Érika Maria de Jesus

**“Estudo de Dois Grupos de Elementos de
Cana-de-Açúcar Homólogos à
superfamília *hAT* de Transposons”**

**São Paulo
2007**

Érika Maria de Jesus

**“Estudo de Dois Grupos de Elementos de
Cana-de-Açúcar Homólogos à
superfamília *hAT* de Transposons”**

Tese apresentada ao Instituto de
Biociências da Universidade de São Paulo,
como parte dos requisitos para a obtenção
do Título de Doutor em Ciências, na área
de Botânica.

Orientadora: Profa. Dra. Marie-Anne Van Sluys

**São Paulo
2007**

Ficha Catalográfica

Jesus, Érika Maria de

Estudo de dois grupos de elementos de cana-de-açúcar homólogos à superfamília *hAT* de transposons / Érika Maria de Jesus. – São Paulo : E.M.J., 2007.

172 páginas.

Tese (Doutorado) - Universidade de São Paulo, Instituto de Biociências, Departamento de Botânica, 2007.

1. genética molecular de plantas 2. transposon 3. transposase 4. superfamília hAT

I. Universidade de São Paulo. Instituto de Biociências. Departamebnto de Botânica. Título.

QH433

**“Estudo de dois grupos de elementos de cana-de-açúcar
homólogos à superfamília *hAT* de transposons “**

Candidata: Érika Maria de Jesus.

Orientadora: Profa. Dra. Marie-Anne Van Sluys

Comissão Julgadora:

Prof.(a) Dr.(a)

Prof.(a) Dr.(a)

Prof.(a) Dr.(a)

Prof.(a) Dr.(a)

Prof.(a) Dr.(a) Marie-Anne Van Sluys

“Tive o prazer de vivenciar a excitação criada por mudanças revolucionárias nos conceitos genéticos, que ocorreram ao longo dos sessenta e poucos anos passados. Eu acredito que nós estamos experimentando uma revolução como esta novamente, que é a alteração de nossos conceitos sobre o genoma: as partes que o compõem, sua organização, mobilidade, e seu modo de operação.”

Barbara McClintock, Science 226:792-801 (1984), após receber o Prêmio Nobel de Medicina.

**“Ando devagar porque já tive pressa
E levo esse sorriso, porque já chorei demais
Hoje me sinto mais forte, mais feliz quem sabe
Só levo a certeza de que muito pouco eu sei, eu nada sei...
...É preciso amor para poder pulsar,
É preciso paz para poder sorrir,
É preciso a chuva para florir...”**

Almir Sater e Renato Teixeira

Dedico esta tese

**A meus pais, Tobias e Cândida
pelo apoio incondicional que sempre me deram,
e por me ensinarem o amor e a curiosidade pela natureza.**

AGRADECIMENTOS

À Profa. Dra. Marie-Anne Van Sluys, pela oportunidade, confiança, incentivo e amizade. Pela sua orientação na realização deste trabalho, sempre trazendo uma idéia à frente do convencional.

À Profa. Dra. Magdalena Rossi, pela preciosa co-orientação e amizade, sempre me ajudando desde que assumi este projeto no laboratório. Também pelas enriquecedoras conversas em que discutimos projetos e vida profissional.

Ao Prof. Dr. Josep Casacuberta, o Pep, pela colaboração no ensaio de PCR-Inversa e pelas discussões sobre o meu trabalho, sempre com uma visão crítica muito aguçada.

À Dra. Paula Araújo, pela confiança em me delegar a responsabilidade pela coleção de clones de cDNA e a continuidade de sua análise.

À Profa. Dra. Marina Passeto Nóbrega, minha primeira orientadora de iniciação científica. Me ensinou muito sobre manipulações em biologia molecular e sempre me incentivou a buscar a independência. Ensinaamentos que levarei para toda a vida.

Ao Prof. Dr. Armando Morais Ventura, que me orientou em meu segundo projeto de iniciação científica, etapa em que pude amadurecer e adquirir confiança quanto ao planejamento e realização de experimentos.

Ao Prof. Dr. Carlos Menck, com quem nunca tive uma colaboração de fato, mas contribuiu de maneira indireta para minha formação. Desde a época da IC, e até hoje, sempre que eu preciso, bato na porta do seu laboratório pra pedir uma mãozinha.

A todos os professores do Departamento de Botânica do IB, pelas aulas da pós e da graduação em geral, que estimularam meu interesse e paixão pela botânica. Também pelo compartilhamento harmonioso de laboratórios, equipamentos e conhecimento.

À Profa. Dra. Nanuza, por poder compartilhar de sua sensibilidade e gentileza incomuns, e por sua contagiante alegria de viver e amor pela natureza.

À Dra. Myna Nakabashi, pela imprescindível ajuda na recuperação de elementos *hAT* do genoma da cana-de-acúcar, sua clonagem e sequenciamento. Também pela contribuição com seu senso crítico nas discussões sobre este trabalho.

À Juliane Karine Ishida, pelo companheirismo e ajuda essencial nos momentos críticos da realização do ensaio de Northern Blot.

À Profa. Dra. Kátia Scortecchi por ceder a sonda da ubiquitina utilizada no ensaio de Northern Blot.

À Daniela Kajihara, por toda a sua dedicação em cuidar do material vegetal utilizado neste trabalho, sempre com muito capricho e carinho.

Ao Alessandro Varani, pelo valioso auxílio na parte de bioinformática.

À Profa. Dra. Ana Paula P. Costa, pelas muitas discussões em que compartilhou seu conhecimento sobre biologia molecular e elementos de transposição, pelas inúmeras dicas científicas e culturais e pelas ótimas conversas sobre cinema e música.

À Profa. Dra. Regina Hashimoto Miura, que me acolheu no laboratório com muito carinho desde o primeiro dia em que cheguei aqui. Com sua atenção e generosidade me ajudou muito.

Ao Douglas Silva Domingues, companheiro de algumas tentativas de cultivo de plantas de cana-de-acúcar, e tantas extrações de DNA genômico vegetal.

À Leonor, pelo precioso trabalho que dá suporte ao nosso dia-a-dia e impede que o laboratório se transforme num caos.

À Profa. Dra. Hana Masuda, pelas discussões enriquecedoras, e por ter sempre uma dica quentíssima na hora certa para compartilhar.

À Silvia Blanco Ribeiro, por toda a ajuda inestimável no sequenciamento de amostras e na resolução de pepinos em geral do laboratório. Também por sua amizade com a qual sempre pude contar nos momentos difíceis.

À Maria Elisa Manetti, de quem pude desfrutar da inteligência e cultura em filosóficas conversas. Nossa amizade é uma prova de que pessoas tão diferentes podem ter tanta afinidade.

À Marisa Mommoli, amiga indispensável que sempre me apoiou e foi companheira em questões científicas ou não. Sempre com uma palavra sensata, me ajudou muito todos estes anos.

À Daniela Milstein, por nossa amizade que se estreitou neste último ano em que ficamos no mesmo laboratório. Suas conversas animadas sempre me colocam pra cima.

Aos amigos do GaTE lab, que direta ou indiretamente me ajudaram neste trabalho, pelas discussões científicas e por fazerem deste laboratório um ambiente de convivência agradável e animado:

Colegas das antigas: Patrícia, Priscila, Paula, Denise, Regina, Adriana, Elis, Ivone, Beth e Luciana Terumi.

Do grupo atual: Alessandro, Ana Paula, Bruno, Dani Kajihara, Dani Milstein, Douglas, Guilherme, Hana, Jonas, Juliane, Leonor, Luisa, Marcelo, Maria Elisa, Marisa, Myna, Nilo, Silvia, Úrsula, Vanessa Sato, Vivian e Wanessa.

Aos colegas do Departamento de Botânica de ontem e de hoje, com destaque para Bia Maia, Rogério Suzuki, Auri, Luciano, Cíntia, Thaís, Kátia Campos, Adélia,

Wagner, Rosete, Camila e Claudete (Fisiologia), Luciana Bastos (Algas) e Fábio (Filogenia Molecular) pela convivência harmoniosa.

A todos os funcionários do Instituto de Biociências que de alguma maneira colaboraram dando suporte para a realização deste trabalho, principalmente Helder, Irineu e Vilma da seção de alunos, sempre muito prestativos, e Ana Maria, Linácia e Suzi, sempre animadas e descontraindo o ambiente.

Às amigas da Bio que seguem firmes e fortes: Juliana e Gabriela, amigas pra vida toda.

Às amigas de casa, Thaís, Michele Minzon, Michele Figueiró, Flavinha e Andréa, que foram minhas irmãs durante todos estes anos de USP, tornando nossa convivência uma amizade.

Às amigas de vida, Sandra e Lucila, com quem sei que posso sempre contar como um ombro amigo.

A toda minha família, tios e primos, pelo apoio e pela torcida em todos os desafios da minha vida, principalmente Walnedie, Cristina, Marli, Ana, Tati, Anderson, Glauco, Gisele e Caio.

Especialmente para meu irmão Gustavo e minha cunhada Muriel, que são verdadeiros amigos e geraram meu sobrinho Guilherme que é a alegria da família. A toda a nova família “americana”, principalmente à Inodi que cuida de todos com muito carinho.

Ao Eduardo Soares, o Edu, pelos muitos anos de convivência, compartilhando afinidades e o gosto pela Ciência e pela música. Por sua paciência e ajuda em momentos cruciais da realização deste projeto.

Aos meus pais por tudo. Ao meu pai, Tobias, que me apoiou de maneira irrestrita e incondicional durante toda minha vida. À minha mãe, Cândida, que é para mim o maior exemplo de pessoa generosa, e o maior apoio emocional da minha vida. Amo vocês!

À Marisa, Silvia, Hana, Marie, Douglas ... Muito obrigada pela força pra escrever a tese !!! Não há palavras que possam agradecer.

Agradeço à FAPESP pelo apoio financeiro e concessão de bolsa que permitiram a realização deste trabalho.

Por fim, agradeço a muitas outras pessoas que foram importantes na minha formação. São muitos anos de estrada e seria impossível listar todos que de alguma maneira influenciaram meu caminho.

Obrigada !!!

ABREVIATURAS

BAC – do inglês “bacterial artificial chromosome”, cromossomo artificial de bactéria

BSA – do inglês “bovinum serum albumine”, albumina de soro bovino

cDNA – DNA complementar

DEPC – dietilpirocarbonato

DNA – ácido desoxirribonucleico

DNase - desoxirribonuclease

dNTP- deoxirribonucleotídeos trifosfato

2,4-D - ácido 2,4-diclorofenilacético

DTT – Ditioneitol

EDTA – ácido etilenodiamonotetracético

EST – do inglês “expressed sequence tag”, fragmento de sequência expressa

°C – grau Celsius

IPTG – isopropil tio-β-D-galactosídeo

Kb – kilo pares de base

μL, μg – microlitro, micrograma

mL, mg – mililitro, miligrama

mRNA – RNA mensageiro

ng - nanograma

ORF – do inglês “open reading frame”, quadro aberto e leitura

pb – pares de base

PCR – do inglês “polymerization reaction chain”, reação de polimerização em cadeia.

