a series of protein kinases and its DNA-binding affinity was strongly decreased after phosphorylation by CKII. SCbZIP1 (66-170) also underwent homo- and heterodimerization with truncated forms of the bZIP transcription factor Opaque 2 from *Coix*.

Keywords: Plant bZIPs; Binding affinity; G-box DNA motifs; Phosphorylation; Dimerization

1. Introduction

Transcription factors regulate the expression of genes and are involved in the control of many intracellular processes. These factors can be defined as sequence-specific, DNA-binding proteins that recognize regulatory sequences in the promoter of a gene and are capable of modulating transcription via interaction with basal components of the transcriptional machinery [1]. Structural and functional studies of transcription factors have shown that they have a modular protein structure consisting of DNA-binding and transcription activation domains, and usually contain nuclear localization signals (NLS) and dimerization or multimerization domains [2–5]. Most transcription factors can be grouped into a handful of different gene families based on their type of DNA-binding domain. The members of each family interact with related or identical DNA motifs found in a variety of different promoters which are regulated by external stimuli, in an endogenous, tissue-specific manner

✠ This paper is dedicated to the memory of Dr. Adilson Leite (08.04.1960–28.02.2003), who passed away after a long fight with cancer before this project was finished. This work would not have been possible without the insight, encouragement and collaboration provided by Dr. Adilson Leite, who will always be remembered as an enthusiastic researcher and an amiable colleague.

* Corresponding author. Tel.: +55 19 3788 1143; fax: +55 19 3788 1089.
E-mail address: menossi@unicamp.br (M. Menossi).

0168-9452/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2004.05.002

or by in response to phytohormones such as abscisic acid (ABA) [6,7].

Members of the basic leucine zipper (bZIP) family of transcriptional regulators have been described in eukaryotes and are characterized by a conserved region rich in basic amino acid residues that binds to the target DNA and contains the NLS [3]. Close to the basic region there is a leucine zipper region which consists of several heptad repeats of hydrophobic residues. The leucine zipper region is alpha-helical and prone to dimer formation via a coiled-coil arrangement [3]. The crystal structure of the bZIP domain of the yeast GCN4 protein complexed with DNA demonstrated that the bZIP motif resembled a helical forceps "gripping" the major groove of the DNA [8].

Mutational analyses and domain swapping studies have shown that the DNA-binding specificity of bZIP proteins is determined primarily by the amino acid sequence of the basic region [9–11]. However, the hinge and the leucine zipper regions, as well as residues outside the basic and leucine zipper domains, may also be involved in determining the binding specificity by juxtaposition of the basic region or by other mechanisms [6]. Thus, the sequence of the basic region alone is insufficient for predicting the DNA-binding specificity of this class of proteins [11,12]. Genetic and biochemical studies of bZIPs in plants have shown that they are important regulators of several processes, including the responses to hormones, light, and developmental programs [6,7].

A phylogenetic classification of 50 higher plant bZIP factors has been reported [13]. The *Arabidopsis* genome contains 75 members of the bZIP family classified into 13 sub-families [14]. Another classification has been obtained based on the sequence similarity of the basic region and shared domains [6], and the results are in agreement with the phylogenetic classification. Comparison of the amino acid sequence of the basic regions of plant bZIPs has shown that they are more similar to each other than to bZIPs from other organisms, thus indicating that they are an evolutionary-related subfamily [13].

Consistent with this high level of conservation, all plant bZIPs can bind sequences with a conserved ACGT core, even though they may do so with different affinities [15–18]. To facilitate comparisons among different eukaryotes, the nomenclature used for the DNA-binding sequence of the yeast transcription factor GCN4 has been adopted, in which the two central nucleotides, C and G, in the ACGT element are designated −0 and +0, respectively [19]. The nucleotide at +2 defines the box class, with many plant bZIPs binding preferentially to either G- (ACGTG), A- (ACGTA) or C-boxes (ACGTC), all of which are present in the promoters of a wide variety of plant genes [17,18]. There are two classes of G-boxes and of the corresponding G-box-binding factors (GBFs) in cauliflower nuclear extract [15]. Type A activity interacts with the class I G-box, whereas type B activity binds to the class II G-box. The positions ±3 define the class of the G-box, with the palindromic class I having either C or A at −3, while the class II element has G or T at −3. The different affinity observed for each class is dictated primarily by the nucleotides at positions ±4 [17]. The ACGT element is necessary for maximal transcriptional activation, and plant bZIPs have been classified according to their binding affinities for ACGT elements and their flanking sequences [17]. Thus, characterization of the specificity of a newly isolated DNA-binding protein is a key prerequisite for understanding its physiological function.

All bZIPs can form homo and/or heterodimers, but the mechanisms that determine whether homo or heterodimerization occurs are still poorly understood. Nevertheless, the amphipathic nature of the leucine zipper domain favors the hypothesis that charged or polar residues at positions *a*, *e* and *g* of the heptad of one monomer interact with the corresponding residues at the same position on the opposite monomer, probably via hydrogen bonds [20,21].

After data-mining the Sugarcane EST Genome Project (SUCEST; http://sucest.lad.ic.unicamp.br) [22] for possible sugarcane bZIPs, we cloned and expressed the bZIP region of a novel bZIP family protein, which we named SCbZIP1. We confirmed the correct molecular weight of this novel protein by mass spectroscopy of tryptic peptides and assessed the secondary structure content of the recombinant protein by far UV CD spectroscopy. We also studied the phosphorylation of SCbZIP1 in vitro, its binding to different DNA probes via gel-shift assays and its hetero-dimerization with different truncated forms of Opaque 2 from *Coix lacryma-job*. SCbZIP1 binds strongly to G-box DNA elements, has a basic region and a leucine zipper motif with at least eight heptad repeats, and lacks some motifs that are well conserved in other G-box-binding factors. Northern blot analysis showed that this new sugarcane bZIP was expressed in the early stages of plant development and was not induced by abscisic acid or exposure to cold.

2. Materials and methods

2.1. Bioinformatic

Forty-three plant G-box-binding factors were aligned using 70 amino acid residues corresponding to the basic and leucine zipper domains containing at least six heptad repeats. The protein sequences were aligned with the CLUSTAL X program [23]. The full length protein was used to search for motifs with the MEME program (http://meme.sdsc.edu/meme/website). The theoretical protease digestion profile, the determination of the peptide masses (http://ca.expasy.org/tools/peptide-mass.html) and the theoretical p*I* calculations were done using the ExPASy Proteomic Tools (http://ca.expasy.org/tools/pi_tool.html). The putative phosphorylation sites were predicted using the NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/).

2.2. Plant treatment and RNA extraction

Sugarcane plantlets (*Saccharum* sp. cv SP80-3280) employed in this works were grown like described previously [24]. The cold exposure was done according described formerly [24]. The ABA treatment was done adding 100 μM of ABA to the medium and the leaves of the plantlets were harvested 0, 3, 6 and 12 h afterwards. Total RNA was isolated from different plant organs of the sugarcane plantlets using Trizol reagent (GibcoBRL, USA) according to the manufacturer's instructions to analyze expression of SCbZIP1 transcripts.

2.3. RNA blot analysis

Ten micrograms of total RNA were run on a 1% (w/v) agarose gel containing formaldehyde and transferred to a Hybond-N+ filter (Amersham Pharmacia Biotech., USA) as described elsewhere [25]. The filters were hybridized with an α-^{32}P dCTP radiolabeled *SCbZIP1* cDNA at 42 °C. Blots were washed at high stringency and exposed to imaging plates. Signals of the digitalized images were quantified using the Image Gauge software (Fuji Film, Japan).

2.4. Cloning and expression vector construction

A cDNA encoding a truncated polypeptide of 12.5 kDa, representing the bZIP motif of SCbZIP1 [=SCbZIP (66-170)] was amplified by a standard polymerase chain reaction (PCR) using DNA of the EST clone isolated from the SUCEST cDNA-library (SUCEST; http://sucest.lad.ic.unicamp.br). Primers containing the cloning restriction sites (in bold) for NDEI and *Bam*HI in the sense and anti-sense primers were used: 5′-CAAATACCATATGGAGGAGTC-3′ and 5′-CTGGATCCTTCAAGAGTCAG-3′. The amplified cDNA insert was cloned into the pET3c expression vector (Novagen, USA).

2.5. Protein expression and purification

Escherichia coli BL21 (DE3) plysS cells were transformed with the pET3c vector containing the SCbZIP1 (66-170) cDNA. The bacteria were cultured in 50 mL of 2× YT broth [25] with 50 μg of ampicillin/mL and grown overnight at 37 °C and 300 rpm. The pre-culture was transferred to 2 L of 2× YT broth with ampicillin and grown to an OD_{600} of 0.6 at 37 °C and 300 rpm. The culture was then induced for protein expression by the addition of 50 mM lactose (Merck, Germany) for 4 h. The cells were harvested, resuspended and lysed in: 50 mM Tris–HCl pH 7.5, containing 5 mM EDTA (Gibco-BRL, USA), 5 mM Benzamidine (Sigma, USA) and 5 mM DTT (Gibco-BRL, USA). After five cycles of sonication the bacterial extract was centrifuged at 13,000 × *g* for 30 min at 4 °C. The supernatant was heated to 80 °C for 3 min and centrifuged at 13,000 × *g* for 30 min at 4 °C. Ammonium sulfate was added to the supernatant to a final concentration of 80%, which was then stirred at 4 °C overnight. After a further centrifugation at 13,000 × *g* for 30 min at 4 °C, the pellet was dissolved in 20 mM Tris–HCl, pH 7.5, containing 5 mM EDTA, 2 mM DTT, 1 M NaCl and 5% of polyethyleneimine (PEI; Sigma, USA) and then stirred at 4 °C for 15 min and centrifuged again at 13,000 × *g* for 30 min at 4 °C. The ammonium sulfate precipitation step was repeated three times to ensure total removal of the PEI. The final pellet was dissolved in 20 mM Tris–HCl, pH 7.5, containing 5 mM EDTA, 2 mM DTT and 80% ammonium sulfate. The supernatant was centrifuged, filtered and applied to an AP-1 column (10 mm × 100 mm, Waters, USA) filled with 6 mL of Poros 50R2 resin (PerSeptive Biosystems, USA) for further purification by reversed phase chromatography. The column was washed and equilibrated with 50 mL of buffer A [2% acetonitrile and 0.065% trifluoroacetic acid (both from Merck, Germany)]. The protein was eluted with a gradient of acetonitrile (2–100%). The fractions obtained in all pre-purification steps and the chromatographic fractions were analyzed by SDS–PAGE. The Opaque 2 proteins from *Coix* [O2_BL (180–319) and O2_L (239–319)] were generated and expressed as described in a previous report [14]. The expression and purification steps were the same as described here for SCbZIP1. Lyophilized SCbZIP1 protein was dissolved in 500 μl of water and the protein concentration was determined by the Bradford assay [26].

2.6. MALDI-TOF

The recombinant protein (1.5 μg) was analyzed by mass spectrometry (MS). MS was done using matrix-assisted laser desorption ionization (MALDI) and time-of-flight (TOF) mass measurements in a MALDI LR spectrometer (Micromass, UK) operated in the reflector mode (10,000 *m/z* resolution with 50 ppm mass accuracy). The recombinant protein was desalted with a C18 ZipTip pipette tip (Millipore, USA), eluted with acetonitrile-TFA and then co-crystallized on a MALDI target plate with α-cyano-4-hydroxycinnamic acid (CHCA). The mass spectrum was acquired in the positive ion mode. Recombinant protein (1.5 μg) was also digested with 1 μg of trypsin (Sigma, USA) in 10 mM ammonium bicarbonate at 37 °C for 18 h and then analyzed by mass spectrometry (MS) under the same conditions as for the intact recombinant protein.

2.7. Circular dichroism spectroscopy

Circular dichroism (CD) experiments were done using a JASCO J-715 (Jasco, USA) spectropolarimeter with a 2 nm bandwidth and an optical path length of 0.2 cm. For thermal denaturation experiments, the temperature ramp (1 °C/min) was produced with a Peltier unit.

2.8. Phosphorylation

One microgram of SCbZIP1 was incubated with 20 mM Tris–HCl, pH 7.5, containing 10 mM $MgCl_2$, 1 mM DTT, 2 μCi [γ-^{32}P]-ATP, 50 nM of unlabeled ATP, and 1 μl of protein kinase in a final volume of 10 μl. For the kinase P-34^{cdc2} (Calbiochem, USA) the buffer was supplemented with 1 mM EGTA and 0.01% Brij. The reactions were done at 30 °C for 30 min and were then heated to 100 °C for 3 min with Laemmli buffer [27]. All samples were electrophoresed by SDS–PAGE using 15% gels that were subsequently stained with Coomassie blue, dried and autoradiographed.

2.9. Electrophoretic mobility shift assay (EMSA)

Single strand sense oligonucleotides were synthesized (Invitrogen, USA) and the sequences are indicated in Fig. 6, panel I. The bold region in each of these sequences corresponds to the core motif that is highly conserved in most plant bZIP DNA targets (Fig. 6, panel I). The anti-sense oligonucleotide used to produce a double-strand radiolabeled DNA oligonucleotide for the electrophoretic mobility shift assays (EMSA) was designed as follows: 3′-TGAGCTC-5′. One nmol of each of the sense anti-sense oligonucleotides were mixed with 10 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA and heated to 94 °C and gradually cooled to room temperature. The annealed probes were extended with Klenow enzyme (Gibco-BRL, USA) in the presence of ^{32}P-labelled dCTP and were purified on Sephadex® G-25 (NAP^{TM}-5, AP-Biotech, USA) columns.

The protein–DNA-binding reaction was done by mixing different amounts of SCbZIP1 protein (15–500 ng) with 18 pmol of radiolabeled probe and 100 ng of salmon sperm DNA with gel-shift buffer (150 mM Tris–HCl, pH 7.5, 250 mM NaCl, 2.5 mM EDTA, 5 mM DTT, 10% glycerol and 10% Ficoll) for 30 min at 30 °C. After formation of the complex 5 μL of running buffer (250 mM Tris–HCl pH 7.5, 40% glycerol, 0.2% bromophenol blue and 0.2% xylene cyanol) were added and the samples were run on a 6% PAGE with 1% glycerol in 0.25 × TBE buffer at 400 V and 4 °C for 1 h. The gel was dried and autoradiographed. The competitive assays were done by mixing 150 ng of SCbZIP1 with 18 pmol of radiolabeled probe and a 20, 30, 40 or 50-fold molar excess of non-radiolabeled probe using the same conditions as described above. The EMSA with phosphorylated SCbZIP1 was done using the G-box1 and C-box1 ^{32}P-labeled probes under the same conditions.

2.10. Dimerization reactions

For the homodimerization reactions, 1 μg of SCbZIP1 was mixed with 0.5 mM of BS3 (bis-sulfosuccimidyl suberate, Pierce, USA) in dimerization buffer (100 mM Hepes pH 7.5, 50 mM $MgCl_2$, 5 mM DTT) in a final volume of 10 μL and incubated for 30 min at 30 °C. For the heterodimerization reactions 1 μg of the O2 proteins was added. The reaction products were separated by SDS–PAGE on 15% gels that were subsequently stained with Coomassie blue and then documented with an Eagle Eye II system (Stratagene, USA).

3. Results

3.1. Purification and characterization of SCbZIP1

The strategy of data mining was used to identify the putative sugarcane bZIP proteins in the SUCEST program, as described previously [14]. As an initial target for further detailed studies, we chose to clone the cluster SCSGAM1095E12 because of its long leucine zipper domain and because of its expression in the lateral bud during the early stages of plantlet development.

The cDNA sequence of the clone SCSGAM1095E12 coded for a small protein with 170 amino acid residues (20 kDa) and a theoretical p*I* of 7. The encoded protein contained a leucine zipper with at least eight heptad repeats $(abcdefg)_8$, in which the leucine was replaced by valine in the 4th and 5th heptads at position "d" (Figs. 1 and 2A). The bZIP domain was located close to the carboxy-terminal end. Based on the DNA sequencing results and on computer searches for related proteins in GenBank, the corresponding cDNA appeared to be full length and shared sequence identity (given in %) with the clone AAK25822 from *Phaseolus vulgaris* (43%), clone AAK01953 from *Phaseolus acutifolius* (43%), AtbZIP58 (51%), AtbZIP42 (55%), and AtbZIP48 (39%) from *Arabidopsis*. According to the phylogenetic classification [14], all these plant bZIP proteins, including SCbZIP1, belong to the same family. These proteins are not yet well characterized functionally, but they are small proteins with little homology outside their basic region (data not shown). Compared to other characterized plant bZIPs (e.g. G-box-binding factors), SCbZIP1 lacks many of the conserved motifs such a proline-rich transcriptional activation domain [28]. Nevertheless, SCbZIP1 showed seven non-clustered Pro residues in its N-terminal domain (Fig. 1, residues 1–65).

Alignment of the bZIP motifs showed that all bZIPs had a well-conserved basic region with a more variable leucine zipper domain (Fig. 2A). The leucine residues were well conserved at position "d" in the heptad repeats of the zipper domain in the most of the bZIP proteins, except for the 4th and 5th heptads that had Phe, Ile, Ala or Val at this position (Fig. 2A). SCbZIP1 had two Val at these positions. Fig. 2B shows the putative nuclear localization signal (NLS), the amino acid residues that contact DNA, the sites for the regulation of DNA binding by phosphorylation, the amino acid residues indispensable for DNA binding, and the putative phosphorylation target sites for CKII and PKC.

The bZIP region (amino acids 66-170) of the cDNA clone SCSGAM1095E12 was sub-cloned in pET3c. SCbZIP1 (66-170) was expressed and purified by a procedure

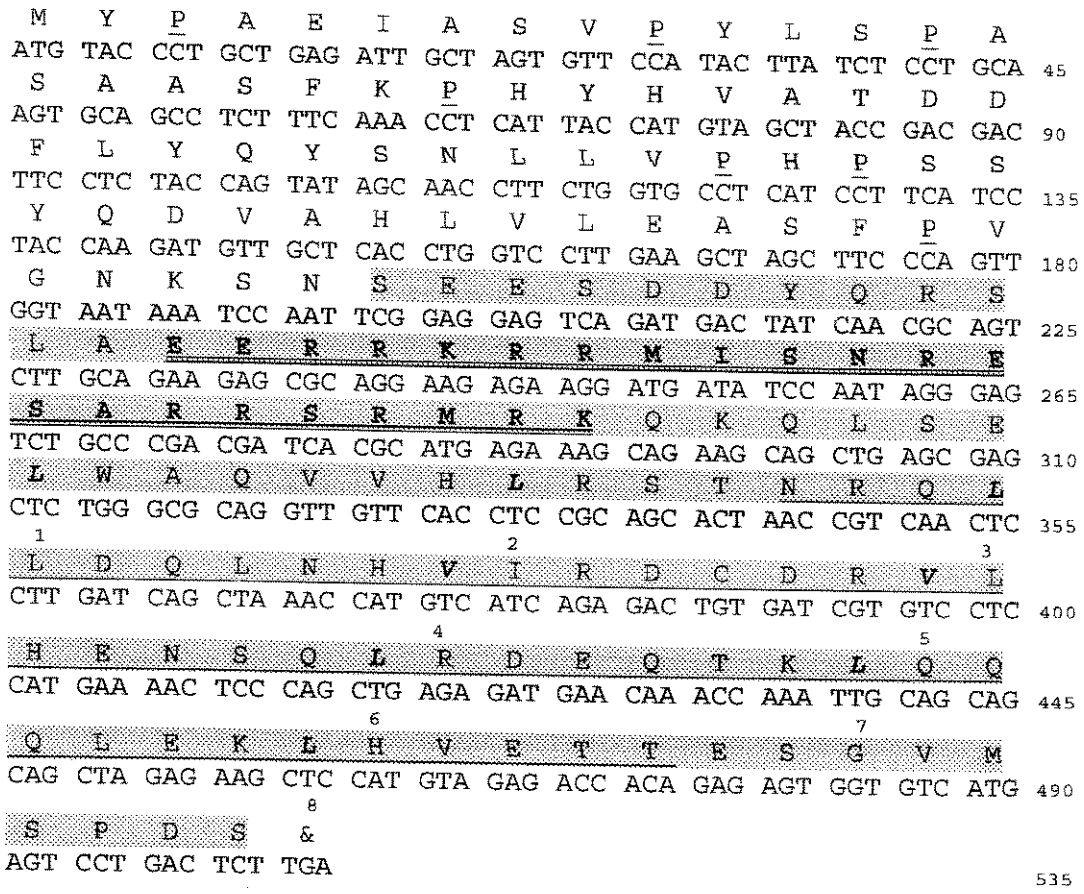


Fig. 1. Sequence of the *SCbZIP1* cDNA and predicted amino acid sequence. Amino acids corresponding to the expressed protein fragment SCbZIP1 (66-170) are highlighted in gray. The basic region is shown in bold type with double underlining. The Leu and Val of the putative zipper domain are numbered and shown in italics. The region with single underlining corresponds to the putative coiled coil region. The prolines of the putative trans-activation domain are underlined.