PVP – Polivinilpirrolidona

PVPP – Polivinilpolipirrolidona

RNA – ácido ribonucleico

RNase - ribonuclease

SDS – dodecil sulfato de sódio

siRNA – do inglês “small interference RNA”, pequeno RNA de interferência

SNP – do inglês “single nucleotide polymorphism”, polimorfismo de nucleotídeo único

TBE - Tris-borato-EDTA

TE – Tris-EDTA

TE – do inglês “transposable element”, elemento de transposição

“TE221- RP” – reconstrução parcial das porções 5’ e 3’ do cDNA TE221

X-Gal – 5-bromo-4-cloro-3-indolil-β-D-galactosídeo

ÍNDICE

RESUMO.....	17
ABSTRACT.....	18
I. INTRODUÇÃO.....	19
I.1. Elementos de Transposição – O Genoma Dinâmico.....	19
I.2. Diversidade dos Elementos de Transposição.....	24
I.3. Superfamília <i>hAT</i> de transposons.....	28
I.4. A cana-de-açúcar (<i>Saccharum</i> spp.).....	32
II. OBJETIVOS.....	39
II.1. Objetivo Geral.....	39
II.2. Objetivos Específicos.....	39
III. MATERIAIS E MÉTODOS.....	40
III.1. Material Vegetal e Condições de Crescimento.....	40
III.1.1. Material Vegetal Utilizado para Extração de DNA.....	40
III.1.2. Material Vegetal Utilizado para Extração de RNA.....	40
III.2. Cepa Bacteriana Utilizada (<i>E. coli</i>)	41
III.3. Meios de Cultura.....	41
III.3.1. Meio de Cultura para Calos.....	41
III.3.2. Meios de Cultura para Bactérias.....	41
III.4. Clones de cDNA de Cana-de-açúcar.....	42
III.5. Eletroforese e Géis de Agarose.....	42
III.6. Manipulações Gerais com Bactérias <i>Escherichia coli</i>	43
III.6.1. Obtenção de Bactérias Competentes.....	43
III.6.2. Transformação de Bactérias Competentes.....	44
III.7. Procedimentos Gerais de Manipulação de DNA.....	44
III.7.1. Extração de DNA Plasmidial em Pequena Escala (Miniprep)	44
III.7.2. Extração de DNA Plasmidial em Pequena Escala (Miniprep) em Placas de 96 Poços.....	45
III.7.3. Extração de DNA Genômico de Plantas.....	47
III.7.4. Sequenciamento Automático de Nucleotídeos.....	48
III.8. Sequenciamento Completo dos Clones de cDNA do SUCEST.....	49
III.8.1. Estimativa do Tamanho dos Clones de cDNA.....	49

III.8.2. Sequenciamento de Clones Menores que 1 Kb.....	50
III.8.3. Sequenciamento de Clones de Tamanho entre 1 Kb e 3,5 Kb.....	52
III.8.4. Sequenciamento de Clones de Tamanho Maior que 3,5 Kb.	52
III.8.4.1. Sonicação do DNA e Subclonagem dos Fragmentos.....	52
III.8.4.2. Sequenciamento das Sub-Bibliotecas e Montagem das Sequências.....	53
III.9. Análise das Sequências Completas dos Clones do SUCEST.....	53
III.10. Análise da diversidade entre os cDNAs homólogos à superfamília <i>hAT</i>	54
III.10.1. Alinhamento de Domínios Conservados das Transposases do tipo <i>hAT</i> dos cDNAs de cana-de-açúcar...54	
III.10.2. Alinhamento de um Domínio Conservado das Transposases do tipo <i>hAT</i> presente nos cDNAs de cana-de-açúcar e no gene DAYSLEEPER.....	54
III.10.3. Busca por Genes Homólogos aos cDNAs de Cana-de-Açúcar no Genoma do Arroz.....	55
III.10.4. Alinhamento dos Clones de cDNA do grupo I com Transposases Domesticadas do tipo <i>hAT</i>	55
III.10.5. Geração de Árvores Filogenéticas a partir de Alinhamentos.....	56
III.11. Distribuição dos Elementos Homólogos à Superfamília <i>hAT</i> nos Genomas de Cana-de-açúcar e Outras Monocotiledôneas.....	56
III.11.1. Hibridação Molecular.....	56
III.11.1.1. Sondas.....	56
III.11.1.2. Hidrólise do DNA Genômico.....	58
III.11.1.3. Gel de Agarose.....	58
III.11.1.4. Southern Blot (Transferência Alcalina).....	60

III.11.1.5. Pré-Hibridação.....	60
III.11.1.6. Marcação das Sondas.....	60
III.11.1.7. Hibridação.....	61
III.11.1.8. Lavagens das Membranas.....	61
III.11.1.9. Exposição e Revelação.....	61
III.12. Estudo da Expressão dos Elementos Homólogos à Superfamília <i>hAT</i> de Cana-de-açúcar.....	62
III.12.1. Hibridação Molecular de RNA.....	62
III.12.1.1. Extração de RNA Total.....	62
III.12.1.2. Gel de Formaldeído.....	63
III.12.1.3. Transferência (Northern Blot).....	64
III.12.1.4. Sondas.....	65
III.12.1.5. Marcação das Sondas.....	65
III.12.1.6. Hibridação.....	65
III.12.1.7. Lavagem das Membranas.....	65
III.12.1.8. Exposição e Revelação.....	66
III.12.1.9. Análise das Autorradiografias Geradas com a Hibridação.....	66
III.13. Caracterização da Estrutura de Elementos do Genoma da Cana-de-açúcar Homólogos à Superfamília <i>hAT</i>	66
III.13.1. Ensaio de PCR-Inversa sobre o DNA Genômico da Cana-de-Açúcar a partir do Clone de cDNA TE221 (Grupo II).....	67
III.13.1.1. Hidrólise do DNA Genômico com Endonuclease de Restrição.....	67
III.13.1.2. Precipitação do Produto da Hidrólise.....	67
III.13.1.3. Reação de Ligação (Circularização).....	68
III.13.1.4. PCR-Inversa.....	68
III.13.1.4.1. Iniciadores.....	68
III.13.1.4.2. Reações de PCR-Inversa.....	69

III.13.1.5. Purificação das bandas de interesse da PCR-Inversa.....	70
III.13.1.6. Subclonagem do Produto de PCR-Inversa.....	70
III.13.1.7. Transformação de Bactérias Competentes pelo Método do Choque Térmico.....	71
III.13.1.8. Minipreparação de DNA plasmidial.....	71
III.13.1.9. Sequenciamento do Produto de PCR-Inversa.....	71
III.13.1.10. Montagem das Sequências Consenso dos Clones.....	71
III.13.1.11. Análise das Sequências Obtidas a Partir de PCR-Inversa.....	72
III.13.1.12. Validação do “TE221-RP”	72
III.13.2. Ensaio de PCR sobre o DNA Genômico de Cana-de-açúcar com Iniciadores Desenhados Sobre as TIRs Relativas ao TE221.....	73
III.13.2.1. Amplificações por PCR, Subclonagem e Sequenciamento.....	73
III.13.2.2. Hibridação molecular de SChAT1 e SChAT2 com Sonda do Grupo II.....	74
IV. RESULTADOS.....	76
IV.1. Obtenção das Sequências Completas dos Clones de cDNA e Identificação das Famílias de Elementos Transponíveis Expressas em Cana-de-açúcar.....	76
IV.2. A superfamília <i>hAT</i>	82
IV.2.1. A superfamília <i>hAT</i> é Constituída por Dois Grupos de Sequências.....	86
IV.2.2. A perfil de expressão da superfamília <i>hAT</i> em membrana de macroarray.....	91
IV.2.3. Padrão de Hibridação de Sondas dos Grupos I e II em Genomas de Gramíneas.....	92
IV.2.4. Estudo dos homólogos a <i>hAT</i> do Grupo I.....	93

IV.2.4.1. Os cDNAs do Grupo I de Cana-de-açúcar e Dois Homólogos Encontrados em Arroz Apresentam Alta Similaridade de Sequência com Transposases Domesticadas.....	96
IV.2.4.2. Os Elementos do Grupo I estão Representados com Baixo Número de Cópias nos Genomas das Gramíneas.....	100
IV.2.4.3. O Grupo I Apresenta Diferentes Padrões de Expressão em Tecidos de Cana-de-açúcar.....	102
IV.2.5. Estudo dos Homólogos a <i>hAT</i> do Grupo II.....	103
IV.2.5.1. Análise da distribuição de dois elementos do grupo II nos genomas de mono e eudicotiledôneas.....	105
IV.2.5.2. Estruturas de Elementos Homólogos à Superfamília <i>hAT</i> Amplificadas do Genoma da Cana-de-Açúcar.....	107
IV.2.5.2.1. Identificação de Potenciais TIRs para os Elementos <i>hAT</i> do Grupo II pela Técnica de PCR-Inversa.....	107
IV.2.5.2.2. Obtenção das regiões Genômicas 5' e 3' não transcritas do TE221 - reconstrução parcial de um elemento <i>hAT</i> -like.....	113
IV.2.5.2.3. Amplificação genômica de elemento SChAT do Grupo II a partir de suas TIRs.....	120
V. DISCUSSÃO.....	124
V.1. Obtenção das Sequências Completas dos Clones de cDNA e Identificação das Famílias de Elementos Transponíveis Expressas em Cana-de-Açúcar.....	124
V.2. Análise da Diversidade Entre os cDNAs Homólogos à Superfamília <i>hAT</i>	126
V.3. Identificação de Potenciais TIRs e consequente isolamento de uma cópia de SChAT a partir do genoma de cana de açúcar.....	133
V.4. Conclusões.....	136
VI. BIBLIOGRAFIA.....	137
VII. APÊNDICES.....	144

RESUMO

Os elementos de transposição (TEs) são sequências genéticas móveis. Sua capacidade mutagênica faz deles uma importante fonte de variabilidade nos genomas. Outro importante papel dos TEs na evolução dos genomas é o de doadores de domínios protéicos na formação de novos genes. 276 clones de cDNA homólogos a TEs foram previamente identificados no banco de dados do SUCEST (projeto de sequenciamento de etiquetas expressas de cana-de-açúcar). Neste trabalho nós realizamos o sequenciamento completo de 156 destes clones e a classificação e caracterização de suas sequências comparando-as com bancos de dados. Foram identificadas 9 diferentes famílias de transposons e 11 diferentes famílias de retrotransposons. As famílias mais representadas entre os transposons foram MuDr e *hAT* (que engloba os elementos do tipo Ac e Tam3), para os quais foram identificados 43 e 32 clones de cDNA, respectivamente. Entre os retrotransposons, a família mais representada foi *Hopscotch*, apresentando 25 clones de cDNA. Após esta análise global, o foco das investigações voltou-se para os cDNAs do tipo *hAT*. Uma análise comparativa destes cDNAs revelou que as sequências homólogas a *hAT* estão distribuídas em dois grupos. O grupo I, é composto por sequências com alta conservação no nível de nucleotídeos, está presente no genoma de todas as gramíneas analisadas (híbridos e parentais da cana-de-açúcar, milho e arroz) com um baixo número de cópias, teve a sua expressão detectada em folhas, raízes e mais intensamente em calos cana. Além disso, apresenta alta similaridade de sequências com transposases domesticadas descritas na literatura. O grupo II, por sua vez, é composto por sequências mais heterogêneas, que apresentam similaridade com os transposons originais que constituem a superfamília *hAT*: *hobo* (de *Drosophila melanogaster*), Ac (de *Zea mays*) e Tam3 (de *Antirrhinum majus*). Sua distribuição é restrita ao genoma de *Saccharum*, com um número de cópias maior que o grupo I. Um ensaio de PCR-Inversa identificou terminações inversas repetidas (TIRs) para o cDNA TE221 do grupo II. A partir de iniciadores desenhados sobre estas TIRs foi possível recuperar dois elementos, de 3,5kb e 4,2kb, respectivamente, e um MITE de 250 pb, todos homólogos a *hAT*. Este resultado demonstrou que a estratégia utilizada para recuperar elementos do genoma da cana-de-açúcar a partir do cDNA TE221 mostrou-se eficiente. Homólogos aos grupos I e II de cana-de-açúcar foram identificados em bancos de dados de milho, arroz e *Arabidopsis*. Estes dados sugerem que a separação dos dois grupos ocorreu antes da divergência entre as classes Monocotyledonea e Eudicotyledonea. Com base nos resultados aqui apresentados sugerimos que um transposon ancestral do tipo *hAT*, presente nas angiospermas anteriormente à separação de Monocotyledonea e Eudicotyledonea, teve sua transposase capturada na formação de um gene com função celular. A partir do evento da domesticação, estas transposases seguiram dois caminhos evolutivos distintos, um como gene funcional e outro como um transposon tradicional. Estas duas formas de transposase do tipo *hAT* podem ser encontradas no genoma da cana-de-açúcar, representadas pelos elementos dos grupos I e II, respectivamente.

ABSTRACT

Transposable elements (TEs) are mobile genetic sequences. Their mutagenic capacity makes them important sources of variation in the genomes. These elements have another important evolutionary role as donors of functional protein domains in the formation of new genes. 276 cDNA clones homologous to TEs were previously identified in the Brazilian Sugarcane Expressed Sequence Tag Project (SUCEST) databases. In this work, we have obtained the full sequences of 156 for these clones. These sequences were compared with Genbank database. We have identified 9 families of transposons and 11 families of retrotransposons. The most representative families found amongst the transposons were MuDr and *hAT* (which encompass Ac and Tam3), with 43 and 32 cDNAs, respectively. Amongst the retrotransposons, the most representative family was Hopscotch, with 25 cDNAs. After this global analysis, we have focused our investigation in the *hAT*-like cDNAs. A comparative analysis of these cDNAs has revealed a profile of two distinct groups. Group I is composed of sequences with high conservation at nucleotide level, it is present in the genome of all grasses analysed (hybrids and parentals of sugarcane, maize and rice) with low copy number, it is expressed in leaves and roots of sugarcane, and more intensely in callus. In addition, group I sequences have clustered with domesticated transposases. The group II is composed of more heterogeneous sequences similar with the original elements that constitute the *hAT* superfamily: hobo (from *Drosophilla melanogaster*), Ac (from *Zea mays*) and Tam3 (from *Antirrhinum majus*). This group was shown to be restricted to the genome of *Saccharum*, with higher copy number than group one. Inverse-PCR assays have identified terminal inverted repeats (TIRs) to the cDNA TE221 from group II. Primers based on the sequences of the TIRs allowed us to recover three elements *hAT*-like from sugarcane's genomic DNA: one of 3,5kb and another of 4,2kb, and a MITE of 250 bp. These results corroborate the strategy applied in order to recover elements from the sugarcane's genome. Sequences homologous to both sugarcane group I and group II were found also in maize and rice, as well as in arabidopsis databases. These data suggest that the divergence of the two groups occurred before the separation between the classes Monocotyledonea and Eudicotyledonea. Based on our results, we suggest the existence of an ancestral transposon *hAT*-like, present in angiosperms before the separation between Monocotyledonea and Eudicotyledonea, of which the transposase was captured to compose a new gene with some cellular function. Since the domestication event, these transposases followed distinct evolutive pathways, one as a regular gene and another as a *bona fide* transposon. These two forms of *hAT*-like transposases could be found in the sugarcane's genome, represented by the elements from groups I and II, respectively.

I. INTRODUÇÃO

I.1. Elementos de Transposição – O Genoma Dinâmico

Os elementos de transposição (TEs) são sequências genéticas móveis presentes no genoma dos seres vivos, amplamente distribuídos entre bactérias, algas, fungos, plantas e animais (Bennetzen, 2000). Estes elementos são abundantes nos genomas, e podem corresponder a mais de 70% do DNA nuclear em algumas gramíneas (SanMiguel et al., 1998). A atividade dos TEs e sua inerente habilidade de mudar de lugar no cromossomo foram descritas pela primeira vez em plantas, pela geneticista Barbara McClintock no final da década de 40. Ela verificou em seus estudos o envolvimento dos TEs na origem de mutações espontâneas e instáveis em milho, e denominou-os inicialmente de “elementos controladores dos genes” (McClintock, 1956). A atividade mutacional dos TEs pode gerar alterações estruturais e funcionais nos genomas, como a interrupção de um gene por inserção dentro do mesmo, ou a modificação do nível ou padrão de expressão gênica, quando a inserção ocorre em regiões não codificantes (Bennetzen, 1996). Em decorrência destas características os TEs podem ser considerados uma fonte de variabilidade genética, e consequentemente podem influenciar no funcionamento e evolução dos genomas (McClintock, 1984).

É comum que as mutações geradas pela inserção de TEs dentro de genes sejam instáveis, e ocorra a reversão do fenótipo durante o desenvolvimento de partes da planta ou grãos. A reversão ocorre devido à excisão do elemento e geralmente permite a recuperação da função do gene e consequentemente, o fenótipo original. A observação de fenótipo variegado pode estar correlacionada com a atividade de TEs, em particular quando observa-se a excisão de um dado TE de um alelo mutante em uma determinada fase do desenvolvimento, gerando setores com um padrão de expressão do genótipo selvagem em um fundo com genótipo mutante (Feschotte *et al*, 2002). O tamanho e forma do setor contendo as células revertentes é determinado pelo momento do desenvolvimento em que ocorre a excisão do elemento, e pelo padrão de divisão celular específico do tecido em que ocorreu a reversão (Fedoroff, 1989). O mecanismo de formação do

fenótipo variegado está esquematizado na figura 1, juntamente com exemplos encontrados na natureza.

Figura 1 – Mecanismo da formação de fenótipo variegado e exemplos. **A:** A inserção de um TE em um gene da rota de síntese de pigmento leva ao fenótipo mutante sem pigmento. A excisão do TE do alelo mutado, pela ação de uma transposase funcional, leva à reversão para o fenótipo selvagem, com pigmentação. A excisão precoce no desenvolvimento do grão leva à formação de um setor grande de fenótipo revertente, enquanto a reversão tardia leva à formação de pontos pequenos. A ocorrência da excisão em células germinais leva à formação de um grão revertente por completo (adaptado de Feschotte *et al*, 2002). **B:** Exemplos de fenótipo variegado, da esquerda para a direita: grãos de *Zea mays*, flor de *Ipomea sp.* e inflorescência de *Daucus carota*.

Algumas características dos elementos de transposição, como a sua persistência dentro dos genomas, ausência de função aparente, geração de mutações deletérias, assim como a utilização da maquinaria celular de replicação para a manutenção de suas cópias levaram os pesquisadores Orgel e Crick (1980) a propor o conceito de “DNA egoísta”. Durante muito tempo estes elementos foram considerados parasitas moleculares, não tendo nenhuma relação com seu genoma hospedeiro a não ser a de manter ou aumentar seu número de cópias, a despeito de prejuízos causados ao mesmo. Porém, um novo paradigma tem se consolidado, sugerindo que os TEs, enquanto fonte genômica de mutações, devem ter um papel fundamental na evolução dos genes e na estrutura genômica dos eucariontes (Lönnig & Saedler, 1997, Kalendar *et al*, 2000 e Cordaux *et al*, 2006). Foi observada a ativação dos TEs perante situações de estresse de origem tanto biótica quanto abiótica, como infecção por patógenos ou estresse hídrico. Se esta ativação dos TEs implica no aumento das taxas de mutação no genoma hospedeiro, isto é um indício de que podem atuar no processo de adaptação de seu organismo hospedeiro frente a pressões evolutivas (Grandbastien, 2004 e Kalendar *et al*, 2000). Nesta condição, os elementos poderiam gerar alterações no genoma que resultassem em indivíduos capazes de sobreviver a pressões seletivas que estivessem a ponto de extinguir toda a população (McClintock, 1984).

O paradoxo do “C-value” pode em parte ser explicado pela amplificação dos TEs, onde observa-se que o conteúdo de DNA de um organismo não reflete sua complexidade (Thomas, 1971). Um exemplo em particular são os retrotransposons com LTRs, que são os mais abundantes nas gramíneas. No caso da diferença no tamanho do genoma do milho e do sorgo, sugere-se que o grande acúmulo de retrotransposons no milho seja o responsável por esta diferença. Através da comparação da região do gene *adh1* (desidrogenase do álcool) do milho e do sorgo, demonstrou-se que o sorgo não tem retrotransposons nesta região, o que sugere que eles se inseriram após a divergência destas duas espécies (SanMiguel *et al.*, 1998). No caso de duas espécies de *Oryza*, *O. sativa* e *O. australiensis*, a variação no número de cópias do elemento RIRE1 pode explicar a diferença de 1/3 no tamanho de seus genomas (Vicient e Schulman, 2002).