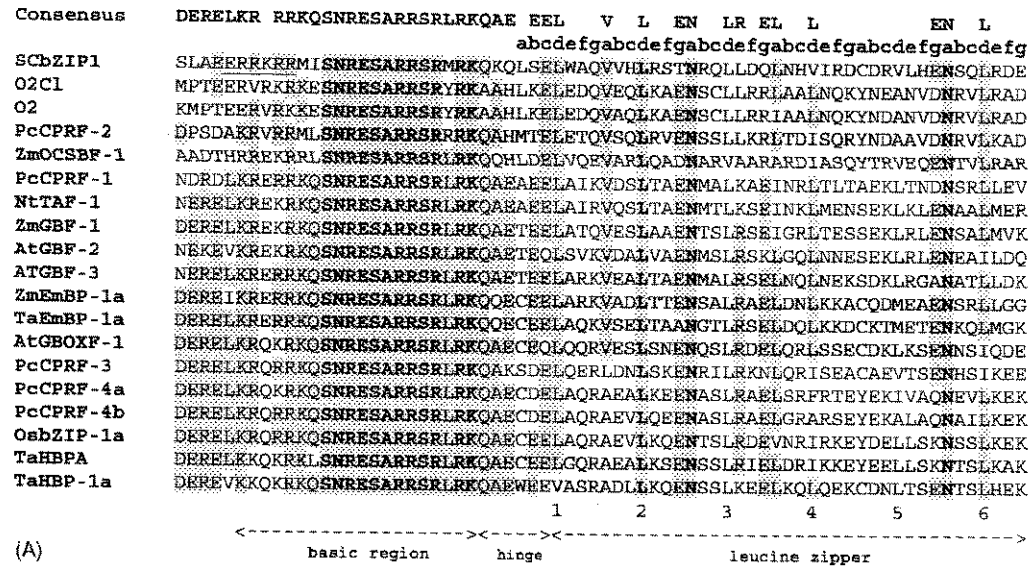
that allowed the rapid, efficient purification of relatively thermo-stable bZIPs with 90% purity. Fig. 3 shows that the putative SCbZIP1 protein fragment appeared almost as a single major band after the various purification steps (lane 7).

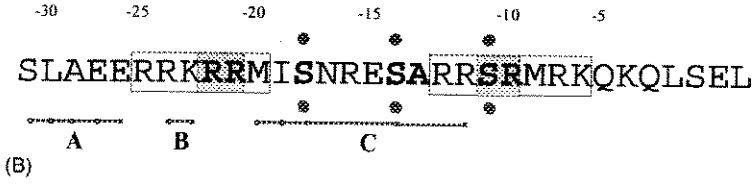
3.2. MALDI-TOF analysis

To determine the molecular weight and identity of the recombinant protein, a MALDI-TOF analysis was done using undigested and trypsin-digested protein. The primary sequence of SCbZIP1 was analyzed using the program PeptideMass (http://ca.expasy.org/tools/peptide-mass.html) to obtain the theoretical peptide masses expected from trypsin digestion, and the predicted masses were compared with the experimental data. The molecular mass was determined to be 12,525 Da, which is close to the theoretical molecular mass calculated from the primary amino acid sequence cloned (12,529 Da). The molecular weight determined by SDS–PAGE (~14.7 kDa) was larger than obtained by MALDI-TOF and the discrepancy between these values maybe reflects technical differences between the methods. The mass spectrum resulting from MALDI-TOF analysis of the digested protein is shown in Fig. 4A. Three of the expected peptides resulting from the trypsin digestion of SCbZIP1 were clearly detected in their monoprotonated form: VLHENSQLR at *m/z* 1095.5, QLLDQLNHVIR at *m/z* 1348.8 and LHVETTESGVMSPDS at *m/z* 1579.1 (Fig. 4A).

3.3. Circular dichroism

The structural integrity and self-association of SCbZIP1 (66-170) were analyzed by far-UV CD spectroscopy. SCbZIP1 was found to be approximately 33% alpha-helical based on the CD ellipticity minimum at 222 nm (Fig. 4B). The helical content of SCbZIP1 (66-170) corresponded to about 45 amino acid residues in contrast to the 56 residues predicted by primary sequence analysis of the corresponding region. This suggested that the SCbZIP1 leucine zipper domain may not be fully structured or may present some conformational fluctuations, as previously proposed for other eukaryotic bZIP transcription factors [29, V. H. Moreau, unpublished observation]. The thermal denaturation of SCbZIP1 (66-170) was accompanied by a decrease





	-25 -20 -15 -10 -5	
EmBP-1a	DEREIKRERRKQSNRESARRSRLRKQQECEEL	group 1
CPRF-1	NDRDLKRERRKQSNRESARRSRLRKQAEAEEL	group 1
TAF-1	NERELKREKRKQSNRESARRSRLRKQAEAEEL	group 1
CPRF-3	DERELKRQRRKQSNRESARRSRLRKQAKSDEL	group 1
OCSBF-1	AADTHRREKRRLSNRESARRSRLRKQQHLDEL	group 1
CPRF-2	DPSDAKRVRRMLSNRESARRSRRRKQAHMTEL	group 2
O2	KMPTEERVRKKESNRESARRSRYRKAAHLKEL	group 2
SCbZIP1	SLAEERRKRRMISNRESARRSRMRKQKQLSEL	group 2
TGA1a	SKPVEKVLRRLAQNREAARKSRLRKKAYVQQL	group 3

- Bold letters in gray indicate the amino acids that could be the major determinants of plant bZIP DNA binding specificity.

(C)

Fig. 2. Sequence comparisons of plant bZIPs. (A) Alignment of bZIP regions including up to the first six heptad repeats of the leucine zipper domain. Only 19 of the 43 aligned sequences are shown in the figure. The consensus sequence is shown on the top of the alignment. The heptads are depicted according to Lupas [45]. Partially conserved amino acids are highlighted in gray. Amino acids conserved throughout all sequences are shown in bold letters. The "d" positions of the Leu or other hydrophobic residues are numbered. (B) Schematic depiction of the basic region of SCbZIP1. The open boxes contain the bipartite nuclear localization signal (BR-A left, BR-B right); the grey highlighted boxes and bold case letters indicate the amino acids in contact with DNA; the black points (●) below and above the letters indicate residues involved in the putative regulation of DNA binding/nuclear shuttling by phosphorylation; black dots and lines below sequence identify residues indispensable for DNA binding. (C) Sequence comparison between different classes of plant bZIPs. Bold letters in gray indicate the amino acids that could be the determinants of DNA-binding specificity.

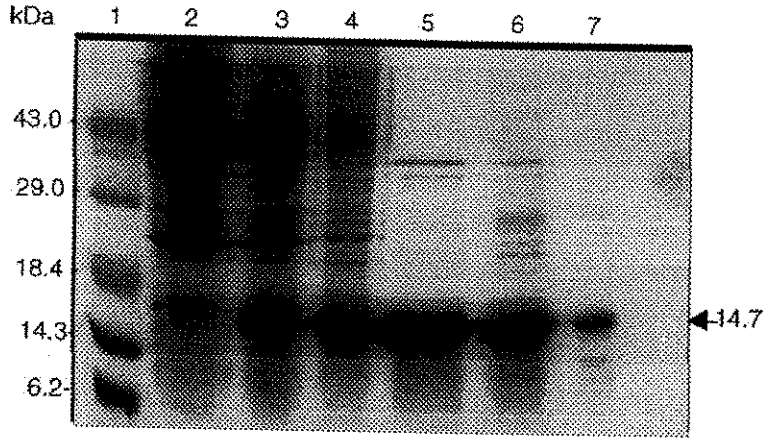


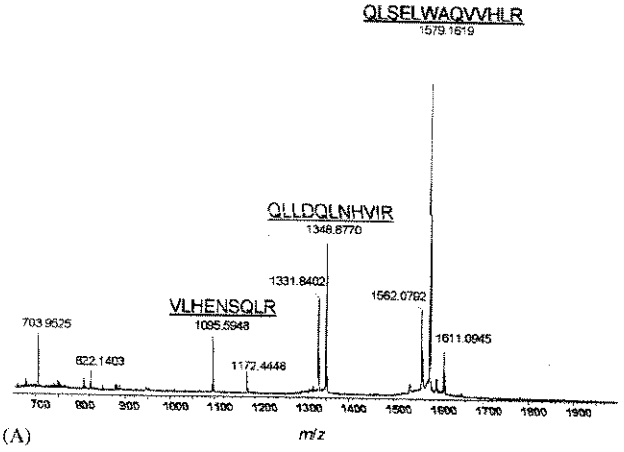
Fig. 3. Expression and purification of SCbZIP1 (60-170). Fractions from the different purification steps were subjected to SDS–PAGE in 15% gel and stained with Coomassie blue. Lane 1: low molecular mass protein markers (Gibco-BRL, USA); lane 2: non-induced *E. coli* extract; lane 3: lactose-induced *E. coli* extract; lane 4: soluble fraction of lactose-induced *E. coli* extract; lane 5: soluble, lactose-induced *E. coli* extract, heated to 80 °C; lane 6: same extract after 80% ammonium sulfate precipitation; lane 7: purified protein after reverse phase chromatography (peak fraction). The arrow indicates the band corresponding to SCbZIP1 (66-170).

in the negative signal minimum at 222 nm, and displayed a cooperative transition that was dependent on the protein concentration in the range between 8 and 80 μM, as predicted by the law of mass action for dissociation processes (Fig. 4B). At the melting temperature, SCbZIP1 (66-170) showed a linear loss of the negative CD signal. This process is characteristic of leucine zippers and has been attributed to fraying of the extremities of the zipper [29–31].

The formation of coiled-coil structures was assessed using the COILS-Prediction of Coiled Coil Regions in Proteins program (http://www.ch.embnet.org/software/COILS_form.html). The result showed that SCbZIP1 contained at least six repetitions of amino acids with a high probability of forming coiled-coils in the leucine zipper domain, which is a common characteristic of bZIP transcriptional factors (Fig. 1).

3.4. Protein phosphorylation

The phosphorylation of transcription factors is a common modification which can regulate their functional activities, including multimerization, compartmentalization or DNA binding [6,32–34]. Fig. 5 shows the results of the phosphorylation of SCbZIP1 (66-170) with different protein kinases and sugarcane nuclear extract. With the exception of PKA (lane 1), all the other kinases and the nuclear extract were capable of phosphorylating the recombinant protein SCbZIP1 (66-170) (lanes 2-6). Although SCbZIP1 (66-170) was not phosphorylated by PKA, we found at least four putative theoretical sites for phosphorylation by this kinase using the NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) (data not shown). PKC was the most efficient of the kinases



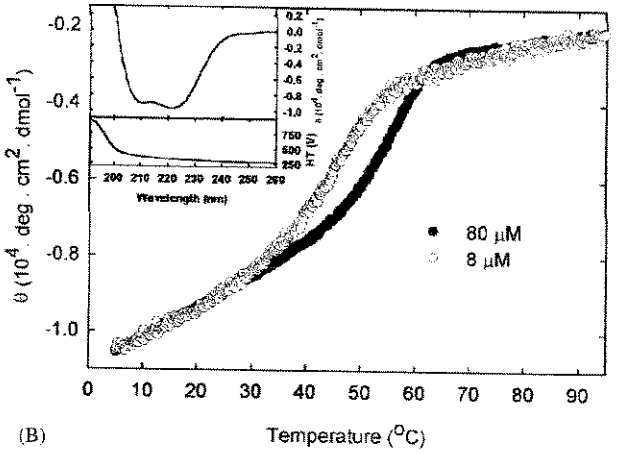


Fig. 4. MALDI-TOF and circular dichroism analysis of the recombinant protein SCbZIP1 (66-170). (A) The masses of monoisotopic peaks with a relative intensity greater than 5% of the most intense peak in the spectrum were used for comparison to a theoretical digestion of the protein by trypsin. The peaks that fitted the theoretical peptide masses are shown, together with the sequences of the peptides. (B) Thermal denaturation of 8 μM (open symbols) or 80 μM (closed symbols) SCbZIP1 (66-170) was followed by the decrease in the negative ellipticity band at 222 nm as described in Section 2. The inset shows the far-UV CD spectrum of 8 μM SCbZIP1 (66-170) at 10 °C and the horizontal transmittance (HT) compensation in the photomultiplier. The raw ellipticity values were corrected to absolute ellipticity; and the CD spectra for 8 and 80 μM SCbZIP1 (66-170) at 10 °C were superimposed.

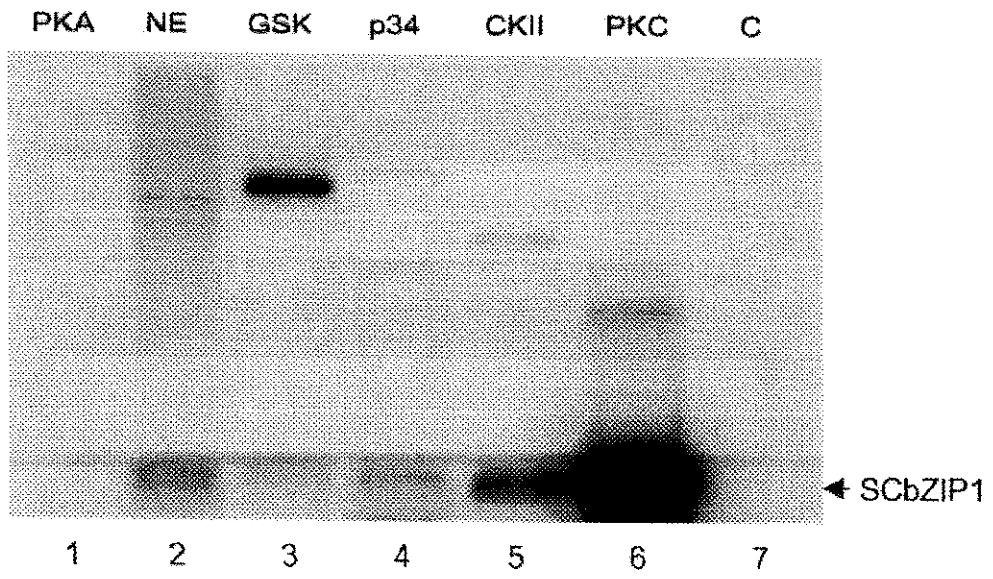


Fig. 5. In vitro phosphorylation of the recombinant SCbZIP1 (66-170). Phosphorylation reactions were done as described in Section 2. Radiolabeled phospho-proteins were separated by SDS–PAGE 15% gels. The gel was dried and analyzed by autoradiography. Lane 1: PKA, bovine protein kinase A; lane 2: NE, sugarcane leaf nuclear extract; lane 3: GSK, human protein kinase α-3 glycogen synthase; lane 4: p34, human cyclin B-dependent protein kinase; lane 5: CKII, *Zea mays* casein protein kinase type II; lane 6: PKC, rat protein kinase C; lane 7: C, negative control reaction without the protein CbZIP1. The arrowhead indicates the position of the phosphorylated SCbZIP1 (66-170).

in phosphorylating SCbZIP1 (66-170) and produced a band of high intensity (Fig. 5, lane 6). This could be due to the presence of two putative PKC sites in the basic region of SCbZIP1 (RMISNRE and NRESARRS) and a third site in the second heptad repeat (LRSTNRQ) (Fig. 2A and B).

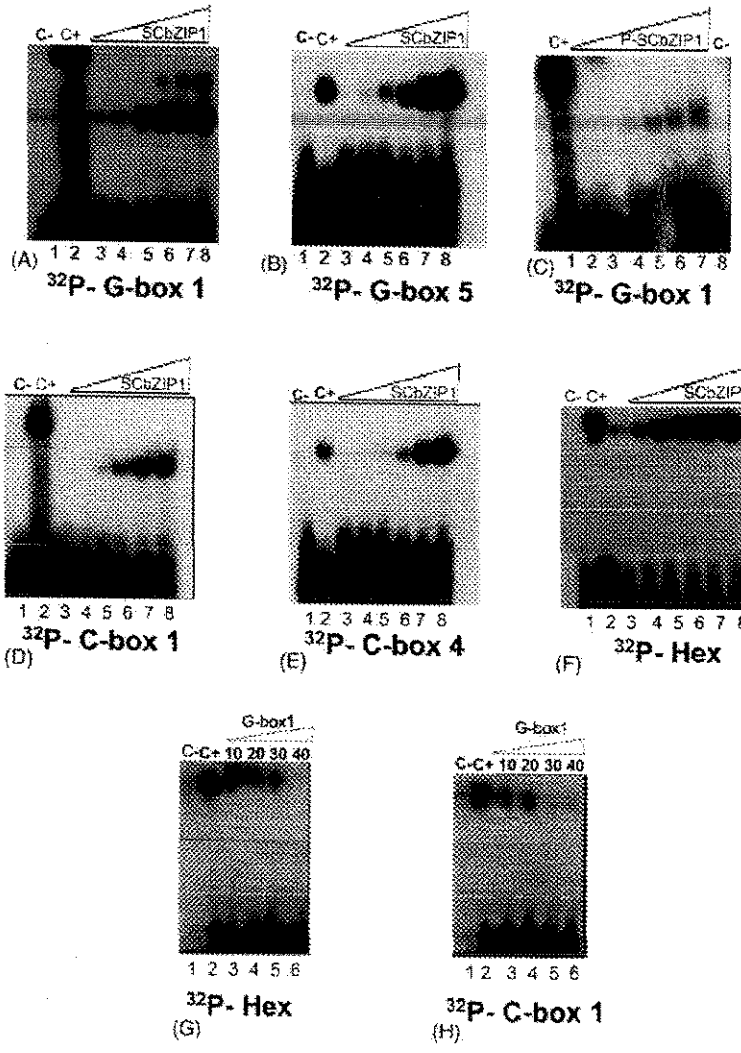
The activity of several bZIP transcription factors is modulated by CKII [4,34–36]. To investigate whether the DNA-binding activity of SCbZIP1 (66-170) could be modulated by CKII, we phosphorylated the recombinant protein and analyzed its DNA-binding capacity by EMSA using the G-box1 probe (Fig. 6). The SCbZIP1-binding affinity for the G-box1 DNA decreased strongly after phosphorylation by CKII. Similar results have also been reported for the HY5 protein from *Arabidpsis* [32]. After phosphorylation, HY5 protein showed a decreased affinity for the light-responsive G-box element of the CHS1 promoter. Opaque 2 was able to bind to the O2 DNA-binding site only in its non-phosphorylated or hypo-phosphorylated form [33]. Our results suggest that the phosphorylation of SCbZIP1 could be a post-translational modification involved in regulating the binding activity or localization of this sugarcane protein in vivo.