Como já mencionado anteriormente, a atividade transposicional dos TEs pode trazer efeitos mutagênicos para o genoma hospedeiro. Pode-se supor, portanto, que um genoma ancorando elementos com altas taxas de transposição acumularia tantas mutações com o decorrer do tempo a ponto de tornar-se inviável. Estes danos poderiam ser atenuados com a diminuição das taxas de transposição, e de fato aparentemente há uma seleção positiva de elementos com baixas taxas de transposição, porém não nulas (Brookfield, 2005). Um exemplo é o retrotransposon L1 de mamíferos. Ele apresenta em sua forma selvagem uma inserção prematura de poliadeninas em sua sequência, o que causa uma interrupção transcricional e diminui suas taxas de transcrição, que ficam muito abaixo das dos genes constitutivos (Perepelitsa-Belancio & Deininger, 2003). A remoção destas sequências que aparentemente interferem na transcrição, sem alterar a sequência de aminoácidos, de fato gerou um elemento com uma taxa de transposição 200 vezes maior que a do elemento selvagem (Han & Boeke, 2004 e Bushman, 2004). As baixas taxas de transposição do elemento L1 selvagem podem ser interpretadas como resultado de uma seleção positiva para esta característica (Brookfield, 2005).

A regulação da atividade dos TEs pode ocorrer no nível transcricional, pós-transcricional ou pós-traducional. A regulação transcricional pode ocorrer através da metilação das sequências dos TEs presentes no genoma. Van Sluys e colaboradores (1993) analisaram os níveis de metilação do elemento Ac introduzido em *Arabidopsis thaliana* na forma de um T-DNA (Ac-T-DNA), através de transformação com *Agrobacterium*. Enquanto em raízes derivadas de calos, provenientes diretamente da transformação, foi detectada a atividade transposicional de Ac, após retrocruzamentos o elemento apresentou-se inativo nas plantas regenerantes. Usando endonucleases de restrição isoesquizômeros sensíveis e não sensíveis à metilação, os autores detectaram um aumento nos níveis de metilação do Ac-T-DNA, e propuseram que este aumento na metilação poderia estar relacionado com a inativação do elemento. Subsequentemente, Scortecci e colaboradores (1997) trataram as plantas transgênicas, em que o transposon Ac estava inativo, com 5-azacitidina, um agente desmetilante. Foi observado que aproximadamente 13% das plantas tratadas apresentaram eventos de transposição do elemento Ac contido no T-DNA, enquanto as plantas não tratadas não apresentaram evidências de

transposição. Estes resultados confirmaram a relação entre metilação e inativação da capacidade transposicional do elemento Ac, através da inibição de sua transcrição.

A regulação pós-transcricional pode se dar através do silenciamento por interferência de RNA. Um exemplo deste tipo de regulação é o locus relativo ao elemento Mutator, descrito por Slotkin e colaboradores (2005). Plantas de milho contendo o elemento MuDr transposicionalmente ativo, quando cruzadas com plantas contendo este locus, denominado “Mu killer” (Muk), geravam uma progênie que apresentava o elemento MuDr silenciado. Observou-se que o locus Muk apresenta a duplicação invertida de um elemento MuDr parcialmente deletado. Esta estrutura, após transcrita, forma um grampo de dupla fita de RNA que é processado por uma enzima do tipo Dicer em pequenos RNAs de interferência (siRNAs). Estes dão início a um ciclo de reconhecimento e processamento dos transcritos do elemento MuDr, gerando siRNAs de 25 nucleotídeos e de 21-22 nucleotídeos. Os autores sugerem que a presença destes siRNAs leva ao silenciamento do transposon MuDr tanto pela degradação de seus transcritos quanto pela ativação de uma rota de silenciamento gênico por metilação de DNA dependente de RNA. Muk foi o primeiro exemplo da ocorrência natural de uma estrutura derivada de um transposon capaz de iniciar o silenciamento de uma família ativa de transposons.

A regulação pós traducional está relacionada com as concentrações da transposase. Lampe e colaboradores (1999), estudaram os parâmetros que afetam as taxas de transposição do elemento Himar1, isolado da mosca *Haematobia irritans*. Em ensaio de atividade do elemento Himar1 *in vitro* demonstraram que concentrações elevadas da transposase de Himar1 diminuem a frequência de transposição do elemento.

Atualmente, em função do elevado número de elementos de transposição identificados nos mais diversos organismos a partir dos projetos de sequenciamento de genomas, nota-se que o papel de genes egoístas destas sequências móveis está sendo reconsiderado. Uma função atribuída aos elementos de transposição, bastante relevante do ponto de vista da evolução dos genomas, é a de doadores de domínios protéicos no processo de formação de novos genes. Neste caso segmentos dos TEs que codificam para

domínios protéicos de interesse, como domínios de ligação ao DNA, são incorporados na formação de novos genes que codificam para novas proteínas.

Resultados publicados por Bundock and Hooykaas (2005) sobre o mutante DAYSLEEPER descreveram pela primeira vez em plantas uma transposase domesticada derivada de um transposon do tipo *hAT*. DAYSLEEPER tem função essencial para o crescimento e desenvolvimento de *arabidopsis*. Os autores demonstraram que a proteína codificada por este gene atua como um fator de transcrição, ligando-se a regiões promotoras de diversos genes, e que a interrupção do gene gera má formação de plântulas transgênicas, que têm o seu desenvolvimento comprometido. Com base na sequência de DAYSLEEPER, outras transposases domesticadas do tipo *hAT*, denominadas w-gary e b-gary, foram descritas em trigo e cevada, respectivamente (Muehlbauer *et al*, 2006). Um outro exemplo é a transposase domesticada do tipo MuDr descrita por Cowan e colaboradores (2005), denominada Mustang. Ela foi identificada nos genomas de *arabidopsis* e arroz através de busca eletrônica.

1.2. Diversidade dos Elementos de Transposição

Os elementos de transposição (TEs) apresentam uma grande diversidade em sua estrutura, fato que se reflete no grande número de famílias de elementos já descrito. Estas famílias apresentam características peculiares como o tipo e a ordem de seus domínios protéicos internos, o que caracteriza as enzimas necessárias para a sua transposição. São também características das diversas famílias de TEs as terminações repetidas que se encontram em suas extremidades. Estas terminações, que podem ser reverso-complementares ou diretas, contém as regiões promotoras completas ou parciais do elemento, e são reconhecidas no momento de sua transposição. Os TEs são divididos em duas grandes classes de acordo com seu modo de transposição: os de classe I ou retroelementos transpõem-se através da transcrição reversa de um cDNA a partir de um RNAm intermediário, e os de classe II, os transposons clássicos, movem-se diretamente a partir de uma molécula de DNA (Smyth, 1993).

Classe I: retroelementos com LTRs, LINEs e SINEs

O grupo dos retroelementos compõem-se pelos retrovírus, pelos retrotransposons com LTRs (longas terminações repetidas) e pelos retrotransposons sem LTRs, cujas estruturas estão esquematizadas na figura 2. Os retrotransposons com LTRs dividem-se em dois grupos, Ty3/gypsy e Ty1/copia, distinguíveis pela organização interna de seus domínios protéicos da transcriptase reversa, da RNase H e da Integrase. Os retrotransposons com LTRs codificam para várias proteínas necessárias para sua transposição, através de um único RNA mensageiro. As LTRs são diretamente repetidas, e não codificam nenhuma proteína, mas contém as regiões promotoras e terminadoras associadas com a transcrição do mensageiro (Smyth, 1993). Já os retrotransposons sem LTRs subdividem-se em LINEs ("long interspersed nuclear elements") e SINEs ("short interspersed nuclear elements") (Xiong e Eickbush, 1990). Os retrotransposons sem LTRs geralmente possuem em suas extremidades regiões não traduzidas e uma cauda poli-A. Os LINEs codificam para as proteínas do nucleocapsídeo, uma endonuclease e uma transcriptase reversa, necessárias para sua transposição. Os SINEs são sequências pequenas, de 100 a 300 pb, que não codificam nenhuma das proteínas com funções necessárias para sua transposição.

Figura 2 – Diversidade estrutural dos retroelementos. **A:** Grupo dos retrotransposons com LTRs e retrovírus. Apresentam as “Longas Terminações Repetidas” (LTRs) em suas extremidades, que são reconhecidas no momento da transposição e possuem suas regiões promotoras. Os retrotransposons dos tipos Ty1/copia e Ty3/gypsy diferenciam-se pela ordem interna de seus domínios protéicos. Os retrovírus diferenciam-se dos retrotransposons por apresentarem o domínio ENV que codifica para a proteína de envelope. CAP: capsídeo, PROT: protease, INT: integrase, RT: transcriptase reversa, RH: RNase H. **B:** Retrotransposons sem LTRs. Apresentam uma cauda poli-A em sua extremidade 3'. Enquanto os LINES codificam as proteínas necessárias para sua transposição, os SINES não tem esta capacidade.

Todos os SINES descritos até o momento são caracterizados por apresentarem um promotor interno da RNA polimerase III, o que lhes confere uma habilidade de serem eficientemente replicados (Deininger, 1989; Kumar e Bennetzen, 1999; Bennetzen, 2000). A principal diferença entre os retrotransposons e os retrovírus é que estes últimos possuem o domínio *env*, que codifica para a proteína de envelope que lhes confere caráter infeccioso

(Smyth, 1993). Recentemente verificou-se a presença de um domínio *env* em retrotransposons da família Ty3/gypsy em plantas (Peterson-Burch *et al*, 2002; Wright *et al*, 2002).

Classe II: elementos autônomos, não autônomos e MITEs

Os transposons podem apresentar um ou mais quadros abertos de leitura (ORF; do inglês “open reading frame”) e codificam para uma transposase, enzima responsável pela transposição do elemento. Em suas extremidades localizam-se as TIRs (terminações inversamente repetidas), que são reverso-complementares e contém regiões promotoras. A transposase reconhece sequências específicas próximas e dentro das TIRs quando realiza a mobilização do elemento (Lewin, 2004). Os transposons são encontrados nos genomas como elementos autônomos e não autônomos. Enquanto os elementos autônomos são capazes de codificar uma transposase funcional, os não autônomos dependem da ação em “trans” de uma transposase funcional codificada por elemento autônomo da mesma família para realizarem sua transposição (Fedoroff, 1989). Os MITEs (miniature inverted-repeat transposable elements) possuem TIRs em suas extremidades e não têm a capacidade de codificar qualquer transposase. Suas sequências são curtas, entre 100 e 600 pb. A existência de transposons que potencialmente podem codificar para transposases funcionais e que compartilham similaridade de sequência com alguns MITEs levou à hipótese de que os MITEs podem ser sequências remanescentes, derivadas de deleções de transposons, que são mobilizadas em “trans” pelas transposases codificadas pelos mesmos (Feschotte *et al*, 2002 e Casacuberta & Santiago, 2003). As famílias de transposons são, portanto, compostas por elementos que compartilham as mesmas TIRs, englobando os elementos autônomos e não autônomos, e também os MITEs. A estrutura de uma família de transposons está esquematizada na figura 3.

Figura 3 – Estrutura de uma família de transposons, mostrando as diferentes formas em que podem ser encontrados. O elemento autônomo possui uma ORF íntegra que codifica para uma transposase funcional, enquanto os elementos não autônomos apresentam uma ORF defectiva e dependem da transposase codificada pelo elemento autônomo para sua transposição. Em suas extremidades encontram-se as TIRs (terminações inversas repetidas), que contém as regiões promotoras do elemento. A transposição depende do reconhecimento das TIRs pela transposase.

I.3. Superfamília *hAT* de transposons

A superfamília *hAT* de transposons tem como ícones dois dos transposons melhor caracterizados em plantas, *Ac* do milho (*Zea mays*) e *Tam3* de boca-de-leão (*Antirrhinum majus*), e também o transposon *hobo* descrito em *Drosophila melanogaster* (Kempken *et al*, 2001). Suas principais características são a presença de três domínios protéicos conservados localizados na porção carboxi-terminal de suas transposases, TIRs pequenas de 11 pb e duplicação do sítio de inserção de 8 pb (Kempken *et al*, 2001 e Rubin *et al*, 2001). Um dos domínios característicos das transposases do tipo *hAT* localiza-se entre os resíduos 682 a 720 da transposase de *Ac*, e foi demonstrado estar envolvido na função de dimerização (Essers *et al*, 2000). Outro domínio conservado, utilizado em análises no presente trabalho, localiza-

se entre os resíduos 555 a 594 da transposase de Ac, próximo ao domínio de dimerização. Sua função, no entanto, continua desconhecida. Outros transposons tem sido descritos dentro desta superfamília em diversos organismos, como um elemento do tipo *hAT* em *Silene latifolia*, da classe Eudicotiledonea (Pritham *et al*, 2003), o elemento *Crypt1* do fungo *Cryptonectria* parasítica (Linder-Basso *et al*, 2001) e DodoPi do fungo fitopatígeno *Phytophthora infestans* (Fong & Judelson, 2004), o elemento *Hermes* de *Musca domestica* (Warren *et al*, 1994), a dupla de elementos autônomo e não autônomo Dart e nDart de *Oryza sativa* (Fugino *et al*, 2005), e até mesmo elementos de *Zea mays* que não interagem com Ac, como a dupla Mx e rMx (Xu & Dooner, 2005).

O transposon Ac tem grande importância histórica, pois foi o primeiro elemento genético móvel descrito. Foi descoberto por Bárbara McClintock durante estudos de quebras cromossômicas em milho, em meados da década de 1940. A dupla *Ac/Ds* (Activator/Dissociation), elementos autônomo e não autônomo, respectivamente, são os transposons mais estudados até o momento e correspondem aos “elementos controladores dos genes” descritos em milho por Bárbara McClintock (1956). Ela observou primeiramente o elemento Ds, que estava ligado a quebras cromossômicas. Esta quebra, porém, não era espontânea e necessitava da presença de um segundo elemento genético para ocorrer, denominado Ac. Ambos os elementos apresentavam propriedades incomuns, como a capacidade de mudar de locus, e foi observando este fenômeno que McClintock reportou formalmente que o locus Ds poderia se transpor (McClintock, 1948). Estudos subsequentes levaram à identificação de uma linhagem de milho, contendo o elemento Ds, que apresentava uma mutação no locus *c* do braço curto do cromossomo 9. McClintock sugeriu que o elemento Ds havia se transposto para o locus *c* da linhagem mutante. A nova mutação era estável na ausência do elemento Ac, mas podia ocorrer reversão para o fenótipo selvagem quando Ac era reintroduzido por cruzamento. A reversão podia ocorrer tanto em células somáticas quanto em células germinativas. Os revertentes germinais da mutação do gene *c* não apresentavam mais quebras cromossômicas neste locus, devido à excisão de Ds. Estas observações levaram McClintock a

entender a relação entre a dinâmica de transposição Ac/Ds e a instabilidade das mutações. Ela observou também que o elemento Ac, da mesma forma que Ds, podia gerar mutações por inserção. A diferença é que estas mutações podiam reverter espontaneamente, enquanto as mutações geradas pela inserção de Ds somente revertiam na presença de Ac. (Fedoroff, 1989). A magnitude das descobertas de McClintock sobre a capacidade de se mover, de gerar quebras e rearranjos cromossômicos e a influência dos “elementos controladores” sobre os genes, somente foi entendida nas décadas seguintes, quando os avanços tecnológicos permitiram a identificação de elementos de transposição em outras espécies e a confirmação de suas hipóteses. A profundidade de suas observações e a visão à frente de seu tempo, que teve ao interpretá-las, foi reconhecida em 1984 quando McClintock recebeu o Prêmio Nobel de Medicina.

Os elementos Ac e Ds foram isolados molecularmente por Fedoroff e colaboradores (1983). Ambos estavam inseridos no locus *waxy* (*wx*) do milho, que codifica uma importante enzima da rota biossintética do amido em tecidos do pólen e endosperma. Um ano depois Pohlman e colaboradores (1984) determinaram a sequência nucleotídica do elemento Ac identificando o gene que codifica a transposase e as terminações inversas repetidas (TIRs) imperfeitas de 11 pb. Um único transcrito foi identificado para o elemento Ac, com 3,5 kb de tamanho, compreendendo 5 exons e codificando uma ORF de 2,421 pb relativa à transposase (Kunze *et al*, 1987). Becker e Kunze (1997) demonstraram que a transposase do elemento Ac apresenta um domínio de ligação ao DNA que reconhece tanto as TIRs quanto pequenos motivos de trinucleotídeos, repetidos nos 250 pb subterminais de cada extremidade do elemento. A primeira demonstração de que o elemento Ac poderia se mobilizar em um sistema heterólogo foi realizada com protoplastos de *Nicotiana tabacum* (Baker *et al*, 1986). Em seguida, Van Sluys e colaboradores (1987) demonstraram a transposição de Ac em raízes ou raízes derivadas de calos de plantas transgênicas de *Arabidopsis thaliana* e *Daucus carota*.

O elemento *Tam3* foi identificado em *Antirrhinum majus* através da análise de mutantes para dois genes envolvidos na rota de síntese do pigmento antocianina: *nivea* (*niv*) e *pallida* (*pal*). Em ambos os locus foi identificada uma inserção de 3,5 kb em suas regiões promotoras, relativa ao elemento. Três

diferentes cópias do elemento Tam3 isoladas do locus *niv* apresentaram terminações inversas repetidas (TIRs) de 11 pb, e sítio duplicado de inserção de 8 pb, como o elemento Ac. No entanto, o elemento isolado do locus *pal* apresenta sítio duplicado de 5 pb. A observação do aparecimento de alelos estáveis na progênie de linhagens que apresentavam alelos instáveis dos genes *niv* e *pal* levou à dedução de que o mecanismo de transposição de Tam3 é de excisão/inserção, como a do elemento Ac (Coen *et al*, 1989). A transposição de Tam3 é fortemente afetada pela temperatura. O mesmo é ativo em baixas temperaturas, em torno de 15°C, e inativo em temperaturas mais altas, em torno de 25°C (Harrison & Fincham, 1964). Baseando-se em estudos de padrões de cores que aparecem nos fenótipos variegados, Carpenter e colaboradores (1987) reportaram que as taxas de excisão de Tam3 podem aumentar até 1000 vezes sob a temperatura de 15°C, comparando-se com as taxas de excisão a 25°C. Hashida e colaboradores (2003) investigaram fatores que poderiam estar envolvidos com a inativação de Tam3 em plantas de *Antirrhinum majus* crescidas em diferentes temperaturas. Seus resultados demonstraram que a inativação do elemento à temperatura de 25°C não está relacionada à transcrição da transposase ou à sua atividade enzimática, mas ao estado de metilação do elemento no genoma. Altas temperaturas levaram à hipermetilação do elemento, enquanto baixas temperaturas resultaram na redução da metilação. Frente à observação de uma forte correlação entre temperatura e metilação, os autores sugeriram que esta última deve ser pelo menos um dos fatores que influenciam a transposição dependente de baixa temperatura, peculiar ao elemento Tam3.

O elemento *hobo* foi clonado e caracterizado molecularmente por McGinnis e colaboradores (1983) em *Drosophila melanogaster*. Duas linhagens de *Drosophila* podem ser definidas com relação ao elemento *hobo*, as que contém o elemento em suas formas autônoma e não autônoma, denominadas linhagens H, e linhagens E, que não apresentam nenhuma forma do elemento. O elemento *hobo* autônomo tem 3016 pb, com TIRs de 12 pb e sítio duplicado de inserção de 8 pb. É ubíquo nas linhagens H e apresenta um baixo número de cópias, de 2 a 10 por genoma. Os elementos defectivos resultam de deleções internas do elemento autônomo (Streck *et al*, 1986). Sabe-se que *hobo* é responsável por um dos sistemas de disgênese híbrida em *Drosophila*,

porém, ainda não foi estabelecida uma correlação entre a esterilidade gerada pela disgnese gonadal e a dinâmica de transposição de hobo ou suas características moleculares (Bazin et al, 1999).

I.4. A cana-de-açúcar (*Saccharum* spp.)

A cana-de-açúcar é uma planta nativa das regiões tropicais. A teoria mais aceita sobre sua origem considera que é nativa das ilhas do arquipélago da Polinésia, e que a espécie *Saccharum robustum* teria se originado no centro de expansão da Nova Guiné. Deste centro esta cultura teria se dispersado, ainda no período Neolítico, para Ilhas Fiji e Taiti a leste, e depois a oeste para a Malásia, Índia, Filipinas e China. Nas Américas a cana foi introduzida no final do século XV, através da segunda expedição de Cristóvão Colombo. No Brasil, Martim Afonso de Souza introduziu oficialmente mudas de cana-de-açúcar provenientes da Ilha da Madeira, em 1502. Desde sua descoberta pela humanidade a cana-de-açúcar começou a ser utilizada na alimentação, como fonte adoçante. Posteriormente verificou-se o seu grande valor como reserva energética, e ela passou a ser também utilizada com este fim em viagens marítimas e de migração (Cesnik e Miocque, 2004). Atualmente a cana-de-açúcar é uma das mais importantes culturas das regiões tropical e subtropical do globo, sendo que o Brasil é um dos maiores produtores mundiais. Suas principais utilizações são para a produção de açúcar e como fonte de energia, na forma de etanol e eletricidade (fonte: www.fapesp.br). A crescente conscientização sobre a necessidade de se desenvolver e estimular o uso de fontes de energia renováveis reafirma a importância da utilização da cana-de-açúcar na produção de etanol combustível.