3.5. DNA-binding activity

Since many of the DNA sequence elements used to screen for plant bZIP proteins have a common ACGT core sequence [16], we investigated the influence of different sequences flanking the ACGT core (positions ±3 and ±4) in order to characterize the putative DNA targets for SCbZIP1. The nomenclature used to define the nucleotide positions of the dyad symmetrical ACGT-binding site followed that used for the yeast GCN4-binding site [19]. The DNA binding of the SCbZIP1 (66-170) was assayed by EMSA using six G-box sequences, five C-box element sequences [16], and a Hex [16] and an Em1d [37] sequence. Fig. 6 shows selected results of the EMSA experiments. SCbZIP1 (66-170) had a similar affinity for G-box1 and G-box5 (Fig. 6A and B). This was not surprising since the two probes had a very similar DNA sequence, with the only difference being the substitution of a G in G-box1 for a C in G-box5 at position −4 (Fig. 6I). SCbZIP1 (66-170) did not bind any of the other four G-box probes tested (data not shown, Fig. 6I). Thus, SCbZIP1 displayed a binding specificity similar to that described for the type A cauliflower nuclear G-box-binding protein [15].

Other protein–DNA-binding studies have demonstrated that plant bZIPs can also interact with C-box elements [16]. As shown here (Fig. 6D, E, I) two of the five C-box probes tested bound to SCbZIP1 (66-170). These two C-box sequences had two bases changed at positions +4 and −4 (Fig. 6I). SCbZIP1 (66-170) did not bind any of the other three C-boxes tested, although C-box3 has been shown before to bind other plant bZIPs, including Opaque 2, TAF-1, and CPRF-2 [16].

We next examined the interaction of SCbZIP1 with Em1d [37] and Hex [16], two other oligonucleotide probes which also contain the ACGT core sequence but differ in their flanking regions (Fig. 6I). The Hex motif consists of half a G-box and half a C-box and could therefore be a DNA target for bZIP proteins with affinity for both G- and C-box motifs [16]. As expected, SCbZIP1 (66-170) bound with high affinity to Hex (Fig. 6F), but did not bind to the Em1d probe (data not shown).



Probe	Sequence	Binding by SCbZIP1
GBI	AAACTTGACCACGTGCAACTCGAC	+
GBII	AAACTTGACTCACGTGACACTCGAC	-
GBIII	AAACTTGACTCACGTGAAACTCGAC	-
GBIV	AAACTTGACTCACGTGCAACTCGAC	-
CBV	AAACTTGACTCACGTCCAACTCGAC	+
GBVI	AAACTTGACTCACGTGAAACTCGAC	-
CBI	AAACTTGACTCACGTCCAACTCGAC	+
CBII	AAACTTGACTCACGTCCAACTCGAC	-
CBIII	AAACTTGACTCACGTCCAACTCGAC	-
CBIV	AAACTTGAATCACGTCCAACTCGAC	+
CBV	AAACTTGACTCACGTCCAACTCGAC	-
hmx	AAACTTGACTCACGTCCAACTCGAC	+
Em	AAACTTGACTCACGTGTTACTCGAC	-

(I)

Fig. 6. Electrophoretic mobility shift assays for the protein SCbZIP1 (66-170) with different types of ACGT core-containing DNA probes. (A–F) Proteins were incubated with the radiolabeled double-stranded DNA oligonucleotide probes indicated at the bottom of each panel (see panel I for the corresponding sequences). Increasing amounts of SCbZIP1 (66-170) were incubated with each probe: lane 1: negative control without protein (C−); lane 2: positive control, 500 ng Opaque 2 protein from *Coix* (C+); lanes 3–8: 15, 30, 70, 150, 300, or 500 ng of SCbZIP1 (66-170) protein. (C): SCbZIP1 (66-170) protein phosphorylated in vitro by CKII before EMSA. G, H: Competitive EMSA: Increasing amounts of unlabeled double-stranded DNA probe were incubated with the protein and the radiolabelled probe. I: The DNA sequences of the probes used in the EMSA experiments with SCbZIP1 (66-170). The numbering of the bases relative to the core ACGT was adapted from the nomenclature for the GCN4 transcriptional factor [19]. The probes bound by SCbZIP1 (66-170) are shown by + and those not by −.

Competitive EMSA experiments were used to further investigate the specificity of SCbZIP1 (66-170) for G-box1, Hex and C-box1. Fig. 6G and H shows the results of the competitive assay with ^{32}P-labeled Hex and C-box1 in the presence of unlabeled G-box1. The binding to Hex and C-box1 was sensitive to 40- and 30-fold excess amounts of G-box1 unlabeled probe, respectively. These results confirmed that SCbZIP1 (66-170) had greater affinity for the G-box1 and Hex probes than for C-box1.

3.6. Dimerization experiments

bZIP proteins are able to form homo and heterodimers. Based on the amphipathic nature of the alpha-helix of bZIPs, the charged or polar amino acids at positions *a*, *e*, and *g* (Fig. 2A) of one bZIP can interact with the corresponding positions on another bZIP through hydrogen bonds or hydrophilic interactions. A chemical cross-linking dimerization assay with the cross-linker BS3 was therefore used to assess the dimerization of SCbZIP1 (66-170) with itself and with two different truncated forms of Opaque 2 protein from *Coix lagrima-job* (Fig. 7).

Each of the three protein constructs tested (SCbZIP1, O2_BL and O2_L) had the ability to form dimers (Fig. 7, lanes 3, 5 and 8), as shown by the appearance of bands corresponding to double the molecular mass (29, 18, 34 kDa) of the respective monomeric forms (15, 9, 17 kDa; Fig. 7B, lanes 2, 4 and 7).

Heterodimerization of SCbZIP1 (66-170) with both O2 constructs was also observed (Fig. 7B, lanes 6 and 9). Both the O2 homodimerization and heterodimerization occurred in the same reaction medium when SCbZIP1 (66-170) and O2_L were incubated together with the cross-linker (Fig. 7, lanes 6). This could indicate that the propensity of O2_L to form a heterodimer with SCbZIP1 (66-170) was superior to that of SCbZIP1 (66-170) to form a homodimer, since only a very weak band of SCbZIP1 (66-170) dimer was seen under these conditions (Fig. 7, lane 6). In contrast, O2_BL did not form a homodimer, but could still engage in heterodimer formation with SCbZIP1 (66-170) (Fig. 7, lane 9).

3.7. SCbZIP1 mRNA expression

The expression of SCbZIP1 mRNA during plantlet development, and after treatment with ABA and exposure to cold, were analyzed by Northern blotting (Fig. 8). The transcripts were expressed in the initial stages of sugarcane plantlet development (Fig. 8A). Expression was also seen in the lateral buds and flowers, but not in any of the other tissues analyzed (Fig. 8B). These findings suggested that SCbZIP1

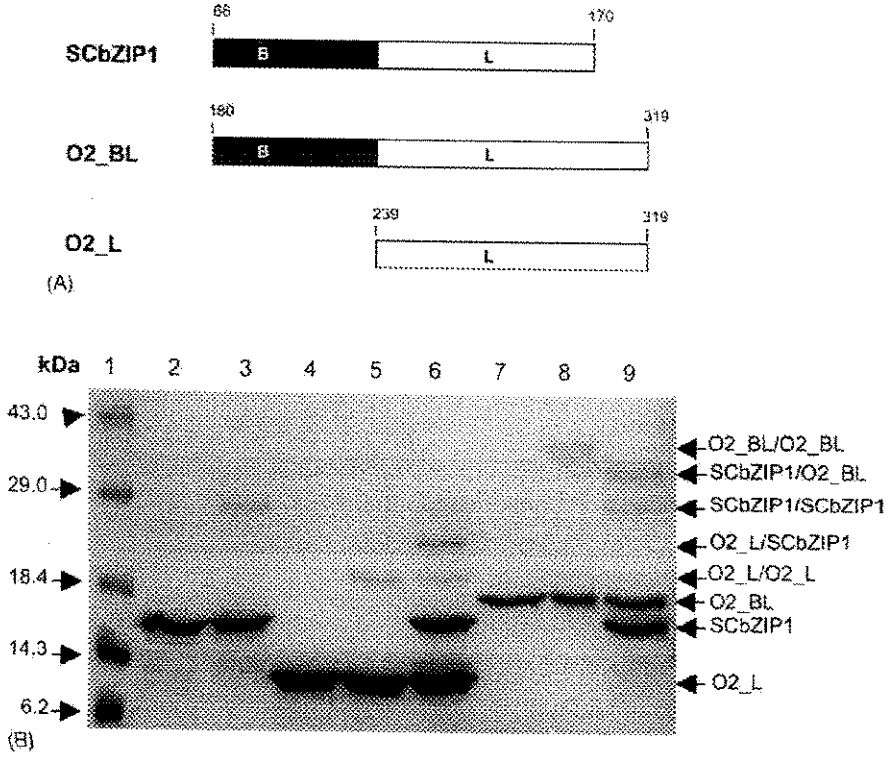


Fig. 7. SCbZIP1 homodimerization and heterodimerization with Opaque 2 from *Coix* in vitro. (A) Schematic representation of the expressed proteins. The basic region (B) is depicted in black and the leucine zipper domain (L) in white. The amino acids spanning the constructs are indicated. (B) SDS–PAGE 15% gel of the proteins or protein-mixtures subjected or not to dimerization in vitro with the chemical cross-linker BS3. Lane 1: low molecular mass protein markers; lane 2: SCbZIP1 without cross-linker; lane 3: SCbZIP1 incubated with 0.5 mM BS3; lane 4: O2_L without BS3; lane 5: O2_L with 0.5 mM BS3; lane 6: SCbZIP1 and O2_L with 0.5 mM BS3; lane 7: O2_BL without BS3; lane 8: O2_BL with 0.5 mM BS3; lane 9: O2_BL and SCbZIP1 with 0.5 mM BS3. The respective protein monomers/dimers are indicated by the arrows.

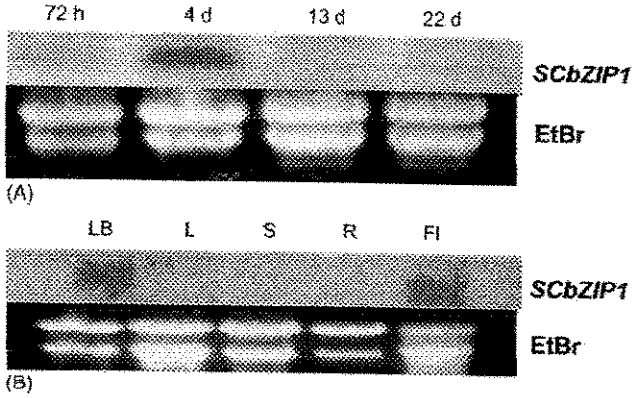


Fig. 8. Northern blot analysis of the expression of *SCbZIP1* mRNA during sugarcane plantlet development and in different plant tissues. (A) Total RNA extracted 72 h, 4, 13 or 22 days after planting. (B) Total RNA extracted from lateral bud (LB), leaves (L), steam (S), roots (R) and flowers (Fl). The total RNA stained with ethidium bromide before being transferred to the membrane is shown under each lane.

could play a role in some stages of plant and flower development. No expression of SCbZIP1 mRNA was observed in leaves from plants treated with ABA or exposed to cold (data not shown), although other plant bZIPs, such ABI5, are induced by abscisic acid or cold exposure [7,38].

4. Discussion

Amino acid sequence analysis of the SCbZIP1 cDNA clone showed that the carboxy terminus contained a long bZIP motif with eight heptad repeats. The CD spectra of SCbZIP1 revealed a α-helical pattern, with a helical content of about 33%. This relatively low helical content may be associated with conformational fluctuations previously reported for bZIP transcriptional factors [29]. Such conformational fluctuations have been proposed to be important in determining the specificity of the formation of heterodimers and, therefore, for biological function [29]. In addition, CD spectra showed that the thermal denaturation of SCbZIP1 (66-170) was dependent on the protein concentration, suggesting that a dissociation reaction occurred (Fig. 4B). The shift observed in the $t_{1/2}$ of the denaturation curves was compatible with the dissociation of a dimeric protein and corroborated the finding that SCbZIP1 (66-170) could form stable dimers in solution. Taken together, CD and MALDI-TOF results confirmed the identity of the recombinant protein and showed that SCbZIP1 (66-170) maintained its secondary structure after the purification protocol which involved denaturation by heating.

The phosphorylation of SCbZIP1 (66-170) by CKII decreased its binding affinity for G-box1 when compared to the non-phosphorylated protein (Fig. 6C). The serine residues conserved in the basic region of all GBF proteins are targets to CKII phosphorylation [4]. The importance of these conserved serines for DNA-binding activity has been studied using gel-shift assays, as in the case of a wheat bZIP transcriptional factor, HBP-1a(17) [39]. All of the serines in this protein were mutated to glutamic acids to mimic the effects of phosphorylation, and resulted in decreased DNA-binding activity, especially when the serines at the PKC sites SAR and RSR were mutated [39]. Thus, the phosphorylation of conserved serines in the basic region would be expected to cause a decrease of DNA-binding activity. Phosphorylation of the basic domain leads to the translocation of GBF to the nucleus where a putative specific phosphatase activity could dephosphorylate the serines at the NLS site, thereby allowing GBFs to bind to G-box DNA target sequences [4]. The serines in the basic region of SCbZIP1 could be regulated by phosphorylation in vivo and therefore be involved in the regulation of nucleo-cytoplasmic translocation and/or DNA binding.

Plant bZIP proteins show specificity in binding DNA sequence elements containing an ACGT core. Gel mobility shift experiments using 10 recombinant plant bZIP proteins have shown that the nucleotides flanking the ACGT core greatly influence the binding specificity and affinity [16]. Three main elements have been identified, namely, A-, C- and G-boxes and bZIPs have been classified into three groups based on a strong affinity for G-boxes (group 1), a similar affinity for G- and C-boxes (group 2), and a stronger affinity to C-boxes (group 3) [16].

The EMSA results showed that SCbZIP1 had a strong affinity for G- and C-boxes, and for Hex, an ACGT element containing G-box/C-box hybrid sites. Analysis of the sequences flanking those G-box bound by SCbZIP1 (66-170) revealed that position −3 was always occupied by a C and there was strong binding only when a G or C was present at position −4. The presence of T or A at position −4 led to a loss of binding activity (data not shown). For plant bZIPs binding to the ACGT elements there was an exclusive preference for C or A at position −3 and a preference for G or T at position −4 in the G-box elements [16]. However, SCbZIP1 (66-170) bound strongly to an ACGT element with C at position −4. This difference in binding preference could reflect sequence differences in the amino acids in the basic region (Fig. 2).

SCbZIP1 (66-170) bound with high affinity to C-box elements containing a T at −3 and an A at −4 in agreement with results obtained by others [16], who showed that most plant bZIPs can bind to DNA sequences containing the same flanking sequences. Furthermore, SCbZIP1 (66-170) was able to bind to the Hex motif, a probe with half a G-box and half a C-box. Thus, based on the classification according to affinity previously reported [16], the SCbZIP1 protein can be classified into group 2 because of its affinity for the G-, C-box and Hex probes.

Our results support the hypothesis that bZIP-binding activity could depend on the binding affinity of protein dimer subunits for ACGT half sites [16]. Although all plant bZIP proteins exhibit differential binding for the high affinity G-

and C-box elements, they all have a similar DNA-binding specificity in that they prefer almost the same flanking sequences. Fig. 2C shows a comparison of the plant bZIP protein DNA-binding basic regions. There are four strictly conserved amino acids that directly contact DNA: Arg −11, Ser −12, Ala −15 and Asn −19 [17]. Arg at −11 is essential for DNA binding because an Opaque 2 Arg to Lys point mutant at this position was unable to bind to promoter of 22 kDa zein genes [40]. Ala −15 could be the major determinant for plant bZIP DNA-binding specificity, and the presence of Ala −15 could differentiate group 3 from groups 1 and 2 [17]. It is not yet clear how bZIP proteins can discriminate between different G-box elements, but it is likely that amino acids present in BR-A (Fig. 2B), or in sequences flanking the basic region or hinge region can affect binding specificity [16,17].

The binding specificities of bZIP proteins in vivo may be regulated by a combination of heterodimerization with other bZIP proteins, accessory proteins, co-factors, chaperones and/or post-translational modification, including phosphorylation [5,16]. The *Arabidopsis* GBF and RSG form heterodimers with other proteins, a mechanism that might have evolved to generate additional functional diversity [41,42]. As shown here, SCbZIP1 (66-170) interacted with two truncated forms of Opaque 2 from *Coix*. These results indicate that SCbZIP1 may interact with other sugarcane transcriptional factors.

GBFs are involved in the regulation of gene expression during plant development. The expression of the ribulose-1, 5-bisphosphate carboxylase small subunit gene (*Rbcs2*), for instance, is regulated during tomato fruit development via a G-box within the *Rbcs2* promoter [43]. The presence of G-box elements in the light-regulated promoters of the gene *Chs* indicates that GBF proteins may participate in light-induced gene activation [44]. *SCbZIP1* mRNA was expressed during the early stages of sugarcane development and the EMSA and phosphorylation results suggested that SCbZIP1 could bind to G-box elements present in various gene promoters. The ability of SCbZIP1 (66-170) to form dimers with other bZIP transcription factors could also affect the affinity for different DNA elements and amplify the participation in various physiological processes in plant development. Future studies should clarify the function of SCbZIP1 in sugarcane development and lead to the characterization of physiologically relevant dimerization partners.

Acknowledgements

The authors thank Eduardo Kiyota and Daniela Stancato for technical assistance and Cláudia Bandeira Kobarg for help with gel-shifts. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through a doctoral fellowship to P.S.S. (process 00/02480-0), and a grant to M.M. (process 01/07546-1).