As espécies do gênero *Saccharum* pertencem à família *poacea*, dentro da classe das monocotiledôneas. Os cultivares comerciais atuais de cana-de-açúcar, *Saccharum* spp. ($2n=100-130$), são híbridos provenientes de cruzamentos interespecíficos entre as espécies *Saccharum officinarum* ($2n=80$) e *Saccharum spontaneum* ($2n=40-128$), seguidos de novos cruzamentos com *Saccharum officinarum*. A contribuição de *Saccharum spontaneum* para o número de cromossomos do híbrido é de 15 a 25% (D'Hont, 1996). *Saccharum officinarum* é uma espécie pertencente às ditas “canas nobres”, que

apresentam alto teor de açúcar e baixa porcentagem de fibra, em colmos grossos. É bastante suscetível a doenças. Já *Saccharum spontaneum* pertence às “canas selvagens”, com colmos curtos e finos e alto teor de fibra, mas apresenta uma boa resistência vegetativa em campo, mesmo sob condições adversas (Cesnik e Miocque, 2004). Dentro das características dos híbridos, o alto teor de açúcar é atribuído a *S. officinarum*, enquanto considera-se como contribuição de *S. spontaneum* o vigor vegetativo e a resistência a estresses bióticos e abióticos (Ming *et al*, 1998). A cana-de-açúcar é propagada comercialmente por via assexuada, através da germinação de suas gemas. Um segmento do colmo contendo de 2 a 4 gemas, denominado tolete, é utilizado em sua multiplicação vegetativa. O processo sexuado é utilizado em programas de melhoramento, através da germinação de sementes, visando obter variabilidade genética (Cesnik e Miocque, 2004).

O genoma dos híbridos é alopoliplóide, ou seja, resultante do cruzamento de duas espécies distintas constituídas de dois genomas distintos (Grivet *et al*, 1996). Esta situação de choque genômico nos alopoliplóides pode levar à ativação de elementos de transposição quiescentes (Chen e Ni, 2006). Foi verificada a reativação tanto de transposons como de retrotransposons em alotetraplóides de *arabidopsis* (Comai *et al*, 2000 e Madlung *et al*, 2005), assim como a ativação de retrotransposons em alotetraplóides de trigo (Kashkush *et al*, 2003). Madlung e colaboradores (2005) detectaram, através de microarranjo de DNA, a reativação transcricional de um retrotransposon do tipo Ty1/copia e de dois transposons do tipo Em/Spm pertencentes à subfamília Suf, em alotetraplóides sintéticos de *arabidopsis*. Esta reativação foi correlacionada com a redução da metilação nas sequências destes TEs (Madlung *et al*, 2002). Enquanto o elemento Suf é metilado nos parentais tetraplóides, aparece desmetilado e reativado transcricionalmente nos alotetraplóides. Mudanças na metilação do DNA foram detectadas também em decorrência da recente hibridização e poliploidia em *Spartina* (*Poacea*), associadas com mudanças na expressão gênica e variações fenotípicas (Salmon *et al*, 2005). Estes dados sugerem que a alopoliploidização pode provocar uma perturbação da estrutura genômica e o remodelamento da cromatina, gerando tanto a reativação como o silenciamento de TEs e de genes codificantes de proteínas (Chen e Ni, 2006).

Frente a estas informações pode-se concluir que o conhecimento sobre o conteúdo de TEs na cana-de-açúcar, também um aloploplóide, deve contribuir para o entendimento do funcionamento e evolução deste extenso e complexo genoma.

O Projeto SUCEST (“Sugarcane Expressed Sequence Tag”) teve como objetivo a geração de bibliotecas de cDNA de cana-de-açúcar e o sequenciamento parcial destas sequências expressas (ESTs – “expressed sequence tags”) em larga escala. Esta abordagem permite o acesso à informação genética de organismos com genomas complexos e a identificação de genes preferencialmente expressos em determinados tecidos, de uma maneira mais rápida e econômica que os métodos tradicionais de sequenciamento, tornando o processo de anotação também mais ágil.

Foram geradas 26 bibliotecas de cDNA a partir de uma variedade de tecidos recolhidos de 13 cultivares distintos de cana-de-açúcar, como descrito em Vettore *et al* (2001), de modo a acessar o maior conjunto de transcritos possível. Bibliotecas construídas a partir de RNA extraído de tecidos cultivados *in vitro* e infectados com bactérias endofíticas fixadoras de nitrogênio foram geradas, de modo a acessar genes envolvidos na interação planta-bactéria e na assimilação do nitrogênio em plantas de cana. Bibliotecas construídas a partir de meristema apical de plantas jovens e gemas laterais de plantas adultas visaram a identificação de genes expressos em estágios iniciais da diferenciação de órgãos. As bibliotecas oriundas de calos, gerados a partir de meristemas e submetidos a diferentes tratamentos de temperatura, tiveram como objetivo a identificação de genes induzidos por calor ou frio. Foram também construídas bibliotecas a partir de tecidos em diferentes estágios do desenvolvimento de flores, assim como de folhas de plantas adultas e de folhas estioladas de plantas crescidas *in vitro*, de modo a identificar genes expressos nestes importantes órgãos da planta. Diversas bibliotecas provenientes de raízes foram geradas, como as de amostras de raízes de plântulas crescidas *in vitro*, ou de plantas crescidas em estufa, e as geradas a partir da porção que vai da raiz à zona caulinar de plantas jovens crescidas em casa de vegetação. Foram ainda preparadas bibliotecas a partir do colmo, para a identificação de genes envolvidos na síntese de parede celular, assim como de sementes

germinadas. Por fim, foram geradas bibliotecas de tecidos provenientes do primeiro e quarto entrenós de plantas adultas, coletados no momento de maior síntese e acúmulo de sacarose.

Estas diversas bibliotecas de cDNA foram sequenciadas em 23 laboratórios, localizados em universidades e institutos de pesquisa do estado de São Paulo. Como resultado bruto do projeto foram sequenciadas 291.689 ESTs. Deste total, 259.325 clones de cDNA foram sequenciados parcialmente em sua porção 5', e 32.364 destes clones foram parcialmente sequenciados também em sua porção 3'. Após a remoção de sequências de baixa qualidade, ou seja, de qualidade menor que 20 no programa computacional "Phred", e de sequências relativas a vetores ou a RNAs ribossomais, restaram 237.954 ESTs potencialmente derivadas de RNAs mensageiros relativos a genes codificantes de proteínas. Estas 237.954 ESTs analisadas foram agrupadas, por identidade de sequência, em 43.141 "clusters" (Vettore *et al*, 2001). Os dados gerados por este projeto, incluindo as sequências consenso dos clusters e dos clones de cDNA sequenciados, estão depositados no banco de dados do SUCEST.

Os dados em larga escala gerados pelo SUCEST podem propiciar a realização de análises genômicas comparativas entre a cana-de-açúcar e outras espécies de plantas, como *arabidopsis* e o arroz, que já tiveram seus genomas completamente sequenciados. Vincentz *et al* (2004) realizou uma análise comparativa entre o transcriptoma da cana e sequências de DNA e proteínas de eudicotiledôneas e monocotiledôneas, depositadas em bancos de dados de acesso público. O objeto de estudo deste trabalho foram 42.982 sequências únicas e "contigs" que podem representar transcritos únicos da cana (Telles e Da Silva, 2001), denominados SASs ("sugarcane assembled sequences" - sequências montadas de cana-de-açúcar). Estas sequências foram submetidas a uma cascata de comparações com bancos de dados através de BLASTN e TBLASTX. Esta cascata comparou as sequências das ESTs da cana com as sequências do genoma de *arabidopsis*, em seguida com as sequências de ESTs de outras eudicotiledôneas, e finalmente com as sequências do genoma do arroz e de ESTs de outras monocotiledôneas. Como resultado desta análise, 70,5% das SASs apresentaram identidade com sequências de *arabidopsis*. Entre estas, 99,1% também apresentaram

identidade com sequências de arroz, e podem representar o conjunto de genes conservado entre eudicotiledôneas e monocotiledôneas. Das sequências de SASs remanescentes, 2% apresentaram similaridade com sequências de ESTs de outras eudicotiledôneas, e podem representar genes perdidos no genoma de *Arabidopsis*. Por outro lado, 14% das sequências apresentaram identidade ou similaridade com sequências do genoma do arroz e de ESTs de outras monocotiledôneas, o que pode indicar tratarem-se de genes específicos de monocotiledôneas. Admitiu-se que os 13,5% de sequências ainda remanescentes foram encontradas somente no transcriptoma da cana, e podem representar sequências altamente variáveis (com altas taxas de evolução) em relação a seus homólogos em outras monocotiledôneas, e que são específicas da cana-de-açúcar. Entre estas sequências de rápida evolução são destacados fatores de transcrição, genes de resposta de defesa e resposta a estresse e elementos de transposição (TEs). Como não atuam como os genes constitutivos das células, que codificam proteínas necessárias para sua manutenção, os TEs toleram o aparecimento de mutações em suas sequências em taxas maiores que os genes constitutivos, uma vez que não são submetidos a uma seleção negativa devido a estas mutações. Devido a esta característica é esperado que haja sequências de TEs exclusivas de cana, como encontrado por Vincentz *et al* (2004).

Um trabalho de mineração dos dados do banco do SUCEST detectou 276 clones de cDNA que apresentaram homologia com elementos de transposição, entre retrotransposons e transposons (Rossi *et al*, 2001). Esta anotação foi realizada através de comparação com o banco de dados do Genbank, que foi disponibilizada através de busca por palavras chave (Telles *et al*, 2001). Foram utilizadas as palavras chave “transposable element”, “transposase”, “transposon” e “retrotransposon”. Para manter alta a estricência da busca, foram admitidos somente os resultados com valor de corte de e^{-50} ou melhores. Os 276 clones identificados como TEs foram classificados em 21 famílias, de acordo com o melhor alinhamento com um elemento completo descrito na literatura, utilizando o programa BLAST. A distribuição das famílias de TEs encontradas está representada na figura 4. Foram encontrados retrotransposons homólogos a Bs1, Opie2, Hopscotch,

Sto4, Cinfu1 e CentA, todos descritos originalmente em *Zea mays*; MAGGY de *Magnaporthe grisea*; Retrofit de *Oryza longistaminata*; RICCOPIA e RIRE de *Oryza sativa*; Tnt1 de *Nicotiana tabacum*; Retrosor1 de *Sorgum bicolor* e Dea1 de *Ananas comosus*. Entre os transposons também foram encontrados homólogos às principais famílias, incluindo MuDr, Ac e Em/Spm, originalmente descritos em *Zea mays*; Soymar de *Glicine Max*; Tag de *Arabidopsis thaliana*; Candystripe de *Sorgum bicolor*; Tam de *Antirrhinum majus* e Tip100 de *Ipomea purpurea*.