References

[1] F.C.P. Holstege, R.A. Young, Transcriptional regulation: contending with complexity, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 2–4.
[2] A.D. Frankel, P.S. Kim, Modular structure of transcriptional factors: implications for gene regulation, Cell 65 (1991) 717–719.
[3] H.C. Hurst, in: P. Sheterline (Ed.), Transcription Factors: bZIP "Protein Profile", vol. 1 (issue 1), Academic Press, London, UK, 1995.
[4] Y. Sibéril, P. Doureau, P. Gantet, Plant bZIP G-box-binding factors: modular structure and activation mechanisms, Eur. J. Biochem. 268 (2001) 5655–5666.
[5] M. Kuhlmann, K. Horvay, A. Strathmann, T. Heinekamp, U. Fisher, S. Bötner, W. Dröge-Lesser, The α-helical D1 domain of the tobacco bZIP transcription factor BZI-1 interacts with the ankyrin-repeat protein ANK1 and is important for BZI-1 function, both in auxin signaling and pathogen response, J. Biol. Chem. 278 (2003) 8786–8794.
[6] M. Jakoby, B. Weisshaar, W. Dröger-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj, F. Parcy, BZIP transcription factors in *Arabidopsis*, Trends Plant Sci. 7 (2002) 106–111.
[7] R.R. Finkelstein, T.J. Lynch, The *Arabidopsis* abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor, Plant Cell 12 (2000) 599–609.
[8] T.E. Ellenberger, C.J. Brandl, K. Struhl, S.C. Harrison, The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein–DNA complex, Cell 71 (1992) 1223–1237.
[9] M. Neuberg, M. Schuermann, J.B. Hunter, R. Muller, A Fos protein containing the Jun leucine zipper forms a homodimer which binds to the AP1 binding site, Nature 338 (1989) 589–590.
[10] R. Gentz, F.J. Rauscher III, C. Abate, T. Curran, Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains, Science 243 (1989) 1695–1699.
[11] L.J. Ransone, P. Wamsley, K.L. Morley, I.M. Verma, Domain swapping reveals the modular nature of Fos, Jun, and CREB proteins, Mol. Cell. Biol. 10 (1990) 4565–4573.
[12] X. Niu, M.J. Guiltinan, DNA binding specificity of the wheat bZIP protein EmBP-1, Nucleic Acids Res. 22 (1994) 4969–4978.
[13] A.L. Vettore, J.A. Yunes, G. Cord Neto, M.J. da Silva, P. Arruda, A. Leite, The molecular and functional characterization of an *Opaque-2* homologue protein from *Coix* and new classification of plant bZIP proteins, Plant Mol. Biol. 36 (1998) 249–263.
[14] M. Vincentz, P.S. Schlögl, L.G. Corrêa, F. Kuhne, A. Leite, Phylogenetic relationships between *Arabidopsis* and sugarcane bZIP transcriptional regulatory factors, Genet. Mol. Biol. 24 (2001) 55–60.
[15] M.E. Williams, R. Foster, N-H. Chua, Sequences flanking the hexameric G-box core CAGGTG affect the specificity of protein binding, Plant Cell 4 (1992) 485–496.
[16] T. Izawa, R. Foster, N.-H. Chua, Plant bZIP protein DNA binding specificity, J. Mol. Biol. 230 (1993) 1131–1144.
[17] R. Foster, T. Izawa, N.-H. Chua, Plant bZIP proteins gather at ACGT elements, FASEB J. 8 (1994) 192–200.
[18] A.E. Menkens, U. Schindler, A.R. Cashmore, The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins, Trends Biochem. Sci. 20 (1994) 506–510.
[19] A.R. Oliphant, C.J. Brandl, K. Struhl, Defining the sequence specificity of DNA-biding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein, Mol. Cell. Biol. 9 (1989) 2944–2949.
[20] P. Lamb, S.L. McKnight, Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization, Trends Biochem. Sci. 16 (1991) 417–422.
[21] J.N. Glover, S.C. Harrison, Crystal structure of heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA, Nature 373 (1995) 257–261.

[22] A.L. Vettore, F.R. da Silva, E.L. Kemper, G.M. Souza, A.M. da Silva, M.I. Ferro, F. Henrique-Silva, E.A. Giglioti, M.V. Lemos, L.L. Coutinho, M.P. Nobrega, H. Carrer, S.C. Franca, M. Bacci Junior, M.H. Goldman, S.L. Gomes, L.R. Nunes, L.E. Camargo, W.J. Siqueira, M.A. Van Sluys, O.H. Thiemann, E.E. Kuramae, R.V. Santelli, C.L. Marino, M.L. Targon, J.A. Ferro, H.C. Silveira, D.C. Marini, E.G. Lemos, C.B. Monteiro-Vitorello, J.H. Tambor, D.M. Carraro, P.G. Roberto, V.G. Martins, G.H. Goldman, R.C. de Oliveira, D. Truffi, C.A. Colombo, M. Rossi, P.G. de Araujo, S.A. Sculaccio, A. Angella, M.M. Lima, V.E. de Rosa Junior, F. Siviero, V.E. Coscrato, M.A. Machado, L. Grivet, S.M. Di Mauro, F.G. Nobrega, C.F. Menck, M.D. Braga, G.P. Telles, F.A. Cara, G. Pedrosa, J. Meidanis, P. Arruda, Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane, Genome Res. 13 (2003) 2725–2735.

[23] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 15 (1997) 4876–4882.

[24] F.T. 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.

[25] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

[26] M.M. Bradford, A rapid and sensitive method for quantification of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.

[27] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 277 (1970) 680–685.

[28] M. Sprenger-Haussels, B. Weisshaar, Transactivation properties of parsley proline-rich bZIP transcription factors, Plant J. 22 (2000) 1–8.

[29] C. Bracken, P.A. Carr, J. Cavanagh, A. Palmer, Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA, J. Mol. Biol. 285 (1999) 2133–2143.

[30] D. Mohanty, A. Kolinski, J. Skolnick, De novo simulations of the folding thermodynamics of the GCN4 leucine zipper, Biophys. J. 77 (1999) 54–69.

[31] J.R. Stone, J.L. Maki, S.C. Blacklow, T. Collins, The SCAN domain of ZNF174 is a dimer, J. Biol. Chem. 277 (2002) 5448–5452.

[32] C.S. Hardtke, K. Gohda, M.T. Osterlund, T. Oyama, K. Okada, X.W. Deng, HY5 stability in *Arabidopsis* is regulated by phosphorylation in its COP1 binding domain, EMBO J. 19 (2000) 4997–5006.

[33] P. Ciceri, E. Gianazza, B. Lazzari, G. Lippoli, A. Genga, G. Hoschek, R.J. Schmidt, A. Viotti, Phosphorylation of Opaque-2 changes diurnally and impacts its DNA binding activity, Plant Cell 9 (1997) 97–108.

[34] D.A. Jans, The regulation of protein transport to the nucleus by phosphorylation, Biochem. J. 311 (1995) 705–716.

[35] L.J. Klimczak, U. Schindler, A.R. Cashmore, DNA binding activity of Arabidopsis G-box-binding factor GBF1 is stimulated by phosphorylation by casein kinase II from broccoli, Plant Cell 4 (1992) 87–98.

[36] W. Dröge-Laser, A. Kaiser, P. Lindsay, B.A. Halkier, G.J. Loake, P. Doerner, R.A. Dixon, C. Lamb, Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses, EMBO J. 16 (1997) 726–738.

[37] A. Hill, A. Nantelis, C.D. Rock, R. S Quatrano, A conserved domain of the *viviparous*-1 gene product enhances the DNA binding activity of the bZIP protein EmBP-1 and other transcription factors, J. Biol. Chem. 271 (1996) 3366–3374.

[38] M. Seki, J. Ishida, M. Narusaka, M. Fujita, T. Nanjo, T. Umezawa, A. Kamiya, M. Nakajima, A. Enju, T. Sakurai, M. Satou, K. Akiyama, K. Yamaguchi-Shinozaki, P. Carninci, J. Kawai, Y. Hayashizaki, K. Shinozaki, Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray, Funct. Integr. Genomics 2 (2002) 282–291.

[39] T. Meshi, I. Moda, M. Okanami, M. Iwabuchi, Conserved Ser residues in the basic region of the bZIP-type transcription factor HPB-1a (17): importance in DNA binding and possible targets for phosphorylation, Plant Mol. Biol. 36 (1998) 125–136.

[40] M.J. Aukerman, R.J. Schmidt, B. Burr, F.A. Burr, An arginine to lysine substitution in the bZIP domain of an Opaque-2 mutant in maize abolishes specific DNA binding, Genes Dev. 5 (1991) 310–320.

[41] U. Schindler, A.E. Menkens, H. Beckmann, J.R. Ecker, A.R. Cashmore, Heterodimerization between light regulated and ubiquitously expressed *Arabidopsis* GBF bZIP proteins, EMBO J. 11 (1992a) 1261–1273.

[42] D. Igarashi, S. Ishida, J. Fukazawa, Y. Takahashi, 14-3-3 Proteins regulate intracellular localization of the bZIP transcriptional activator RSG, Plant Cell 13 (2001) 2483–2497.

[43] K. Baum, U. Wienand, I. Meier, Reduction of G-box-binding factor DNA binding activity, but not G-box-binding factor abundance, causes the down regulation of RBCS2 expression during early tomato fruit development, FEBS Lett. 54 (1999) 95–99.

[44] B. Weisshaar, G.A. Armstrong, A. Block, O. da Costa e Silva, K. Hahlbrock, Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness, EMBO J. 10 (1991) 1777–1786.

[45] A. Lupas, Coiled coils: new structures and new functions, Trends Biochem. Sci. 21 (1996) 375–382.

Artigo 4

Expression profiles of bZIP in sugarcane

Paulo Sérgio Schlögl [a], Fábio Tebaldi Nogueira [a], Rodrigo Drummond[a], Juliana Felix[a], Vicente E. De Rosa Jr.[a], Adilson Leite [a], Eugênio C. Ulian[b], Jörg Kobarg[c] and Marcelo Menossi [a*].

[a] Centro de Biologia Molecular e Engenharia Genética, Departamento de Genética e Evolução, Instituto de Biologia, Unicamp, CEP 6010, 13083-875, Campinas, SP, Brasil.

[b] Centro de Tecnologia Copersucar, Piracicaba, São Paulo, Brazil.

[c] Centro de Biologia Molecular Estrutural, Laboratório Nacional de Luz Síncrotron, Rua Giuseppe Máximo Scolfaro 10.000, CP 6192, 13084-971, Campinas, SP, Brazil

* Corresponding author:

Centro de Biologia Molecular e Engenharia Genética

Universidade Estadual de Campinas (UNICAMP),

13083-875, Campinas, SP, Brazil

Telephone number: (+55) (19) 3788-1143

Fax number: (+55) (19) 3788-1089

E-mail: menossi@unicamp.br

Abstract

Sugarcane is generally propagated by stalk cuttings containing one or more buds whose germination constitutes a critical period in the life of a sugarcane plant. The transition from the dormant into the active stage constitutes a complex phenomenon characterized by changes in gene expression, accumulation of phytohormones and other important physiological changes. ABA and MeJA are major signaling molecules which influence plant development and stress responses. These plant regulators modulate gene expression at the level of mRNA transcription, with the participation of many families of transcriptional factors that function like switchers to regulate gene expression. bZIPs proteins form a large family of transcriptional factors that are spread in plant kingdom. They are involved in a variety of plant physiological processes, such as development, and responses to abiotic and biotic stress. We characterized 85 sugarcane assembled sequences (SAS) corresponding to the full set of bZIP proteins found in the sugarcane EST project (SUCEST, http://sucest.lad.ic.unicamp.br/public). We applied cDNA arrays and RNA blots to study the expression of sugarcane bZIPs during plantlet development and in response to abscisic acid and methyl-jasmonate. Seven bZIP genes were found to be differentially expressed during development and eight were found to be modulated by ABA and MeJA. Moreover, digital mRNA expression profile analysis using the SUCEST database was performed to gain further information about the differentially expressed bZIP genes. Our findings provide novel insights into understanding the expression of this large protein family in sugarcane.

Keywords: signaling, plant bZIPs, cDNA arrays, development

Introduction

Development and the responses to ambiental stimulus are based on the cellular capacity for differential gene expression and are often modulated by transcription factors acting as switches in regulatory cascades (for review see, Riechmann et al., 2000; Riechmann, 2002). Transcription factors form an intricate network of protein-protein and/or protein-DNA interactions controlling the expression of the genome (Riechmann, 2002).

The *Arabidopsis* genome codes for at least 1500 transcriptional regulators, which account for 5% of the estimated total number of genes (Riechmann et al., 2000). The need of analytical approaches becomes clear when it is considered that less than 10% of these factors have been genetically characterized. bZIPs transcriptional factors are spread in the plant kingdom, but there are only a few characterized plant bZIPs that participate in plant development (Izawa et al., 1994; Walsh et al., 1997; Chuang et al., 1999).

Phytohormones regulate and integrate plant development, overall growth and reproduction. Abscisic acid (ABA) and methyl-jasmonate (MeJA) are some of these phytohormones which have important roles in plant development and plant response to biotic and abiotic stress (Himmelbach et al., 2003; Devoto and Turner, 2003).

ABA modulates the expression of a wide array of genes during seed development and germination (Seki et al., 2002), and in responses to various adverse environmental conditions (Shinozaki et al., 1996). Many ABA-induced genes contain a conserved, ABA-responsive *cis*-acting element named ABRE (ABA-Responsive Element) in their promoter region (Grill and Himmelbach, 1998). The ABRE elements of these genes are targets for many plant bZIPs (Guiltinan et al., 1990; Uno et al., 2000; Kang et al., 2002). Some of these transcriptional activators are themselves induced by ABA and/or stress treatments. ABI5 for example, regulates the expression of some *LEA* genes mainly expressed during seeds development, and is also regulated by ABA (Finkelstein and Lynch, 2000).

Jasmonate signaling molecules affect a variety of plant process including fruit ripening, senescence, production of viable pollen, root growth and defenses against insects, pathogens and environmental stress, such as cold, salinity and drought (Devoto and Turner, 2003; Cheong and Choi, 2003). Although the biological function of several jasmonate-responsive genes is not known yet, some up-regulated genes are involved in cell-wall formation, secondary metabolism,

protection against stress, and even in jasmonate biosynthesis, while down-regualted genes are involved in photosynthesis and light harvesting complex II (Cheong and Choi, 2003). Several *cis*-acting elements responsible for gene activation have been identified in the promoter regions of jasmonate-responsive genes, including G-box sequences that are target of bZIP transcriptional factors (Wasternack and Hause, 2002).

Sugarcane is an important industrial crop of tropical and subtropical regions and is generally propagated by cuttings of the stalk containing one or more buds. The bud is a miniature stem with its growing point and primordia of leaves and roots. The transition from the dormant into the active stage constitutes a complex phenomenon characterized by changes in gene expression, accumulation of phytohormones and other important physiological changes (Van Dillewijn, 1952).

In the sugarcane ESTs project (SUCEST, http://sucest.lad.ic.unicamp.br/public) 90% of the sugarcane expressed genes were tagged (Vettore et al., 2003). In this work we used cDNA-array technology to get insights on the expression of a set of sugarcane bZIPs during development of sugarcane plantlets and in response to abscisic acid (ABA) and methyl-jasmonate (MeJA). We found that seven sugarcane bZIP genes were differentially expressed during plantlet development. We also observed that six other genes were modulated by ABA and two by MeJA.

MATERIALS AND METHODS

Plant Growth and Treatments

Sugarcane development

Sugarcane stalk sets (*Saccharum* sp. cv SP80-3280, Copersucar, Brazil) were cultivated under greenhouse conditions in 200 ml plastic cups containing a commercial planting mix (Plantmax, Eucatex, Brazil) plus ammonium sulfate and commercial 4-20-20 fertilizer. Plant material was harvested 2, 4, 13 and 22 days after planting. Twenty lateral buds were collected after 2 and 3 days and six plantlets (stem and leaves) were collected after 13 and 22 days after planting. The tissues: lateral buds (Lb), leaves (Lv), stem (St), roots (Rt) and flowers (Fl) were obtained as described by Vettore et al. (2003).

ABA and MeJA treatments

Sugarcane plantlets (*Saccharum* sp. cv SP80-3280, Copersucar, Brazil) were grown *in vitro* in MS media as previously described (Nogueira et al., 2003). Plants were kept in a growth chamber at 26°C on a 16h/8h day/night cycle with photon flux density of 70 $\mu E\ m^{-2}\ s^{-1}$. After four weeks 100 µM of ABA were added to the medium and the leaves of the plantlets were harvested at 0, 6 and 12h after treatment. MeJA (100 µM) was added to the medium and the leaves were harvested at 0, 1 and 12h. Six plantlets were used for each time point.

RNA Extraction

Total RNA was isolated using Trizol Reagent (GibcoBRL, USA) and in the case of the lateral buds, a high salt precipitation (0.8 M of sodium citrate and 1.2 M NaCl) step was used according to the manufacturer's instructions, due to the high polysaccharide content.

Digital mRNA Expression Profile Analysis

To analyze the digital expression profile of the bZIP transcripts in sugarcane tissues, SUCEST cDNA libraries (Vettore et al., 2003) were grouped into seven library pools, and named as follows: infected (I): AD1 and HR1libraries; meristematic tissues (M): AM1, AM2,

LB1 and LB2; flower (F): FL1, FL3, FL4, FL6 and FL8; leaves (L): LV1, LR1 and LR2; root (R): RT1, RT2 and RT3; seeds (Sd): SD1 and SD2; stem (St): ST1, ST3, SB1, RZ1, RZ2 and RZ3. The relative abundance of each bZIP was estimated based upon EST counts per corresponding SAS in relation to the total number of ESTs of each SUCEST library pool. Additionally, statistical equations described by Audic and Claverie (1997) were used to differentiate between random EST sampling fluctuations and significant change in EST frequencies in the digital mRNA expression profile analysis. For this analysis, the statistical analysis program for digital transcription profiles was downloaded from http://igs-server.cnrs-mrs.fr and the analysis was performed as described by Koo and Ohlrogge (2002), using a p-value of 0.05.

Construction of sugarcane EST arrays

cDNA clones were spotted on nylon membranes using a hand-held tool with a 96-pin printhead (V&P-Scientific, USA). Eight sets of high-density filters were produced, each one containing 85 bZIP EST targets. Additionally, 96 ESTs representing putative house-keeping genes selected based in their digital expression profile data were used as putative non variant controls (data not published), and other 96 spots representing the empty plasmid vector were used in each filter as a negative control to assess nonspecific hybridization. Each EST was spotted at four positions on the filters to assess the reproducibility of spotting.

Probe preparation and hybridizations

Before cDNA probe hybridization the high-density filters were hybridized with a probe corresponding to the plasmid vector used in the cDNA libraries. This oligonucleotide probe recognize the sequence of the *Amp*^{r} gene of the pSPORT1 vector and was used to estimate the DNA variations in the spots. This probe was synthesized with the primers 5'-GTGGTCCTGCAACTTTATCCGC-3' and 5'-TAGACTGGATGGAGGCGGATAA-3' in the presence of [α-^{33}P] dCTP, according to the protocol described by McPherson (2000; http://www.tree.caltech.edu/protocols/overgo.html). The oligonucleotide probe was stripped with boiling SDS and the filters were reused with cDNA probes.

cDNA probes were produced as described by Schummer et al. (1999) with slight modifications. In brief, 30 μg of total RNA was reverse transcribed with Superscript II (Invitrogen, USA) using an oligo-dT18V (3 μM) primer, with 3,000 Ci $mmol^{-1}$ [α-^{33}P]dCTP and unlabeled dATP, dGTP, and dTTP (1 mM each) for 20 min at 42°C. Unlabeled dCTP was then added to a final concentration of 1 mM, and the reaction continued for another 40 min. The cDNA probes were purified by using ProbeQuant G-50 microcolumns according to the manufacturer's instructions (Amersham Biosciences, USA). Hybridizations were done according to Nogueira et al., 2003. Digitized images of the hybridization signals were quantified using the Image Gauge software (Fujifilm, Japan).

Data analysis

The median value from all spot intensities was determined. The coefficient of variation (CV) of these median values were used to assess fluctuations in the amount of DNA between replicate filters. Only filters with CV values lower than 10% for all ESTs were used for subsequent analysis.

After cDNA probe hybridization the mean signal plus 1.65 sd from the spots containing the empty pSPORT vector was used as a threshold to discard genes with low expression intensities and this criterion was also applied to the housekeeping genes. The pSPORT spots consistently produced a hybridization signal above background, probably because of some degree of similarity between sugarcane cDNA and the vector sequences. In addition, to reduce the variation among replicate filters caused by differences in experimental conditions, the average of all signal intensities obtained with the cDNA probe from each filter was set to 1 (Schummer et al., 1999).

In the ABA and MeJA experiments, the median signal of the housekeeping genes in each membrane was also used as normalization factors. In the experiment with developing buds a subset of housekeeping genes was used because most of them were found to change their expression. The logarithms of the normalized signals (logsignals) were used, since they are supposed to have a normal distribution. To compare different treatments the Student t-test and the ISER test (Intensity-dependent Selection of Expression Ratios, Drummond et al., submitted) were used with a 95% confidence level. The later test was designed to optimize the selection of

genes that showed the most significant variations in expression among two RNA samples under study, with a single hybridization for each sample.

In the experiment with different developmental stages only one cDNA array hybridization was performed and three comparisons of samples were done: 4, 13, and 22 days against the first time point (2 day) , using both ISER and the t-test. The logsignals of the four spots of each gene on each membrane were the input of the t-tests, and the mean normalized signal of each gene on each membrane were the ISER input. All genes that were selected as differentially expressed in at least two of the three treatments by both tests were considered as having true differential expression.

In the experiments using phytohormones, the comparisons were done considering the samples colected at 0h as control against the 1 and 12h for MeJa, and 6 and 12 h for ABA in two replicated hybridizations. In each comparison, t-tests were applied taking all logsignals of each gene from both replicates as input, and ISER tests were applied separately in each replicate hybridization. For each treatment, all genes that were selected as differentially expressed by ISER in both replicate hybridizations or selected by the t-test and by ISER in one of the replicate hybridizations were considered as having true differential expression.

RNA-Blot Analysis

Ten micrograms of total RNA were electrophoresed in a 1 % (w/v) agarose gel containing formaldehyde and transferred to a Hybond-N+ filter (Amersham Pharmacia Biotech., USA) as described by Sambrook et al. (1989). The filters were hybridized with the cDNA inserts of development, and ABA- and MeJA-induced sugarcane ESTs labeled with α-^{32}P dCTP and hybridizations were done at 42°C. The blots were then washed at high stringency and exposed to imaging plates. Digitized images of the RNA-blot hybridization signals were quantified using the Image Gauge software (Fujifilm, Japan).

Results

Data mining of bZIPs in the SUCEST database

To initiate an expression profile data of sugarcane bZIPs transcriptional factors we undertook an extensive data mining in the SUCEST database (Vicentz et al., 2001). An ordered

non redundant and complete set of *Arabidopsis* bZIP transcriptional factors was previously employed to screen and classify phylogenetically the 121 sugarcane assembled sequences (SAS) coding for bZIPs (Vincentz et al., 2001). After a new clusterization in the SUCEST database (Telles and da Silva, 2001) the number of bZIP SAS was corrected to 85 (Table 1). These analyses could help to assign putative functions to sugarcane bZIPs since homologous proteins might show similar functional activities. Altrough many bZIP genes have been discovered, only a minority has been genetically or biochemically characterized (Riechmann, 2000). More than 50% of the sugarcane bZIPs shown in Table 1 had a similar protein not characterized in the literature. This highlights the need for a better characterization to get a further understanding of these proteins in several aspects of sugarcane development.

Digital mRNA expression profile analysis

EST frequencies in a standard cDNA library can be used to extract information on gene-expression levels (Koo and Ohlrogge, 2002). This assumption is based in the fact that the number of EST clones from any gene is proportional to the abundance of its mRNA in the cells used to construct the cDNA library (Audic and Claverie, 1997).

To obtain information on which tissues sugarcane bZIP genes are preferentially expressed, we performed a digital mRNA expression profile analysis to discriminate significant changes in EST frequencies (Figure 1). The sub-family VI had the highest representation in all SUCEST libraries especially in stem, flower, roots, and seed. All other sugarcane bZIP sub-families showed low levels of expression in all tissues, with the exception of sub-families II, IX and X, which were highly expressed in infected, stem and seed libraries, respectively. The sub-families I and V were poorly expressed in all tissues and absent in seed and root libraries.

Sugarcane bZIPs modulated during development

The changes in bZIP expression during bud germination were evaluated with nylon cDNA arrays containing 85 sugarcane bZIPS. Seven sugarcane bZIP ESTs showed altered expression during sugarcane plantlet development, five being up-regulated and two down-regulated (Table 2). Figure 2 shows the expression profiling of six ESTs in both the cDNA-array and RNA-blot analysis performed with RNA from biological replicates. Although the absolute

fold-induction values were not identical between samples the expression profiles were very similar. Figure 2 C shows the result from RNA blot analysis in leaves, flowers, stems, roots and lateral buds. SCRUHR1075A10 and SCEQRT3020F04 were up-regulated during development and the highest levels of expression were observed 13 days after planting. SCRUHR1075A10 was highly expressed in leaves and stem, but weakly expressed in lateral bud and flowers, and is not expressed in roots (Figure 2 C), while SCEQRT3020F04 was expressed only in stems. SCBFAD1046H12 was down-regulated during bud germination, showed high levels of expression three and four days after planting (Figure 2 B) and was expressed only in the lateral bud (Figure 2 C). SCSGAM1095E12 was shown to be up-regulated and was described in other study (Schlögl et al., Plant Science, in press). These results indicate that SCRUHR1075A10 and SCEQRT3020F04 are probably functionally active at late stages of development and that SCBFAD1046H12 and SCSGAM1095E12 participates in early developmental stages.

SCCCRT1001E06, SCEZLR1031F04, and SCJLRT1019E07 were up-regulated and had similarly high levels of expression in all stages of plantlets development, although the level of the transcripts from SCCCRT1001E06 decreased at 22 days (Figure 2 A and B). SCCCRT1001E06 and SCEZLR1031F04 were expressed in all tissue with similar levels, except in roots, where lower levels were observed (Figure 2 C). SCJLRT1019E07 was highly expressed in flowers, stems and weakly expressed in leaves (Figure 2 C).

Identification of ABA and MeJA-responsive Sugarcane ESTs with DNA-array

The cDNA arrays used previously were hybridized with cDNA from plantlets grown *in vitro* and treated with ABA and MeJA. The expression profiles of eight SASs were altered in response to ABA and MeJA treatments (Table 3, Figure 3). None of them corresponded to any of the seven SASs found to be modulated during plantlet development. The average expression ratios of ABA and MeJA-responsive sugarcane bZIPs and their homologous best hit from GenBank database are shown in Table 3. Most of the ABA- and MeJA-responsive sugarcane bZIP showed similarity with non characterized bZIPs from rice (Table 3). The expression profiling of six ABA-regulated and two MeJA-regulated SASs are shown in Figure 3. RNA-blot analyses with selected genes showed good agreement with the cDNA array data for the sugarcane bZIPs selected to perform this analysis, confirming that these bZIPs are responsive to ABA and MeJA (Figure 4).

Discussion

Sugarcane bZIPs involved in sugarcane development

An ordered set of *Arabidopsis* bZIP factors was used to identify and classify the sugarcane SASs coding for bZIP transcriptional factors from the SUCEST database. This strategy allowed the identification of 85 sugarcane SASs encoding bZIP factors. All sugarcane bZIPs SASs were classified in one of the *Arabidopsis* sub-families described in Vicentz et al., (2001).

A digital mRNA expression profile analysis showed that the bZIP sub-families are differentially expressed in the sugarcane tissues. The high levels of bZIPs from sub-family VI in all tissues evaluated suggest that these transcriptional factors may have a more general function that takes place in most sugarcane tissues. Most sub-families had higher levels in a few tissues, while others were not even detected in some sugarcane organs (Figure 1). These findings reinforce the wide array of functions that bZIPs play in plants.

We used nylon cDNA arrays to assess the expression profile of 85 sugarcane bZIP during plantlet development and in response to ABA and MeJA. Five genes were up-regulated and two were down-regulated during the development of sugarcane plantlets (Table 2). SCRUHR1075A10 was strongly up-regulated during plantlet development and was expressed in all the tissues sampled, except in roots. Based in these results we can specultate that SCRUHR1075A10 is involved not only in development, but also in other functions that take place in adult organs, such as leaves. This bZIP had no close homolog, showing a weak similarity of 45% to a tobacco bZIP, TGA2.1, also expressed in leaves (Niggeweg et al., 2000).

SCEQRT3020F04 had 85% similarity to liguleless2, a bZIP that participates in the leaf development in maize (Harper and Freeling, 1996). The expression of this sugarcane gene increased as the young leaves were formed, but was not detected in adult leaves, suggesting a similar role to the maize *liguleless2*. This gene was also expressed in the stem, indicating that this bZIP might also be involved in other functions yet to be elucidated.

SCBFAD1046H12 was active in the early stages of development and was detected only in lateral buds (Figure 2). This bZIP had 93% similarity with OBF3.2 from maize (Foley et al., 1993) and 73% to the HBP-1b from wheat (Mikami et al., 1994). HBP-1b is a promoter-binding protein, which specifically binds to the ACGT core sequence from histone genes. These data indicate that SCBFAD1046H12 might act controling the expression of sugarcane histone genes.

SCCCRT1001E06 was up-regulated and expressed during all developmental stages and presented the highest level at day 13, then dropped drastically at 22. It was ubiquitously expressed in all sugarcane tissues evaluated (Figure 2). Furthermore, this SASs had 86% of identity to HBP-1a, another wheat bZIP involved in the transcriptional regulation of wheat histone genes in a cell cycle-dependent way (Tabata et al., 1991). These results suggest that SCCCRT1001E06 might have a similar function during sugarcane plantlets development. However, transcripts were also detected in adult leaves, indicating that this SASs participates in other process in sugarcane.

SCEZLR1031F04 showed higher levels of expression at 13 and 22 days after planting and was expressed in all tissues evaluated. This sugarcane bZIP had 87% identity to the maize GBF1, which is induced by hypoxia and seems to be involved in the activation of *Adh1* (de Vetten & Ferl, 1995). These findings indicate that SCEZLR1031F04 is involved in the late stages of bud germination and that might also participate in stress responses.

SCJLRT1019E07 was expressed in in all stages of the sugarcane development, with high levels 13 days after planting. This bZIP seems to be acting also in the stem and in the flower, where the highest expression was observed. Due to the low similarity (48%) with AtbZIP60, an uncharacterized bZIP from *Arabidopsis*, it is hard to specify another role for this bZIP. It is worth to mention that the number of bZIP modulated during sugarcane development may be underestimated because most transcriptional factors presented very low expression levelsand could be lost during data analysis, or could be expressed at different conditions than those analyzed here.

Hormonal control of bZIP expression

Many aspects of plant development, abiotic and biotic stress responses are mediated by phytohormones such as ABA and MeJA (Cheong and Choi, 2003). Transcription factors play a key role in mediating the action of phytohormones. Eight bZIP genes were modulated by ABA or MeJA treatments. ABA induced the expression of two genes and repressed other four, and Meja induced two genes (Table 3). Blast analysis showed that the best hits of these sugarcane SASs were uncharacterized bZIPs from other plant species. We used other hits with low E-value to add some information from characterized plant bZIPs (Table 3).

SCRURT2012D03 and SCCCRT1003G04 were up-regulated by ABA and showed 76% and 77% sequence similarity with STGA1, respectively. STGA1 is expressed in mature hypocotyls from soybean and binds to a motif found in an auxin-responsive promoter (Cheong et al., 1998). Since both bZIPs were not detected in control plantlets and considering that ABA and auxin signaling seems to be interconected in drought response in *Arabidopsis thaliana* (Bianchi et al., 2002), it is tempting to speculate that the two sugarcane bZIPs might be involved in the cross talk between these hormones uder stress conditions. SCJFLR1035B10 was down-regulated by ABA and had 79% similarity to BZI-2 from tobacco (Strathmann et al., 2001). Since this tobacco protein dimerizes with another bZIP, BZI-1, whose lower expression in transgenic plants affected the development of the stamen and the petals, the sugarcane bZIP might be involved in the development of sugarcane flowers. ABA also down-regulated the expression of SCCCCL4005C09. This bZIP had 71% similarity with RISBZ5, a rice bZIP that was weakly expressed during seed maturation (Onodera et al., 2001), indicating that this protein could modulate gene expression in sugarcane seeds. SCMCST1051A01, SCCCAM1003G07 and SCJFLR1035B10 were also down-regulated by ABA, but both had no strong similarity with any characterized plant bZIP and no function could be assigned to them.

MeJA treatment resulted in the up-regulation of two sugarcane bZIPs. SCJFLR1035G11 showed 95% similarity to OCSBF1 a maize bZIP whose transcripts are abundant in the basal region from leaves and in young roots, both rich in dividing and differentiating cells (Singh et al., 1990). This gene belongs to the sub-family VI, being detected in all sugarcane tissues sampled in the cDNA libraries (Figure 1), suggesting that the MeJA-modulation might be involved in the regulation of several process in a wide array of cell types. SCEPAM1053B09 was also induced by MeJA treatment but showed no strong similarity to other bZIPs.

These results indicate that gene regulation by ABA and MeJA in sugarcane probably involves the action of bZIPs. Several studies in other species show that these phytohormones control many aspects not only in plant development and also stress responses such as drought and pathogen atack (Seki et al., 2002). In fact, we have previously found a sugarcane bZIP that is induced by cold stress (Nogueira et al., 2003). Experiments to evaluate the expression of these bZIPs in response to biotic and abiotic stress are underway and will be helpful to further understand their role in sugarcane.

In summary, we used mRNA profiles generated by DNA-arrays and GeneBank searches to deduce functions of sugarcanes ESTs encoding putative bZIP transcription factors. We have identified a number of sugarcane bZIPs that are expressed in particular organs, at particular developmental stages and modulated by ABA or MeJA. Our findings highlight the relevance of bZIPs in the regulation of sugarcane development and their putative role in the activation of genes modulated by ABA and MeJA.

Acknowledgements

The authors thank Eduardo Kiyota and Edna Rosa dos Santos by technical assistance. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through a doctoral fellowship to P.S.S. (process 00/02480-0).

References

Aeschbacher RA, Schrott M, Potrykus I, Saul MW. (1991). Isolation and molecular characterization of PosF21, an Arabidopsis thaliana gene which shows characteristics of a b-Zip class transcription factor. The Plant Journal 1: 303 - .

Aguan K, Sugawara K, Suzuki N, Kusano T. (1991). Low-temperature-dependent expression of a rice gene encoding a protein with a leucine-zipper motif. Mol Gen Genet. **240**:1-8.

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

Audic, S., Claverie, J.M. (1997). The significance of digital gene expression profiles. Genome Res. **7**: 986-995.

Bianchi, M.W., Damerval C., Vartanian, N. (2002). Identification of proteins regulated by cross-talk between drought and hormone pathways in Arabidopsis wild-type and auxin-insensitive mutants, axr1 and axr2. Funct. Plant Biol.. **29**:55–61

Bonetta, D., and McCourt, P. (1998). Genetic analysis of ABA signal transduction pathways. Trends Plant Sci **3**: 231-235.

Carlini LE, Ketudat M, Parsons RL, Prabhakar S, Schmidt RJ, Guiltinan MJ. (1999). The maize EmBP-1 orthologue differentially regulates opaque2-dependent gene expression in yeast and cultured maize endosperm cells. Plant Mol Biol. **41**:339-49.

Casaretto J, Ho TH. (2003).The transcription factors HvABI5 and HvVP1 are required for the abscisic acid induction of gene expression in barley aleurone cells. Plant Cell. **15**:271-84.

Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, et al. (2002). Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell. **14**:559-74.

Cheong JJ, and Choi YD. (2003). Methyl jasmonate as a vital substance in plants. Trends Genet. **19**:409-13.

Cheong YH, Yoo CM, Park JM, Ryu GR, Goekjian VH, Nagao RT, Key JL, Cho MJ, Hong JC. (1998). STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP domain which heterodimerizes with GBF proteins. Plant J. **15**: 199-209.

Choi H, Hong J, Ha J, Kang J and Kim SY. (2000). ABFs, a family of ABA-responsive elements binding factors. J. of Biological Chemistry **21**: 1723-1730.

Chuang CF, Running MP, Williams RW, Meyerowitz EM. (1999). The PERIANTHIA gene encodes a bZIP protein involved in the determination of floral organ number in Arabidopsis thaliana. Genes Dev. **13**: 334-44.

Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbuhl P, Ellero C, Goff SA, Glazebrook J. (2003). A network of rice genes associated with stress response and seed development. Proc Natl Acad Sci U S A. 100:4945-50.

Creelman RA, Mullet JE. (1997). Oligosaccharins, brassinolides, and jasmonates: nontraditional regulators of plant growth, development, and gene expression. Plant Cell. **9**:1211-23.

Dai,S., Zhang,Z., Chen,S. and Beachy,R.N. (2004). RF2b, a rice bZIP transcription activator, interacts with RF2a and is involved in symptom development of rice tungro disease. Proc. Natl. Acad. Sci. U.S.A. **101**: 687-692

Desprez T, Amselem J, Caboche M, Hofte H. (1998). Differential gene expression in Arabidopsis monitored using cDNA arrays. Plant J. 14:643-52.

Devoto A, Turner JG. (2003). Regulation of jasmonate-mediated plant responses in arabidopsis. Ann Bot (Lond).**92**:329-37.

Droge-Laser W, Kaiser A, Lindsay WP, Halkier BA, Loake GJ, Doerner P, Dixon RA, Lamb C. (1997). Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. EMBO J. **16**:726-38.

Drummond et al. submitted.