Figura 4 – Distribuição das famílias de elementos transponíveis encontradas na anotação das “tags” no projeto SUCEST. Adaptado de Rossi *et al* (2001). As famílias de retrotransposons mais representadas são Hopscotch, Bs1, Opie2 e Maggy. Para os transposons as famílias mais representadas são MuDr, Ac e Tam3.

A abordagem de analisar sequências expressas (ESTs) permitiu a identificação de TEs transcricionalmente ativos, diferentemente das estratégias clássicas em que os TEs são primeiramente analisados em nível de DNA para posteriormente ser avaliada sua transcrição. O trabalho de Rossi e colaboradores (2001) revelou um grande número de TEs expressos no genoma da cana-de-açúcar, distribuídos em um amplo espectro de famílias tanto de

retrotransposons quanto de transposons. Estes elementos normalmente encontram-se silenciados nos genomas, através de sistemas de controle transcricional e pós-transcricional, a fim de preservar a integridade dos genomas hospedeiros e, conseqüentemente, sua própria existência. Sabe-se que TEs silenciados em plantas diplóides podem ser ativados em um novo ambiente genético gerado pela poliploidia (Voytas e Naylor, 1998). Com base nestas informações, a grande quantidade de TEs transcionalmente ativos encontrados na cana-de-açúcar, um híbrido alopoliplóide, sugere que estes TEs podem ter sido ativados devido ao impacto causado em seu genoma pelo evento da união de genomas de duas espécies diferentes e pela poliploidia.

Dentro do universo dos 276 clones de cDNA homólogos a TEs encontrados por Rossi e colaboradores (2001), 156 foram sequenciados completamente como objetivo inicial do presente trabalho. As sequências completas permitiram análises detalhadas dos cDNAs das famílias de transposons mais representadas, Mutator e *hAT*. Os homólogos ao transposon Mutator foram analisados quanto à sua diversidade e apresentaram-se agrupados em quatro classes distintas (Rossi *et al*, 2004). Posteriormente verificou-se que duas destas classes representavam transposons legítimos, enquanto as outras duas estavam relacionadas a transposases domesticadas do tipo Mustang (Saccaro Junior *et al*, 2007).

II. OBJETIVOS

II.1. Objetivo Geral

O presente projeto tem como objetivo geral estabelecer um panorama completo dos TEs transcricionalmente ativos em cana de açúcar, em particular aprofundar a caracterização dos transposons homólogos à superfamília *hAT*.

II.2. Objetivos Específicos

1. Caracterizar a diversidade de sequências com similaridade a elementos de transposição no banco SUCEST de cana-de-açúcar mediante sequenciamento completo dos cDNAs.
2. Associar os cDNAs a famílias bem estabelecidas de elementos de transposição mediante estudo de similaridade em bancos de dados.
3. Avaliar a diversidade dos transposons da superfamília *hAT*.
4. Caracterização molecular dos TEs de grupos I e II identificados dentro da superfamília *hAT* em cana-de-açúcar.

III. MATERIAIS E MÉTODOS

III.1. Material Vegetal e Condições de Crescimento

III.1.1. Material Vegetal Utilizado para Extração de DNA

Foram utilizados para extração de DNA genômico de cana-de-açúcar, dois tipos de materiais. Primeiramente folhas das variedades híbridas de cana-de-açúcar SP-80-3280 e SP-89-1115, e dos parentais *Saccharum spontaneum* (Mandalay) e *Saccharum officinarum* (Badila), cultivadas em campo, sob condições naturais de temperatura e iluminação, gentilmente cedidas pelo Dr. Eugênio Ulian (COPERSUCAR). Estabeleceu-se, depois, o cultivo da variedade de cana-de-açúcar SP-80-3280 em estufa, cultivada em vasos com terra (50% vermiculita), sob condições naturais de temperatura e iluminação. Folhas destas plantas também foram utilizadas para a extração de DNA genômico. Também foi extraído DNA de folhas das gramíneas *Oryza sativa* (arroz) e *Zea mays* (milho), e das monocotiledôneas *Catassatum fimbriatum* (orquídea) e *Ananas comosus* (abacaxi). As sementes de arroz e milho foram obtidas junto à EMBRAPA e germinadas em algodão embebido em água, sob um fotoperíodo de 12 horas à temperatura ambiente. As plantas de orquídea e abacaxi foram gentilmente cedidas pelo laboratório do Prof. Dr. Gilberto Kerbauy (IB - USP). O DNA genômico da eudicotiledônea *Solanum tuberosum* (batata) foi gentilmente cedido pela doutoranda Maria Elisa Manetti (Laboratório GaTE - IB - USP).

III.1.2. Material Vegetal Utilizado para Extração de RNA

O RNA utilizado neste trabalho foi extraído de dois tipos de materiais. Primeiramente foi realizada uma extração de RNA de plantas da variedade de cana-de-açúcar R570, gentilmente cedida pela Profa. Dra. Angelique D'Hont (CIRAD – França). Estas plantas foram crescidas em vasos com terra (50% vermiculita) em estufa, sob condições naturais de temperatura e iluminação. Em seguida foram transferidas para cultivo em hidroponia, ficando com suas raízes submersas em água adicionada de solução nutritiva para plantas 1:1,

sob aeração. Por último foi extraído RNA de calos da variedade de cana-de-açúcar SP-89-1115, crescidos em meio MS suplementado com 20g de sacarose/L, acrescido de 2,4-D para uma concentração final de 1mg/mL. Os calos cresceram protegidos da luz, sob a temperatura de 24°C (+/-2°C).

Solução nutritiva para plantas - $\text{Ca}(\text{NO}_3)_2$ (5mM), KNO_3 (5mM), MgSO_4 (2mM), KH_2PO_4 (1mM), FeEDTA (1mM). Adicionada de 0,6 mL de solução de micronutrientes por litro.

Solução de Micronutrientes - H_3BO_3 (46mM), $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ (9,14mM), ZnCl_2 (1mM), CuCl_2 (0,29mM), $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ (0,1mM).

III.2. Cepa Bacteriana Utilizada (*E. coli*)

DH10B – genótipo F- $\text{mcrA } \Delta(\text{mrr-hsdRMS-mcrBC}) \Phi80\text{d lacZ } \Delta\text{MIS } \Delta\text{lacX74 endA1recA1 deoR } \Delta(\text{ara,leu}) 7697 \text{ araD139 galU galK nupG rpsL}$.

III.3. Meios de Cultura

III.3.1. Meio de Cultura para Calos

MS 20 - Murashige & Skoog (Sigma) – Macro e Micronutrientes (Sigma M5524) 4,3g/L, vitamina MS (Sigma M3900) 1mL/L, acrescido de 2,4-D para uma concentração final de 1mg/mL e sacarose a 20g/L. O pH foi ajustado com KOH para 5,7 e o volume foi ajustado com água destilada. O meio foi acrescido de 6 g/L de fitoágar (GIBCO) e autoclavado a 120°C por 20 minutos.

III.3.2. Meios de Cultura para Bactérias

LB - 10g bactotripton; 5g extrato de levedura; 5g NaCl; pH ajustado para 7,4 com NaOH; água destilada para completar o volume de 1 litro.

SOC – 5g bactotripton; 1,25g extrato de levedura; 500µl NaCl 5M; 620µl KCl 1M; 2,5mL MgCl_2 1M; 2,5mL MgSO_4 1M; glicose para uma concentração de 20mM; pH ajustado para 6,8; água destilada para completar o

volume de 250mL.

SOB – 20g bactotripton; 5g extrato de levedura; 2mL NaCl 5M; 2,5mL KCl 1M; 10mL MgCl₂ 1M; 10mL MgSO₄ 1M; água destilada para completar o volume de 1 litro.

Circle Grow (BIO 101 – Inc.) – de acordo com as instruções do fabricante.

Os meios foram autoclavados a 120°C por 20 minutos. Para os meios sólidos foram acrescentados 20g de ágar por litro. Quando necessário, os meios foram acrescidos do antibiótico de seleção, IPTG na concentração final de 200µg/mL e X-Gal na concentração final de 50µg/mL.

III.4. Clones de cDNA de Cana-de-açúcar

Os clones de cDNA utilizados neste trabalho são provenientes do projeto de sequenciamento de ESTs de cana-de-açúcar da FAPESP (SUCEST). As bibliotecas de cDNA foram construídas a partir de RNA obtido de diferentes tecidos de 13 variedades de cana-de-açúcar, submetidos a diferentes tratamentos, de modo a abranger a expressão dos mais variados genes. As bibliotecas foram montadas como descrito em Vettore *et al*, 2001. Na anotação do SUCEST, 276 clones foram identificados como homólogos a elementos de transposição (Rossi *et al*, 2001). Estes clones bacterianos foram obtidos junto aos laboratórios que participaram do SUCEST, e esta coleção passou a ser mantida em nosso laboratório (GaTE lab – IB - USP) em estoque de glicerol 50%, a uma temperatura de -70°C.

III.5. Eletroforese e Géis de Agarose

Os géis de agarose utilizados no decorrer deste projeto foram preparados da seguinte maneira: agarose (Gibco) numa concentração de 0,7% a 2%, dissolvida em tampão TBE 1/2X, levando-se ao microondas até a mistura entrar em ebulição. Em seguida os géis foram vertidos em molde apropriado para polimerizar à temperatura ambiente. Os DNAs, as reações de PCR e de hidrólise com endonucleases de restrição foram adicionados de tampão de amostra para eletroforese de DNA, e então foram aplicados no gel. As cubas

utilizadas para eletroforese foram os modelos “Horizon 58” e “Horizon 11.14” (Gibco BRL). Após a eletroforese, os géis foram corados em uma solução de brometo de etídeo a 5µg/mL, lavados em água destilada, visualizados no “Image Master®VDS” (Pharmacia Biotech), e fotografados no “Thermal Imaging System FTI-500” (Fuji Film).

TBE 10X – 108g de Tris base; 55g de Ácido bórico; 40mL de EDTA 0,5 M pH8,0. Completar o volume para 1L com água deionizada. Autoclavar a 120°C por 20 minutos.

Tampão de amostra para eletroforese de DNA - azul de bromofenol 0,25% (p/v); xileno cianol 0,25% (p/v) e glicerol 30% (p/v). Acertar o volume com água deionizada.