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosh KA, Gantt JF and Vance CP. (2002) Genome-wide identification of Nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiology **130:** 519-537.
Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, et al. (2000). The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell. **12**:599-609.
Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, et al. (2002) Sequence and analysis of rice chromosome 4. Nature **420**: 316-20.
Foley RC, Grossman C, Ellis JG, Llewellyn DJ, Dennis ES, Peacock WJ, Singh KB. (1993). Isolation of a maize bZIP protein subfamily: candidates for the ocs-element transcription factor. Plant J. **3**:669-79.
Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, Takahashi Y. (2000). Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. Plant Cell 12: 901-915.
Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J. (2000). Microarray analysis of developing Arabidopsis seeds. Plant Physiol. **124**:1570-81.
Grill E, and Himmelbach A. (1998). ABA signal transduction. Curr Opin Plant Biol. **1**:412-8.
Guiltinan, M. J.; Marcotte, W. R. and Quatrano, R. S. (1990). A plant leucine zipper protein that recognizes an abscisic acid response element. Science **250**, 267-271.
Harper L, Freeling M. (1996). Interactions of liguleless1 and liguleless2 function during ligule induction in maize. Genetics **144**: 1871-82.
Heinekamp T, Kuhlmann M, Lenk A, Strathmann A, Droge-Laser W. (2002). The tobacco bZIP transcription factor BZI-1 binds to G-box elements in the promoters of phenylpropanoid pathway genes in vitro, but it is not involved in their regulation in vivo. Mol Genet Genomics **267**:16-26.
Hemberger M, Cross JC, Ropers HH, Lehrach H, Fundele R, Himmelbauer H. (2001). UniGene cDNA array-based monitoring of transcriptome changes during mouse placental development. Proc Natl Acad Sci U S A. **98**:13126-31.
Himmelbach A, Yang Y and Grill, E. (2003). Relay and control of Abscisic acid signaling. Cur. Opin. Plant Biol. **6**: 470-79.
Hoth S, Morgante M, Sanchez JP, Hanafey MK, Tingey SV, Chua NH. (2002). Genome-wide gene expression profiling in Arabidopsis thaliana reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant. J Cell Sci. **115**:4891-900.
Igarashi D, Ishida S, Fukazawa J, Takahashi Y. (2001). 14-3-3 proteins regulate intracellular localization of the bZIP transcriptional activator RSG. Plant Cell. **13**:2483-97.
Izawa, T., Foster, R., Nakajima, M., Shimamoto, K. and Chua, N. (1994). The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. The Plant Cell **6**, 1277-1287.
Iwasaki T, Yamagushi-Shinozaki K and Shinozaki K. (1995). Identification of a cis-regulatory region of a gene in *Arabidopsis thaliana* whose induction by dehydration is mediated by Abscisic acid and requires protein synthesis. Mol. Gen. Genet. **247**: 391-98.
Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance.Science. **280**:104-6.
Jakoby M, Weisshaar B, Dröger-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T And Parcy F. (2002) bZIP transcription factors in *Arabidopsis*. Trends in Plant Science **7**, 106-111.
Johri M. M. and Mitra, D. (2001). Action of plant hormones. Current Science **80**: 199-205.
Kang J-Y, Choi H-I, Im M-Y and Kim S Y. (2002). Arabidopsis basic leucine zipper proteins that mediate stress-responsive Abscisic acid signaling. Plant Cell **14**:343-357.
Kaneko T, Kotani H, Nakamura Y, Sato S, Asamizu E, Miyajima N, Tabata S. (1998). Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned P1 and TAC clones. DNA Res. **5**:131-45.
Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor.Nat Biotechnol. **17**:287-91.

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, and Harper JF. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol **130**:2129-41.
Kim,S.Y. and Thomas,T.L. (1998). A family of novel basic leucine zipper proteins binds to seed-specification elements in the carrot Dc3 gene promoter. J. Plant Physiol. **152**: 607-613.
Koo, AJK and Ohlrogge, JB. (2002). The predicted candidates of Arabidopsis plastid inner envelope membrane proteins and their expression profiles. Plant Physiol. **130**: 823-836.
Khulmann M, Horvay K, Strathmannn A, Heinekamp T, Fisher U, Böttner S and Droge-Laser W. (2003). The α-helical D1 domain of the tobacco bZIP transcription factor BZI-1 interacts with the ankyrin-repeat protein ANK1 and is important for BZI-1 function, both in Auxin signaling and pathogen response. J Biochem. Chem. **10**: 8786-8794.
Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H. (2001). The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4. DNA Res. **8**: 285-90.
Kusano T, Berberich T, Harada M, Suzuki N, Sugawara K. (1995). A maize DNA-binding factor with a bZIP motif is induced by low temperature. Mol Gen Genet. 248:507-17.
Kusano T, Sugawara K, Harada M, Berberich T. (1998). Molecular cloning and partial characterization of a tobacco cDNA encoding a small bZIP protein. Biochim Biophys Acta. **1395**:171-5.
Lara P, Onate-Sanchez L, Abraham Z, Ferrandiz C, Diaz I, Carbonero P, Vicente-Carbajosa J. (2003). Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. J Biol Chem. **278**:21003-11.
Lee S, Reth A, Meletzus D, Sevilla M, and Kennedy C. (2000). Characterization of a major SAS of nif, fix and associated genes in a sugarcane endophyte, *Gluconacetobacter diazotrophicus*. Journal of Bacteriology **182**: 7088-91.
Lin F, Xu SL, Ni WM, Chu ZQ, Xu ZH and Xue HW. (2003). Identification of ABA-responsive genes in rice shoots via cDNA macroarrays. Cell Research **13**: 59-68.
Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, et al. (1999). Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature. **402**: 761-8.
Lockhart DJ, and Winzeler EA. (2000). Genomics, gene expression and DNA arrays. Nature **405**:827-36.
Lu G, Paul AL, McCarty DR, Ferl RJ. (1996). Transcription factor veracity: is GBF3 responsible for ABA-regulated expression of Arabidopsis Adh?. Plant Cell. **8**:847-57.
Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA. (2000). The transcriptome of Arabidopsis thaliana during systemic acquired resistance. Nat. Genet. **26**:403-10.
Menges M, Hennig L, Gruissem W, Murray JA. (2002). Cell cycle-regulated gene expression in Arabidopsis. J Biol Chem. **277**:41987-2002.
Miao ZH, Liu X, Lam E. (1994). TGA3 is a distinct member of the TGA family of bZIP transcription factors in Arabidopsis thaliana. Plant Mol Biol. **25**: 1-11.
Mikami,K., Sakamoto,A. and Iwabuchi,M. (1994). The HBP-1 family of wheat basic/leucine zipper proteins interacts with overlapping cis-acting hexamer motifs of plant histone genes. J. Biol. Chem. **269**: 9974-9985
Minami M, Huh GH, Yang P, Iwabuchi M. (1993). Coordinate gene expression of five subclass histones and the putative transcription factors, HBP-1a and HBP-1b, of histone genes in wheat. **23**: 429-34.
Moseyko N, Zhu T, Chang HS, Wang X, Feldman LJ. (2002). Transcription profiling of the early gravitropic response in Arabidopsis using high-density oligonucleotide probe microarrays Plant Physiol. **130**:720-8.
Mukai Y, Nagasaki H, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Namiki N, Negishi M, Ohta I, Ono N, Saji S, Sakai K, Shibata M, Shimokawa T,
Nakagawa H, Ohmiya K, Hattori T. (1996). A rice bZIP protein, designated OSBZ8, is rapidly induced by abscisic acid. Plant J. **9**:217-27.
Nantel A, Quatrano RS. (1996). Characterization of three rice basic/leucine zipper factors, including two inhibitors of EmBP-1 DNA binding activity. J Biol Chem. **271**:31296-305.

Niggeweg R, Thurow C, Weigel R, Pfitzner U, Gatz C. (2000). Tobacco TGA factors differ with respect to interaction with NPR1, activation potential and DNA-binding properties. Plant Mol Biol. **42**: 775-88.
Nogueira FT, De Rosa VE Jr, Menossi M, Ulian EC, Arruda P. (2003). RNA expression profiles and data mining of sugarcane response to low temperature. Plant Physiol. **132**: 1811-24.
Oeda K, Salinas J, Chua NH. (1991). A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J. **10**:1793-802.
Onodera,Y., Suzuki,A., Wu,C., Washida,H. and Takaiwa,F. (2001). A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J. Biol. Chem.
Phillips J, Artsaenko O, Fiedler U, Horstmann C, Mock HP, Munits K and Conrad U. (1997). Seed-specific immuno-modulation of Abscisic acid activity induces a developmental switch. EMBO J. **16**: 4489-96.
Pysh LD, Aukerman MJ, Schmidt RJ. (1993). OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. Plant Cell. **5**:227-36.
Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**:2105-10.
Riechmann JL, and Ratcliffe OJ. (2000). A genomic perspective on plant transcription factors. Curr Opin Plant Biol. **3**:423-34.
Rook F., Gerrits N., Kortstee A., van Kampen M., Borrias M., Weisbeek P and Smeekens S. (1998).Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene. The Plant Journal 15, 253-263
Ruan Y, Gilmore J, Conner T. (1998). Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. **15**:821-33.
Sambrook, J., Fritsch, E. F., Maniatis, T. (1989). Molecular cloning: a laboratory manual 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sasaki Y, Asamizu E, Shibata D, Nakamura Y, Kaneko T, Awai K, Amagai M, Kuwata C, Tsugane T, Masuda T, Shimada H, Takamiya K, Ohta H, Tabata S. (2001). Monitoring of methyl jasmonate-responsive genes in Arabidopsis by cDNA macroarray: self-activation of jasmonic acid biosynthesis and crosstalk with other phytohormone signaling pathways. DNA Res. **8**:153-61.
Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, et al. (2002) The genome sequence and structure of rice chromosome 1. Nature **420**: 312-6.
Schlögl PS, Kobarg J, Moreau VH, Leite A, Sabino AA, Eberlin MN and Menossi M. (2004). Expression, purification and characterization of a novel bZIP protein from sugarcane. Plant Science **167**: 583-595.
Schummer M, Ng W-L, Bumgarner RE, Nelson PS, Schummer B, Bednarski DW, Hassell L, Baldwin RI, Karlan BY, Hood L (1999). Comparative hybridization of an array of 21500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene **238**: 375-385.
Schiermeyer A, Thurow C, Gatz C. (2003). Tobacco bZIP factor TGA10 is a novel member of the TGA family of transcription factors. Plant Mol Biol. **51**: 817-29.
Strathmann A, Kuhlmann M, Heinekamp T, Droge-Laser W. (2001). BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development. Plant J. **28**:397-408.
Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM. (2000). Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc Natl Acad Sci U S A. **97**:11655-60.
Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM. (2000). Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc Natl Acad Sci U S A.**97**:11655-60.

Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K. (2002). Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genomics. **2**:282-91.

Shinozaki K, Yamaguchi-Shinozaki K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways.

Shinozaki K, Yamaguchi-Shinozaki K. (1996). Molecular responses to drought and cold stress. Curr Opin Biotechnol. **7**:161-7.

Shomura A, Song J, Takazaki Y, Terasawa K, Tsuji K, Waki K, Yamagata H, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong H, Iwama H, Endo T, Ito H, Hahn JH, Kim HI, Eun MY, Yano M, Jiang J, Gojobori T. (2002) The genome sequence and structure of rice chromosome 1. Nature 420: 312-6.

Singh, K.; Dennis, E. S.; Ellis, J. G.; Llewellyn, D. J.; Tokuhisa, J. G.; Wahleithner, J. A. and Peacock J. (1990). OCSBF-1, a Maize Ocs Enhancer Binding Factor: Isolation and Expression during development. The Plant Cell **2**, 891-903.

Siberyl Y, Benhamron S, Memelink J, Giglioli-Guivarc'h N, Thiersault M, Boisson B, Doireau P and Ganet P. (2001). Catharantus roseus G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. Plant Mol Biol. **45**: 477-88.

Tabata T, Nakayama T, Mikami K, Iwabuchi M. (1991). HBP-1a and HBP-1b: leucine zipper-type transcription factors of wheat. EMBO J. 10: 1459-67.

Telles GP, Braga MDV, Dias Z, Lin T-L, Quitzau JAA, da Silva FR and Meidanis J. (2001). Bioinformatics of the sugarcane EST project. Genetics and Molecular Biology **24**:9-15.

Thomashow MF. (1998). Role of cold-responsive genes in plant freezing tolerance. Plant Physiol. 118:1-8.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. (2000). Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci U S A. **97**:11632-7.

Van Dillewijn C., (1952). Botany of sugarcane. Walthan Mass USA.

de Vetten NC, Ferl RJ. (1995). Characterization of a maize G-box binding factor that is induced by hypoxia. Plant J. **7**:589-601

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, et al. (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res. **13**:2725-35.

Vettore, A. L., Yunes, J. A., Cord Neto, G., da Silva M.J., Arruda, P. e Leite, A. (1998). The molecular and functional characterization of an *Opaque-2* homologe protein from *Coix* and new classification of plant bZIP proteins. Plant Mol. Biol. **36**, 249-263.

Vincentz, M., Schlögl, P. S., Corrêa, L. G., Kuhne, F. e Leite, A. (2001). Phylogenetic relationships between *Arabidopsis* and Sugarcane bZIP transcriptional regulatory factors. Genetics and Molecular Biology **24**:55-60.

Yin Y, Zhu Q, Dai S, Lamb C, Beachy RN. (1997). RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development. EMBO J. **16**: 5247-59.

Zhang Y, Tessaro MJ, Lassner M, Li X. (2003). Knockout analysis of Arabidopsis transcription factors TGA2, TGA5, and TGA6 reveals their redundant and essential roles in systemic acquired resistance. Plant Cell. **15**: 2647-53.

Zhang LS, Yu Z, Fan D, Liu X, Lu T, Li C, Wu Y, Sun T, et al. (2002) Sequence and analysis of rice chromosome 4. Nature **420**: 316-20.

Zhu T, and Wang X. (2000). Large-scale profiling of the Arabidopsis transcriptome. Plant Physiol. **124**:1472-6.

Walsh J, Waters CA, Freeling M. (1998). The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf blade-sheath boundary. Genes Dev. **12**:208-18.

Wasternack C, and Hause B. (2002). Jasmonates and octadecanoids: signals in plant stress responses and development. Prog Nucleic Acid Res Mol Biol. **72**:165-221.
Weber H, Vick BA, and Farmer EE. (1997). Dinor-oxo-phytodienoic acid: a new hexadecanoid signal in the jasmonate family. Proc Natl Acad Sci U S A. **94**:10473-8.
Williams ME, Foster R, and Chua N-H. (1992). Sequences flanking the hexameric G-box core CAGGTG affect the specificity of protein binding. Plant Cell **4**, 485-496.

Figure legends: Article 4

Figure 1 – Digital mRNA Expression Profile Analysis. The expression profile of the bZIP transcripts among sugarcane tissues, the relative abundance was estimated based upon EST counts per corresponding sugarcane assembled sequences in relation to the total number of ESTs of each SUCEST library pool.

Figure 2 – Expression profile of developmentally regulated sugarcane bZIP assembled-sequences. **A**: The graphs show the induction kinetics observed in Northern blots (black squares) and macroarrays (white squares). **B**: Northern blots, each lane was loaded with 10 μg of total RNA isolated from plantlets grown at 72 h, 4 days, and 13 and 22 days after planting. **C**: Northern blots, each lane was loaded with 10 μg of total RNA isolated from different plant tissues. Lb: lateral bud; Lv: leafs; St: stem; Rt: roots and Fl: flowers.

Figure 3 - Expression profile of ABA- and Meja-regulated sugarcane genes. The expression was calculated from normalized relative intensities of each time point of each EST from two independent experiments (white and black squares). **A**: ABA-regulated bZIP SASs and **B**: Meja-regulated bZIP SASs.

Figure 4 – Expression profile of ABA and MeJA modulated sugarcane bZIP assembled-sequences. **A**: The graphs show the induction kinetics observed in the RNA blots (black squares) and cDNA arrays (white triangle). **B**: Northern blots, each lane was loaded with 10 μg of total RNA isolated from sugarcane plants: C (control), 6 h, and 12 after ABA treatment, and 1 h and 12 h after MeJA treatments.

Table 1. Sugarcane bZIPs identified by data mining in the SUCEST database.

Clone ID	Ac. Number	E-value	Description	Identity/ Similarity (%)	References
Family I					
SCRLLR1131B01	AAL79736	4e-46	OsbZIP	80/86	-
SCEPLR1051F12	AAD24610	3e-15	AtbZIP23	48/61	-
SCEPAM1053B09	AAL79736	e-114	OsbZIP	80/86	-
SCEZAM2250C05	AAL79736	2e-65	OsbZIP	74/81	-
Family II					
SCCCLR1076F10	BAB72062	e-146	OsbZIP	93/97	-
SCEQAM2038B11	BAB72061	e-111	OsbZIP	93/97	-
SCSGHR1067E05	BAB72061	1e-07	OsbZIP	82/90	-
SCEPRT2044H06	BAD03235	1e-38	OsSTGA1	53/61	-
SCCCRT1003G04	CAD41728	1e-38	OsbZIP	69/77	1
SCRURT2012D03	CAD41728	2e-52	OsbZIP	67/76	1
SCCCCL3001H09	T01415	5e-44	liguleless2	70/83	2
SCEQRT3020F04	T01415	1e-86	liguleless2	72/85	2
SCQSRZ3037F10	T01415	6e-28	liguleless2	98/98	2
SCEZLB1013F09	CAA48904	e-113	OBF3.2	97/98	3
SCBFAD1046H12	CAA48904	2e-75	OBF3.2	89/93	3
SCRUHR1075A10	AAB68661	7e-07	TGA.2	26/45	4
SCCCSD1096H12	BAB11142	1e-24	Atbzip65	36/48	5
Family III					
SCCCLR1070F08	AAL79735	2e-18	Osbzip	43/54	-
SCACFL5025H02	BAC79602	2e-13	OsDPBF3	54/62	-
SCRLAD1100D06	BAC79602	1e-26	OsDPBF3	70/75	-
SCEPFL4173E10	AAO06115	3e-60	HvABI5	71/80	6
SCJFRZ1006A05	AAO06115	3e-33	HvABI5	77/76	6
SCVPLB1018G11	AAO06115	5e-30	HvABI5	45/51	6
SCCCLR1C03C05	AAO06115	6e-44	HvABI5	54/61	6
SCPILB2024H09	BAB91752	1e-39	OsABI5	57/66	7
SCUTAD1032A02	BAB90392	5e-33	Osbzip	62/73	7
SCCCRT1002F09	BAB89789	1e-59	Osbzip	62/69	7
SCJFFL3C08F11	AAO72626	1e-26	OSE2-like	66/77	8
SCJLFL4183H11	T12585	1e-31	Dc3pbf3	67/80	9

Family V					
SCSGAM1095E12	AAK25822	1e-24	PvbZIP	41/63	10
SCCCLR1067B06	BAB39174	2e-05	AtbZIP7	67/77	11
Family VI					
SCJFLR1035B10	AAD21199	2e-18	Ccbzip	55/79	-
SCEPCL6020G07	T03642	1e-24	Ocsbf1	81/84	-
SCRURT3064D09	T03642	7e-22	Ocsbf1	90/96	-
SCJFLR1035G11	T03642	1e-56	Ocsbf1	93/95	12
SCJFRT1011E05	CAA18838	5e-18	ATB2	44/68	13
SCRLLR1059D11	BAA05617	2e-46	mLIP15	83/88	14
SCBGLR1027H03	BAA05617	6e-35	mLIP15	77/85	14
Family VIII					
SCBFRZ2019D02	BAA36492	1e-09	OsbZIP	38/54	-
SCSGFL4196C02	BAB39174	4e-07	RISBZ4	58/67	11
SCCCCL4004B05	BAB39174	4e-45	RISBZ4		11
SCCCRZ2003E12	BAB39174	3e-85	RISBZ4	63/68	11
SCCCCL4005C09	BAB39175	5e-68	RISBZ5	63/71	11
SCQSRT1036D12	JQ2147	6e-93	OHP1	93/96	15
SCACLR2007C09	JQ2148	3e-11	OHP2	68/73	15
SCMCST1051A01	CAA71687	0.024	G/HBF-1	34/48	16
Family IX					
SCRLST3163A02	T08592	9e-25	STF2	54/67	-
SCCCLR1072B02	T03241	e-125	OsGBF1a	84/90	17
SCCCRZ1C02F09	T03241	9e-22	OsGBF1a	95/98	17
SCCCRT1001E06	BAA02304	3e-20	TaHBP1a	86/90	18
SCCCRZ2003B03	BAA02304	e-117	TaHBP-1a	85/90	18
SCCCST3144C04	BAA02304	9e-30	TaHBP-1a	87/91	18
SCBFRZ2048C10	BAA02304	4e-84	TaHBP-1a(17)	71/81	18
SCJFHR1032B02	BAA02304	4e-16	TaHBP1a	72/76	18
SCSFFL4085D03	AAB40291	2e-30	OSBZ8	68/74	19
SCEZLR1031F04	T02084	e-114	GBF1	87/91	
SCBGLR1023B02	BAB85261	4e-42	OsGBF8	74/80	7
Family X					
SCEQLR1093A06	T08591	1e-18	STF1	53/63	-
SCVPRZ2040C12	BAC07004	e-146	OsbZIP	67/75	-
SCCCCL4011E09	AAK76638	2e-11	AtbZIP60	38/51	-
SCJLRT1019E07	AAK76638	4e-07	AtbZIP60	34/48	-
SCVPRZ2036C12	BAC07004	8e-54	OsbZIP	43/57	-
SCAGLB1071F05	BAB62558	2e-14	OsTHY5	98/100	7
SCCCRZ2004A09	BAB62558	3e-08	OsTHY5	55/62	7
Family XI					
SCVPCL6047G09	AAM20732	1e-21	MUK11.16	48/63	-

SCCCRZ2001B01	AAO42021	1e-23	AtbZIP	57/77	-
SCACHR1037A09	BAB83611	6e-23	AtbZIP	36/45	-
SCUTFL1056A06	At2g42380	1e-08	AtbZIP34	72/81	-
SCEQRT2097C12	At2g42380	6e-57	AtbZIP34	48/59	-
SCEQLB1064D03	BAA96162	3e-83	OsRF2a-like	82/86	7
SCSBST3100F06	BAA96162	3e-41	OsRF2a-like	79/84	7
SCSGHR1069F07	BAA96162	1e-32	OsRF2a-like	59/67	7
Family XII					
SCUTAM2008F01	AAL78371	2e-52	OsVsf-1	81/88	-
SCCCRT2002D06	BAC84100	e-101	RF2a-like	87/91	-
SCUTRZ3073E03	AAM94510	2e-32	OsbZIP	81/89	-
SCQGRT1043E05	AAL87668	3e-61	RVS1	81/87	-
SCCCAM1003G07	AAM19114	4e-23	OsbZIP	77/81	-
SCRUAD1061C02	CAD41260	6e-08	POSF21	46/57	21
SCJFRZ2034D03	CAD41260	1e-68	OsbZIP	91/94	22
SCSFLR2016F09	AAR28765	e-111	RF2b	78/82	23
SCRFRZ3056F07	AAR28765	1e-42	RF2b	76/80	23
SCVPCL6065H07	AAC49832	4e-42	RF2a	67/78	24
SCCCAM1001G08	AAO72544	1e-59	OsbZIP	80/85	25
SCJFHR1C05H04	AAO72544	1e-29	OsbZIP	65/75	25
SCACAM2043F07	BAB39175	1e-11	RISBZ5	57/85	11

References: 1- Feng et al., 2002; 2- Walsh et al., 1998; 3- Foley et al., 1993; 4- Niggeweg & Gatz,1997; 5- Kaneko et al.,1998; 6- Casaretto & Ho,2003; 7- Sasaki et al.,2002; 8- Cooper et al.,2003; 9- Kim & Thomas,1998; 10- Schlögl et al.,2004; 11- Onodera et al., 2001; 12- Singh et al.,1990; 13- Rook et al., 1998; 14- Kusano et al.,1995; 15- Pysh et al., 1993; 16- Droger-Laser et al.,1997; 17- Nantel & Quatrano, 1996; 18- Mikami et al., 1994; 19- Nakagawa et al., 1996; 20- de Vetten & Ferl, 1995; 21- Aeschbacher et al., 1991; 22- Feng et al., 2002; 23- Dai et al., 2004; 24- Yin et al., 1997; 25- Cooper et al., 2003.

Table 2. Characteristics of developmentally-regulated bZIP in sugarcane plantlets through cDNA array analysis.

Cluster	Sub-family*	Expression**				Similar	AcNumber
		2d	4d	13d	22d		
Up-regulated genes							
SCRUHR1075A10	II	0.09	0.15	1.21	0.40	TGA.2	AAB68661
SCEQRT3020F04	II	0.06	0.14	0.60	0.20	liguleless2	T01415
SCCCRT1001E06	IX	0.10	0.12	0.42	0.41	TaHBP1a	BAA02304
SCEZLR1031F04	IX	0.05	0.04	0.04	0.14	GBF1	T02084
SCJLRT1019E07	X	0.07	0.12	0.41	0.35	AtbZIP60	AAK76638
SCSGAM1095E12***	V	0.25	0.44	0.12	0.01	PvbZIP	AAK25822
Down-regulated genes							
SCBFAD1046H12	II	0.12	0.09	0.08	0.07	OBF3.2	CAA48904

* The bZIP sub-families are described in Vincentz et al., 2001.
** The values correspond to the average of four spots in the same array.
*** This bZIP has been described in Schlögl et al., 2004.

Table 3 . Characteristics of ABA- and Meja-regulated bZIP in sugarcane through cDNA array analysis.

Cluster	Sub-family	Ratio*		Ac. Number		
ABA						
Up-regulated genes		***6h***	***12h***			
SCRURT2012D03	II	3.1	2.9	OsBZIP	CAD41728	T08813
SCCCRT1003G04	II	2.0	2.6	OsbZIP	CAD41728	T08813
Down-regulated genes		***6h***	***12h***			
SCJFLR1035B10	VI	0.5	0.4	CcbZIP	AAD21199	AAK92213
SCCCCL4005C09	VIII	0.3	0.2	RISBZ5	BAB39175	
SCMCST1051A01	VIII	0.3	0.2	BZI-1	AAL27150	
SCCCAM1003G07	XII	0.5	0.3	OsbZIP	AAM19114	
Meja						
Up-regulated genes		***1h***	***12h***			
SCJFLR1035G11	VI	1.4	1.7	OCSBF1	T03642	
SCEPAM1053B09	XII	1.3	1.8	POSF21	CAD41260	

* Each value represents the average of the expression ratios between the normalized relative intensity of each interval of treatment and control from two independent experiments.

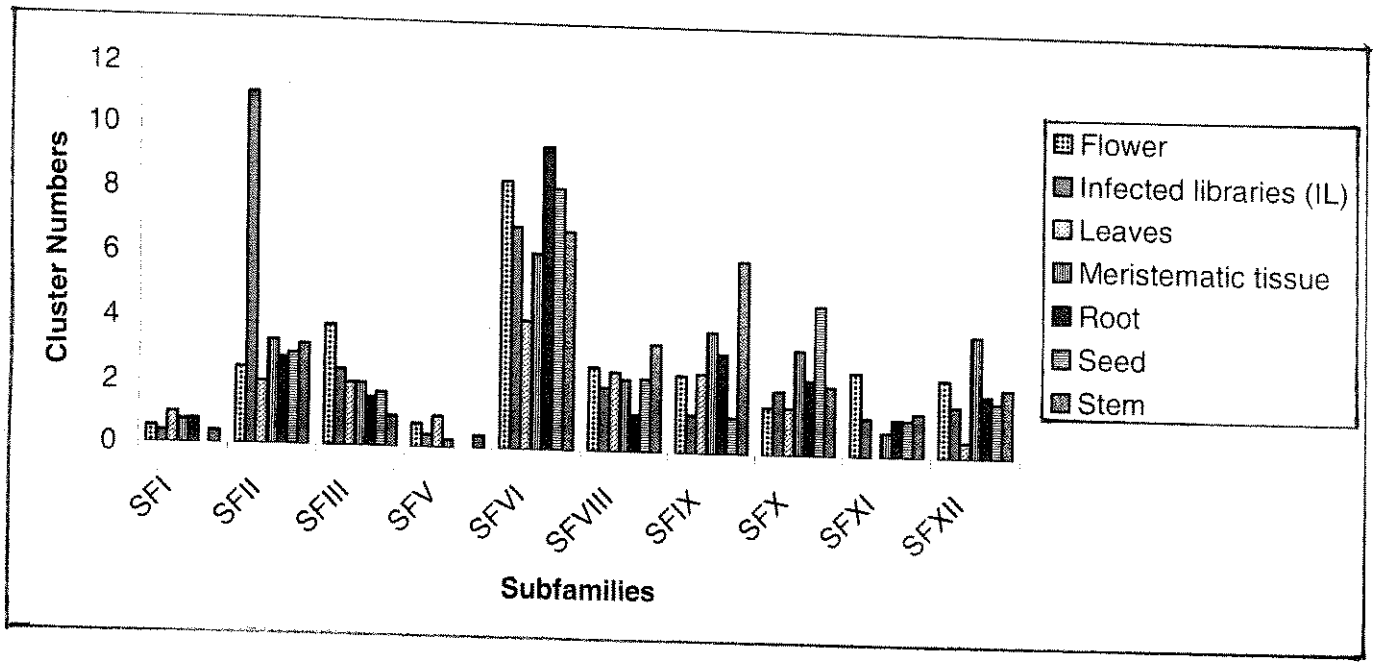

Cluster Numbers
12
10
8
6
4
2
0
SFI
SFII
SFIII
SFV
SFVI
SFVIII
SFIX
SFX
SFXI
SFXII
Subfamilies
Flower
Infected libraries (IL)
Leaves
Meristematic tissue
Root
Seed
Stem

Figure 1

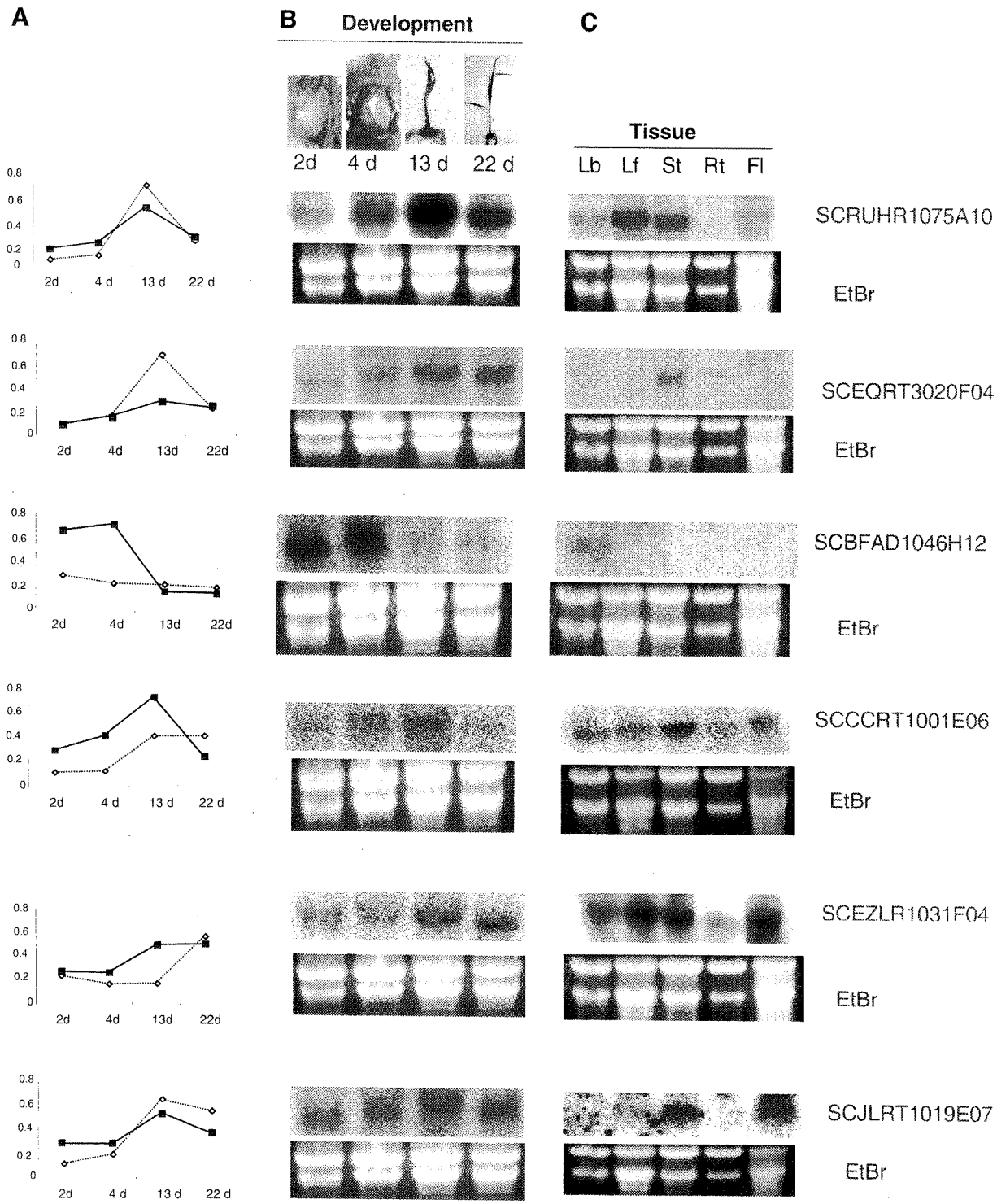
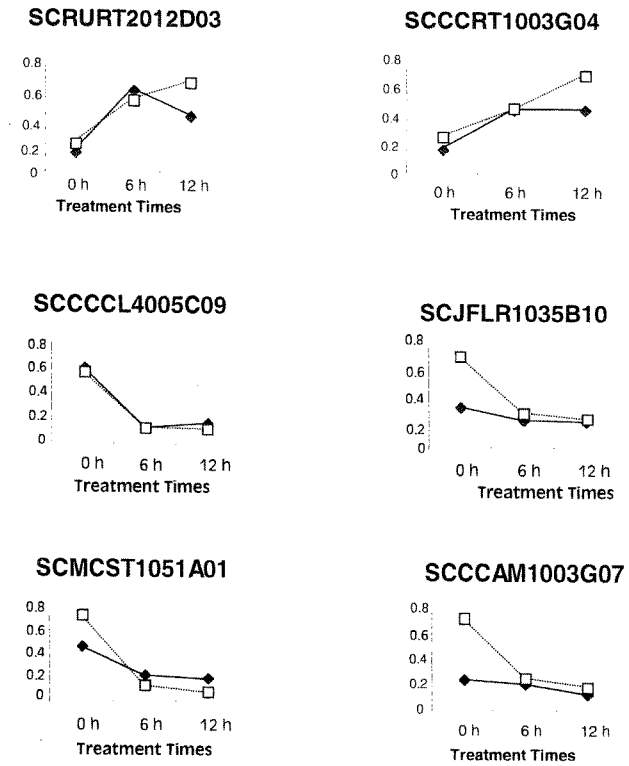

A
B
Development
C
Tissue
2d
4 d
13 d
22 d
Lb
Lf
St
Rt
Fl
0.8
0.6
0.4
0.2
0
SCRUHR1075A10
EtBr
SCEQRT3020F04
EtBr
SCBFAD1046H12
EtBr
SCCCRT1001E06
EtBr
SCEZLR1031F04
EtBr
SCJLRT1019E07
EtBr

Figure 2

A


SCRURT2012D03
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times
SCCCRT1003G04
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times
SCCCCL4005C09
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times
SCJFLR1035B10
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times
SCMCST1051A01
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times
SCCCAM1003G07
0.8
0.6
0.4
0.2
0
0 h
6 h
12 h
Treatment Times

B

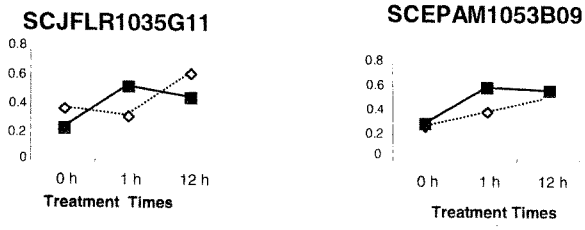

SCJFLR1035G11
0.8
0.6
0.4
0.2
0
0 h
1 h
12 h
Treatment Times
SCEPAM1053B09
0.8
0.6
0.4
0.2
0
0 h
1 h
12 h
Treatment Times

Figure 3

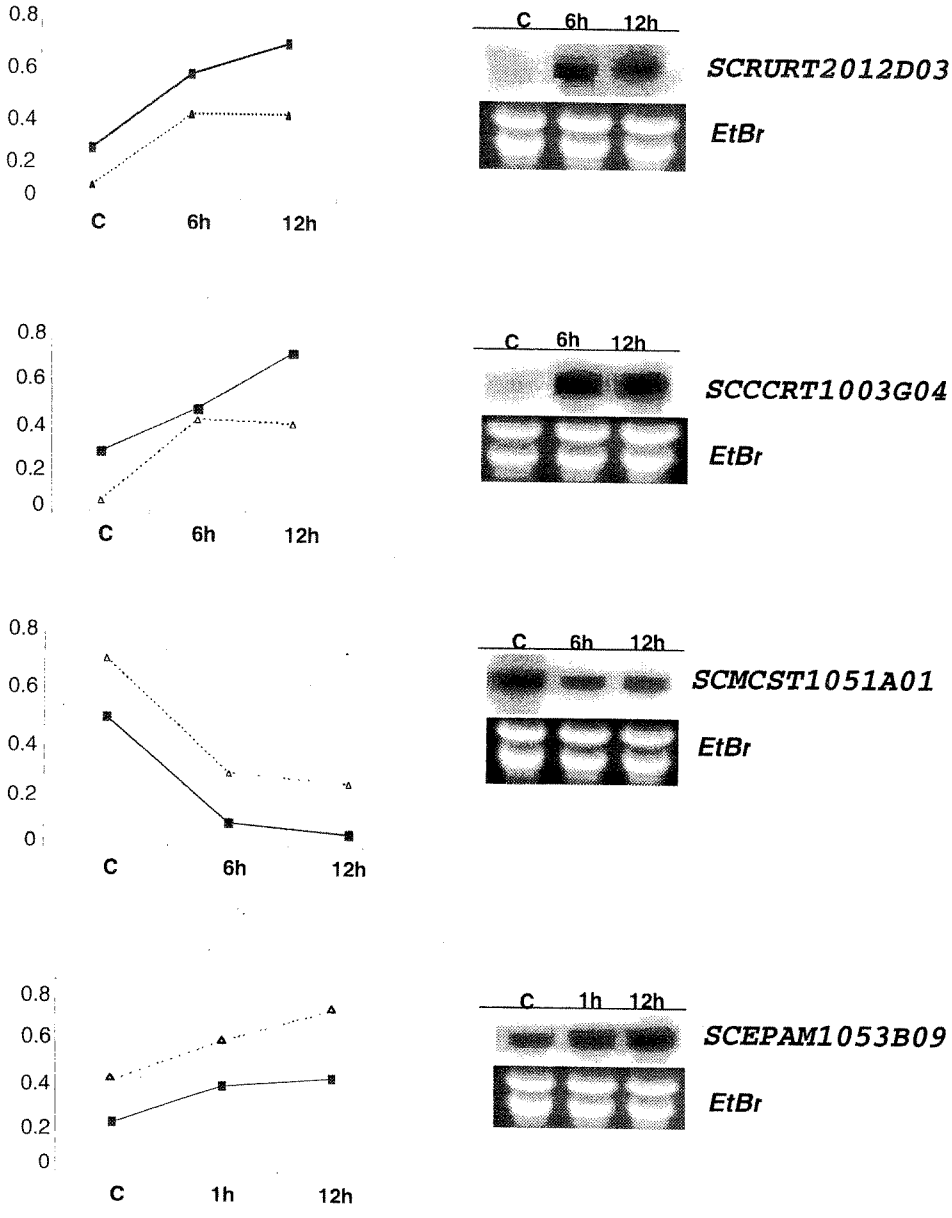

0.8
0.6
0.4
0.2
0
C
6h
12h
C 6h 12h
SCRURT2012D03
EtBr
0.8
0.6
0.4
0.2
0
C
6h
12h
C 6h 12h
SCCCRT1003G04
EtBr
0.8
0.6
0.4
0.2
0
C
6h
12h
C 6h 12h
SCMCST1051A01
EtBr
0.8
0.6
0.4
0.2
0
C
1h
12h
C 1h 12h
SCEPAM1053B09
EtBr

Figure 4

8. Conclusões

- O número de bZIPs encontrado no banco de dados do SUCEST foi similar ao número de bZIPs descritas no genoma de *Arabidopsis*.
- A sub-família Opaco2 é organizada em um grupo ortólogo entre monocotiledôneas e dicotiledôneas, três grupos de ortólogos entre monocotiledôneas e dois grupos entre dicotiledôneas.
- A evolução da sub-família Opaco2 envolveu duplicações gênicas antes da separação entre as monocotiledôneas e dicotiledôneas, além de outras três duplicações nas monocotiledôneas e uma nas dicotiledôneas.
- Opaco2 surgiu das duplicações gênicas que ocorreram especificamente nas monocotiledôneas, originando o que parece ser uma especialização funcional restritas as monocotiledôneas.
- SCbZIP1 apresenta alta afinidade de ligação *in vitro* aos oligonucleotídeos com os motivos G-box, C-box e Hex.
- A fosforilação de SCbZIP1 diminui sua afinidade por DNA e pode ser um mecanismo regulatório desse fator de transcrição em cana-de-açúcar.
- SCbZIP1 é expressa nos estágios iniciais de desenvolvimento das plântulas, não sendo induzida por exposição ao ácido abscísico ou frio. Seus transcritos são detectados apenas nas flores e gemas laterais. Esses resultados sugerem que esse fator de transcrição pode ser importante durante o desenvolvimento das plantas de cana-de-açúcar.
- Algumas bZIPs de cana-de-açúcar são diferencialmente expressas durante o desenvolvimento e outras são moduladas por ácido abscísico e metil-jasmonato. Esses resultados indicam que as bZIPs em cana-de-açúcar apresentam uma variada gama de funções em resposta a diferentes estímulos intrínsecos e externos.

9. Perspectivas

- A clonagem e caracterização bioquímica de outras proteínas bZIPs de cana-de-açúcar, incluindo mutantes, para estudos de interação DNA-proteína e proteína-proteína, certamente melhorarão os conhecimentos sobre esta importante classe de fatores de transcrição.
- Outros estudos de expressão em larga escala poderiam ser realizados com macroarranjos de DNA podendo assim, auxiliar na identificação de funções das bZIPs em cana-de-açúcar.
- Plantas geneticamente modificadas poderiam ser produzidas, de forma que estas seriam uma excelente ferramenta na caracterização funcional das bZIPs de cana-de-açúcar *in vivo*.
- Estudos de duplo híbrido e "one hybrid systems" poderiam ser realizados para encontrar proteínas que interagem com as bZIPs. Isto auxiliaria na caracterização de seus parceiros moleculares durante os diferentes mecanismos celulares em que as bZIPs estariam envolvidas.

10. Referências bibliográficas

Ahringer J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends Genet. **16**, 351-6.

Alland L, Muhle R, Hou H, Potes J, Chin L, et al. (1997). Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. Nature **387**, 49–55.

Amedeo P, Habu Y, Afsar K, Scheid OM and Paszkowski J. (2000). Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature. **405**, 203-6.

Asturias FJ, Meredith GD, Poglitsch CL and Kornberg RD. (1997). Two conformations of RNA polymerase II revealed by electron crystallography. J. Mol. Biol. **272**, 536– 40.

Auble DT, Hansen KE, Mueller CG, Lane WS, Thorner J and Hahn S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibitsTBPbinding to DNA by an ATP-dependent mechanism. Genes Dev. **8**, 1920–34.

Ayer DE. (1999). Histone deacetylases: transcriptional repression with SINers and NuRDs. Trends Cell. Biol. **9**,193–98.

Bentley D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. Curr. Opin. Cell. Biol. **11**, 347–51.

Benfey PN, Ren L and Chua NH. (1990a). Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. EMBO J. **9**, 1685-1696.

Benfey PN, Ren L and Chua NH. (1990b). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. EMBO J. **9**, 1677-1684.

Bordoli L, Netch M, Luthi U, Lutz W and Eckner R. (2001). Plant orthologs of p300/CBP: conservation of a core domain in metazoan p300/CBP acetiltransferase-related proteins. Nucleic Acid Research **29**, 589-597.

Bradsher JN, Tan S, McLaury HJ, Conaway JW and Conaway RC. (1993). RNA polymerase II transcription factor SIII. II. Functional properties and role in RNA chain elongation. J. Biol. Chem. **268**, 25594–603.

Brownell JE and Allis CD. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. **6**,176–84.

Brzeski J, Podstolski W, Olcczak K and Jerzmanowski A. (2001). Identification and analysis of the *Arabidopsis thaliana* BSH gene, a member of the SNF5 gene family. Nucleic Acid Research **27**, 2393-2399.

Burke TW and Kadonaga JT. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. **10**, 711–24.

Burke TW and Kadonaga JT. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev. **11**, 3020–31.

Chavez S and Beato M. (1997). Nucleosome mediated synergism between transcription factors on the mouse mammary tumor virus promoter. Proc. Natl. Acad. Sci. USA **94**:2885–90.

Chen W, Chao G and Singh K B. (1996). The promoter of a H2O2-inducible, Arabidopsis glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. Plant J. **10**, 955-966.

Chesnut JD, Stephens JH and Dahmus ME. (1992). The interaction ofRNApolymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. J.Biol. Chem. **267**,10500–6.

Chiu R, Angel P and Karin M. (1989). Jun-B differs in its biological properties from, and is a negative regulator of, c-jun. Cell **59**, 979-986.
Cho EJ, Takagi T, Moore CR and Buratowski S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. **11**,3319–26.
Cho H, Orphanides G, Sun X, Yang XJ, Ogryzko V, et al. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. Mol.
Cell. Biol. **18**, 5355–63.
Choi H, Hong J, Ha J, Kang J and Kim SY. (2000).ABFs, a family of ABA-responsive element binding factors. The Journal Biol. Chem. **275**, 1723-1730.
Choder M and Young RA. (1993). A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. Mol. Cell. Biol. **13**, 6984–91.
Chuang C-F, Running MP, Williams RW and Meyerowitz EM. (1999). The Perianthia gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. Genes and Development **13**, 334-344.
Ciceri P, Castelli S, Lauria M, Lazzari B, Genga A, Bernard L, Sturaro M, Viotti A. (2000). Specific combinations of zein genes and genetic backgrounds influence the transcription of the heavy-chain zein genes in maize opaque-2 endosperms.
Plant Physiol. **124**, 451-60.
Cirillo LA and Zaret KS. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**:961–69.
Coin F and Egly JM. (1998). Ten years of TFIIH. Cold Spring Harbor Symp. Quant. Biol. **63**, 105–10.
Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B and Kornberg D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science **288**,640–49
Conaway JW, Dvir A, Moreland RJ, Yan Q, Elmendorf BJ, et al. (1998). Mechanism of promoter escape by RNA polymerase II. Cold Spring Harbor Symp. Quant. Biol. **63**, 357–64.
Cosma MP, Tanaka T and Nasmyth K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle– and developmentally-regulated promoter. Cell **97**, 299–311.
Dangl M, Brosch G, Haas H, Loidl P and Lusser A. (2001). Comparative analysis of HD2 type histone deacetylases in higher plants.
Planta **213**, 280-5.
Deyholos MK and Sieburth LE. (2000). Separable whorl-specific expression and negative regulation by enhancer elements within the AGAMOUS second intron.
Plant Cell. **12**, 1799-810.
Dubois MF, Nguyen VT, Bellier S and Bensaude O. (1994). Inhibitors of transcription such as 5, 6-dichloro-1-beta-D-ribofuranosylbenzimidazole and isoquinoline sulfonamide derivatives (H-8 and H-7) promote dephosphorylation of the carboxyl-terminal domain of RNA polymerase II largest subunit. J. Biol. Chem. **269**, 13331–36.
Edwards AM, Kane CM, Young RA and Kornberg RD. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J. Biol. Chem. **266**, 71–75.
Eshed Y, Baum SF and Bowman JL. (1999). Distinct mechanisms promote polarity establishment in carpels of Arabidopsis. Cell. **99**,199-209.

Fan W and Dong X. (2002). In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in Arabidopsis.
Plant Cell **14**, 1377-1389.
Feaver WJ, Gileadi O, Li Y and Kornberg RD. (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. Cell **67**, 1223–30.
Finkelstein RR and Lynch TJ. (2000). The *Arabidopsis* Abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. The Plant Cell **12**, 599-609.
Finnegan EJ. (2001). Is plant gene expression regulated globally?. Trends in Genetics **17**, 361-365.
Foster R, Izawa T and Chua N-H. (1994). Plant bZIP proteins gather at ACGT elements. FASEB J. **8**, 192-200.
Garcia-Ramirez M, Rocchini C, Ausio J. (1995). Modulation of chromatin folding by histone acetylation. J. Biol. Chem. **270**,17923–28.
Ge H and Roeder RG. (1994). The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. J. Biol. Chem. **269**, 17136–40.
Gilmartin PM, Sorokin, L.; Memelink, J. e Chua, N. (1990). Molecular light switches for plant genes. *Plant Cell* **2**: 369-378.
Gille H, Sharrocks A and Shaw P. (1992). Phosphorylation of p62TCF by MAP kinase stimulates ternary complex formation at cFos promoter. *Nature* **358**: 414-417.
Gold MO, Yang X, Herrmann CH and Rice AP. (1998). PITALRE, the catalytic subunit of TAK, is required for human immunodeficiency virus Tat transactivation in vivo. J. Virol. **72**, 4448–53.
Gonzales GA and Montminy M. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell **59**, 675-680.
Hebbes TR, Clayton AL, Thorne AW and Crane-Robinson C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J. 13:1823–30
Hengartner CJ, Thompson CM, Zhang J,Chao DM, Liao SM, et al. (1995). Association of an activator with an RNA polymerase II holoenzyme. Genes Dev. **9**:897–910.
Hill SH and Treisman R. (1995). Transcriptional regulation by extracellular signals: Mechanism and specificity. Cell **80**, 199-211.
Hoecker U, Vasil IK and McCarty DR. (1995). Integrated control of seed maturation and germination programs by activator and repressor functions of Viviparous-1 of maize. Genes Dev. **9**, 2459-69.
Holstege FC, Jennings EG,Wyrick JJ, LeeTI, Hengartner CJ, et al. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell **95**:717–28.
Holstege F, Fiedler U and Timmers H. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J. **16**, 7468–80.
Hunter T and Karin M. (1992). The regulation of transcription by phosphorylation. Cell **70**, 375-387.
Hunter T. (1995). Protein Kinases and Phosphatases: the Yin and Yang of protein phosphorylation and signalling. Cell **80**, 225-236.
Imhof A,Yang XJ, OgryzkoVV, Nakatani Y, Wolffe AP and Ge H. (1997). Acetylation of general transcription factors by histone acetyltransferases. Curr. Biol. **7**,689–92.
International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature **409**, 860-921.

Izawa T, Foster R, and Chua N. (1993). Plant bZIP protein DNA binding specificity. Journal of molecular Biology **230**, 1131-1144.
Keller W, König P and Richmond TJ. (1995). Crystal structure of a bZIP/DNA complex at 2.2 A: determinants of DNA specific recognition. J.Mol.Biol. **254**, 657-667.
Kerstetter R, Vollbrecht E, Lowe B, Veit B, Yamaguchi J and Hake S. (1994). Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. Plant Cell **6**, 1877-1887.
Kornberg RD. (1999). Eukaryotic transcriptional control. Trends in Cell Biol. **9**, M46-M49.
Kouzarides T and Ziff E. (1989). Leucine zipper of *fos, jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. Nature **340**, 568-571.
Kuhlemeier C. (1992). Transcriptional and post-transcriptional regulation of gene expression in plants. Plant Mol.Biol. **19**, 1-14.
Kutach AK and Kadonaga JT. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters.
Mol Cell Biol. **20**, 4754-64.
Latchman DS. (1997). Transcription factors: an overview. Int. J. Biochem. Cell Biol. **12**:1305-1312.
Lee KI and Young RA. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. **37**, 77-137.
Li G, Chandler SP, Wolffe AP and Hall TC. (1998).Architectural specificity in chromatin structure at the TATA box in vivo: nucleossomo displacement upon β-phaseolin gene activation. Proc. Natl. Acad. Sci USA. **95**, 4772-4777.
Li G, Bishop KJ, Chandrasekharan MB and Hall TC. (1999). β-phaseolin gene activation is a two-step process: PvALF-facilitated chromatin modification followed by Abscisic acid-mediated gene activation. Proc. Natl. Acad. Sci USA. **96**, 7104-7109.
Li G, Chandrasekharan MB, Wolffe AP and Hall TC. (2001a). Chromatin structure and β-phaseolin gene regulation. Plant Mol. Biol. **46**, 121-129.
Lohmann JU, Hong RL, Hobe M, Bush MA, Parcy F, Simon R and Weigel D. (2001). A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. Cell **105**, 793-803.
Lupas A. (1996). Coiled coils: new structures and new functions.
Trends Biochem Sci. **21**, 375-82.
Meshi T and Iwabuchi M. (1995). Plant transcription factors. Plant Cell Physiol. **36(8)**, 1405-1420.
Myers LC, and Kornberg RD. (2000). Mediator of transcriptional regulation.
Annu Rev Biochem **69**, 729-49.
Näär AM, Lemon BD and Tijan R. (2001). Transcriptional coactivator complexes. Annu. Rev. Biochem. **70**, 475-501.
Niggeweg R, Thurow C, Kegler C and Gatz C. (2000). Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid-and auxin-inducible expression of as-1-containing target promoters. The Journal Biol. Chem. **275**, 19897-19905.
Ogas J, Kaufmann S, Henderson J and Somerville C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis.
Proc Natl Acad Sci U S A. **96**, 13839-13844.

Osterlund MT, Hardtke CS, Wei N and Deng XW. (2000). Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. Nature **405**, 462-466.
O'Shea EK, Klemm JD, Kim PS and Alber T. (1991). X-ray Structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science **254**, 539-544.
O'Shea-Greenfield A and Smale ST. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J. Biol. Chem. **267**, 1391-1402.
Patthy L. (2003). Modular assembly of genes and the evolution of new functions. Genetica. **118**, 217-31.
Pham AD and Sauer F. (2000). Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in Drosophila.Science **289**, 2357–60.
Purugganan MD, Rounsley SD, Schmidt RJ and Yanofsky MF. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**: 345-356.
Ramachandran S, Hiratsuka K and Chua, N. (1994). Transcription factors in plant growth and development. Curr.Opin.Genet.Dev. **4**, 642-646.
Riechmann JL. (2002). Transcriptional Regulation: a genomic overview. The Arabidopsis book, pg 2-46.
Riechmann JL and Ratcliffe OJ. (2002). A genome perspective on plant transcription factors. Current Opinion in Plant Biology 3, 423-434.
Rook F, Gerrits N, Kortstee A, van Kampen M, Borrias M, Weisbeek P and Smeekens S. (1998). Sucrose-specific signalling represses translation of the *Arabidopsis* ATB2 bZIP transcription factor gene. The Plant Journal **15**, 253-263.
Rossi V, Hartings H and Motto M. (1998). Identification and characterization of an RPD3 homologue from maize (*Zea mays* L.) that is able to complement an *rpd3* null mutant of *Saccharomyces cerevisae*. Mol. Gen. Genet. **258**, 288-296.
Schultz TF, Medina J, Hill A and Quatrano RS. (1998). 14-3-3 proteins are part of an Abscisic acid-VIVIPAROUS (VP1) response complex in the *Em* promoter and interact with VP1 and EmPB-1. Plant Cell **10**, 837-847.
Shua K, Stark GR, Kerr I. and Darnell JEJ. (1993). A single phosphotyrosine residue of Stat91 required for gene activation by interferon-gamma. Science **261**, 1744-1746.
Singh KB. (1998). Transcriptional regulation in plants: the importance of combinatorial control. Plant Physiol. **118**, 1111-1120.
Singer T, Yordan C and Coupland G. (2001). Robertson's mutator transposons in *Arabidopsis thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA methylation (DDM1). Genes Dev. **15**, 591-602.
Sodek L and Wilson CM. (1970). Incorporation of leucine-C^{14} and lysine-C^{14} into protein in the developing endosperm of normal and opaque2 corn. Arch.Biochem.Biophys. **140**, 29-38.
Smale ST and Kadonaga JT. (2003). The RNA polymerase II core promoter. Annu. Rev. Biochem. **72**, 449-479.
Smeal T, Angel P, Meek J and Karin M. (1989). Different requirements for formation of Jun: Jun and Jun: Fos complexes. Genes Dev. **12B**:2091-2100.
Schutte J, Viallet J, Nau M, Segal S, Fedorko J and Minna J. (1989). jun-B inhibits and c-fos stimulates the transforming and trans-activating activities of c-jun. Cell **59**:987-997.
Takatsuji H, Nakamura N and Katsumoto Y. (1994). A new family of zinc finger proteins in petunia: structure, DNA sequence recognition, and floral organ-specific expression. Plant Cell **6**, 947-958.

Tjian R and Maniatis T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell. **77**, 5-8.
Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K and Yamaguchi-Shinozaki K. (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc. Natl. Acad. Sci. USA **97**, 11632-11637.
Vettore AL, Yunes JA, Cord Neto G, daSilva MJ, Arruda P and Leite A. (1998). The molecular and functional characterization of na *Opaque-2* homologe protein from *Coix* and new classification of plant bZIP proteins. Plant Mol. Biol. **36**, 249-263.
Vicente-Carbajosa J, Moose SP, Parsons R L and Schmidt RJ. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. Proc. Natl. Acad. Sci. USA **94**, 7685-7690.
Vincentz M, Schlögl PS, Corrêa LG, Kuhne F and Leite A. (2001). Phylogenetic relationships between *Arabidopsis* and Sugarcane bZIP transcriptional regulatory factors. Genetic and Molecular Biology **24**, 55-60.
Wagner S and Green MR. (1994). DNA-binding domains: targets for viral and cellular regulators. *Curr. Opin. Cell Biol.* **6**: 410-414.
Weis L and Reinberg D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Mol. Cell Biol. **17**:2973-84.
Williams ME, Foster R and Chua N. (1992). Sequences flanking the hexameric G-Box core CACGTG affect the specificity of protein binding. The Plant Cell **4**, 485-496.
Wu K, Malik K, Tian L, Brown D and Miki B. (2000a). Functional analysis of a PPD3 histone deacetilase homologue in *Arabidopsis thaliana*. Plant Mol. Biol. **44**, 167-176.
Yin Y, Zhu Q, Da S, Lamb C and Beachy R. (1997). RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, function in vascular development. EMBO J. **16**, 5247-5259.
Yu CH, Bolouri H and Davidson EH. (2001). Cis-regulatory logic in the *endo*16 gene: switching from a specification to a differentiation mode of control. Development **128**, 617-629.
Zhu Q, Ordiz MI , Dabi T, Beachy R N and Lamb C. (2002). Rice TATA binding protein interacts functionally with transcription factor IIB and the RF2a bZIP transcriptional activator in an enhanced plant in vitro transcription system. Plant Cell. **14**, 795–803.