III.6. Manipulações Gerais com Bactérias *Escherichia coli*

III.6.1. Obtenção de Bactérias Competentes

Para a obtenção de bactérias competentes, um inóculo de colônia isolada de DH10B foi colocado para crescer em 2 mL de meio SOB, a 37°C sob agitação. Em seguida esta cultura foi reinoculada em 50 mL de meio, e incubada sob agitação a 37°C até atingir uma medida de absorbância entre 0,5 e 0,6, medida em espectrofotômetro com o comprimento de onda de 600nm. Foi adicionado 0,5 mL de MgCl₂ 1M à cultura, que ficou no gelo por 15 minutos. As bactérias foram centrifugadas a 3000 RPM por 20 minutos, a 4°C. O sobrenadante foi descartado, e o precipitado ressuscitado em 15 mL de solução RFI gelada, ficando por 10 minutos no gelo. Em seguida uma nova centrifugação foi realizada, e as bactérias foram ressuscitadas em 2 mL de solução RFII gelada. Este volume foi dividido em alíquotas de 50µL, que foram estocadas a -70°C. Toda a manipulação foi realizada em fluxo laminar, com o uso de material estéril.

Solução RFI - acetato de potássio (30mM), cloreto de manganês (50mM), cloreto de potássio (100 mM), cloreto de cálcio (10mM) e glicerol (15%).

Solução RFII - MOPS pH 7,0 (10mM), cloreto de cálcio (75mM) cloreto de potássio (10mM) e glicerol (15%).

As soluções de $MgCl_2$, RFI e RFII foram preparadas com água deionizada e autoclavadas a 120°C por 15 minutos.

III.6.2. Transformação de Bactérias Competentes

Para a transformação de bactérias competentes foi utilizado o método de choque térmico, baseado em Hanahan (1983). As alíquotas de 50µL de bactérias foram retiradas do estoque a -70°C e mantidas em gelo por 10 minutos. Em seguida os vetores desejados foram adicionados às bactérias (1µL de minipreparação plasmidial na concentração de aproximadamente 120ng/µL ou 2µL de uma ligação com volume de 10µL). As bactérias foram então mantidas no gelo por 30 minutos, e passaram por um choque térmico ficando a 42°C por 2 minutos e voltando novamente para o gelo por mais 2 minutos. Após isto, 800µL de meio SOC foram adicionados às bactérias, que foram levadas ao agitador a 37°C por uma hora. Estas culturas foram inoculadas em placas de Petri contendo meio LB sólido na presença do antibiótico de seleção, e de IPTG e X-Gal, quando necessário. As placas ficaram incubadas em estufa a 37°C por 16 a 20 horas para o crescimento das bactérias transformadas. Toda a manipulação foi realizada em fluxo laminar, com o uso de material estéril.

III.7. Procedimentos Gerais de Manipulação de DNA

III.7.1. Extração de DNA Plasmidial em Pequena Escala (Miniprep)

As bactérias contendo os clones de interesse foram colocadas para crescer um dia antes da extração de DNA. Em fluxo laminar, um inóculo de cada clone foi coletado com palito de madeira autoclavado, colocado em tubo de vidro para cultura de bactérias contendo 3mL de meio LB com o antibiótico de seleção na concentração desejada, e levado ao agitador de tubos a 37°C

por um período de 18 horas. Em seguida transferiu-se 1,4mL das bactérias para tubo de 1,5mL e centrifugou-se a 20.000g, por 60 segundos, em temperatura ambiente, para que precipitassem (Centrífuga “Mikro 22 R”, Rotor “1151” – Hettich Zentrifugen). Descartou-se o sobrenadante e adicionou-se o restante das bactérias, repetindo a centrifugação e descartando-se o sobrenadante completamente. Ressuspendeu-se as bactérias em 200 μ L de P1 com auxílio de agitador de tubos (Vortex “AP 56” – Phoenix). Adicionou-se 200 μ L de P2 aos tubos, que foram invertidos gentilmente para homogeneizar seu conteúdo, e deixados a temperatura ambiente por 5 minutos. Adicionou-se 200 μ L de P3 e inverteu-se os tubos gentilmente até que se formou um precipitado branco. Nesta etapa os tubos foram mantidos no gelo. Os tubos foram centrifugados em seguida, a 20.000g por 15 minutos a 4°C. O precipitado branco se separou da fase aquosa. O sobrenadante foi coletado com cuidado para não pegar o precipitado. Adicionou-se à fase aquosa 400 μ L de isopropanol, e homogeneizou-se o conteúdo dos tubos por inversão. Centrifugou-se a 20.000g, à temperatura ambiente por 10 minutos. Descartou-se o sobrenadante. Adicionou-se 500 μ L de etanol 70%, misturando-se por inversão dos tubos. Centrifugou-se a 20.000g, a 4°C, por 5 minutos. Descartou-se o etanol em seguida. O DNA precipitado no fundo do tubo foi seco completamente, ficando por 10 minutos em rotação na câmara de vácuo (“Centrivac Console” – Labconco), e ressuspendido em 30 μ L de água deionizada. Estas minipreparações foram então tratadas com RNase I-A a 50 μ g/ μ L, ficando incubadas por 1 hora a 37°C, e estocadas a -20°C.

P1 - Tris 50mM pH 8,0; EDTA 10mM; Rnase 100g/mL.

P2 - 200mM NaOH; 1% SDS.

P3 - KoAc 3M pH 5,5 acertado com ácido acético glacial.

III.7.2. Extração de DNA Plasmidial em Pequena Escala (Miniprep) em Placas de 96 Poços

Em placas de 96 poços, próprias para crescimento de bactérias,

adicionou-se 1mL de meio Circle Grow com o antibiótico desejado. Com carimbo próprio para fazer réplicas de placas de 96 poços, devidamente flambado e resfriado, o inóculo das bactérias foi coletado das placas de estoque de glicerol, e depositado em placas de cultivo (Mega Titer Plate) contendo o meio de cultura com o antibiótico. Estas foram seladas com adesivos (“Adhesive Plate Sealer” – AB Gene e “Adhesive Sealing Films for Microplates” – Simport). Com agulha flambada, um furo foi feito em cima de cada poço, para aeração das bactérias. As placas foram levadas a uma incubadora horizontal com agitação, a 37°C, por um período de aproximadamente 20 horas. Em seguida as placas foram centrifugadas por 6 minutos a 2.000g (Centrífuga “Universal 30 F”, Rotor “1422” – Hettich Zentrifugen). Os adesivos foram removidos e o sobrenadante descartado, vertendo-se as placas sobre papel absorvente por 5 minutos. Adicionou-se 240µL de GTE em cada poço, utilizando-se micropipetador multicanal Finnpipette. As placas foram seladas com adesivos e levadas ao agitador de tubos por 2 minutos, para ressuspender as bactérias precipitadas. Centrifugou-se as placas por 6 minutos a 2.000g, e o sobrenadante foi descartado, vertendo-se as placas sobre papel absorvente por 5 minutos. Adicionou-se 80µL de GTE em cada poço. As placas foram novamente seladas e levadas ao agitador de tubos, para ressuspender o precipitado. Foram adicionados 2,5µL de RNase a 10mg/mL (RNase A – USB – fonte: pâncreas bovino) em cada poço de placas de polipropileno de 96 poços com fundo em “U” (Costar). Transferiu-se 60µL da suspensão de células para cada poço das novas placas contendo RNase. Em seguida adicionou-se a cada poço 60µL de uma solução de NaOH 0,2N 1% SDS em água deionizada. As placas foram seladas com adesivos, invertidas por 10 vezes para homogeneização, e deixadas à temperatura ambiente por 10 minutos. Adicionou-se então, a cada poço, 60µL de uma solução de KoAc 3M gelada. Novamente as placas foram seladas com adesivos, invertidas 10 vezes e deixadas à temperatura ambiente por 10 minutos. Os adesivos foram removidos, e as placas levadas à estufa a 90°C. Após 30 minutos, as mesmas foram retiradas da estufa e colocadas no gelo por 10 minutos. As placas foram centrifugadas por 4 minutos a 2.000g, temperatura ambiente. Todo o volume de cada poço (180µL), foi transferido para placas de

96 poços (“MultiScreen Filter Plate” (cód. MAGVN 22-5) – Millipore), previamente fixadas no topo de placas de polipropileno de 96 poços com fundo em “V” (Costar), de maneira que os poços ficassem alinhados. Centrifugando-se estes conjuntos, o precipitado ficou retido nas placas Millipore, enquanto a fase aquosa contendo o DNA foi filtrada para as placas com fundo em “V”. Adicionou-se, então, 110µL de isopropanol gelado a cada poço contendo o filtrado. As placas foram seladas com adesivos, invertidas por 15 vezes, e centrifugadas a 2.000g por 45 minutos à temperatura ambiente. O sobrenadante foi descartado, invertendo-se as placas sobre papel absorvente. Adicionou-se 200µL de etanol 70% gelado a cada poço. As placas foram centrifugadas a 2.000g por 10 minutos, à temperatura ambiente, e vertidas sobre papel absorvente para remover-se o sobrenadante. Para secar, as placas foram colocadas em estufa a 37°C por 30 minutos. O DNA precipitado no fundo das placas foi ressuspenso em 80µL de água deionizada. A manipulação de bactérias foi realizada em fluxo laminar.

GTE - 23mL de Glicose Anidra 20%; 10mL de EDTA 0,5M pH8,0; 13mL de Tris-HCl 1M pH 7,0; água deionizada para completar um volume de 500mL.

III.7.3. Extração de DNA Genômico de Plantas

Para a extração de DNA genômico das plantas, suas folhas foram primeiramente pulverizadas com bastão em cadinho de porcelana. Para a pulverização, colocou-se no cadinho nitrogênio líquido e 0,5g de PVPP para cada grama de folhas. Após pulverizado, o material foi então transferido para tubo Corex de 30 mL, ao qual foram adicionados 6mL de tampão CTAB para cada grama de folhas. O material foi misturado por inversão, com os tubos incubados a 65°C por 3 minutos. Em seguida foi realizada uma extração com clorofórmio, adicionando-se o mesmo volume de Clorofórmio:Álcool Isoamílico (24:1), e homogeneizando-se os tubos por gentil inversão. Estes foram centrifugados por 20 minutos a 10.000g, à temperatura ambiente. O sobrenadante foi transferido para novo tubo, para extração com Fenol, ao qual

adicionou-se o mesmo volume de Fenol:Clorofórmio:Álcool Isoamílico [1:1(24:1)]. Homogeneizou-se por inversão e centrifugou-se por 20 minutos a 5.000g, em temperatura ambiente (Centrífuga “Mikro 22 R”, Rotor “1414” – Hettich Zentrifugen). O sobrenadante foi transferido para novo tubo e foram realizadas mais duas extrações com Clorofórmio:Álcool Isoamílico, como descrito acima. O sobrenadante da última extração foi transferido para novo tubo, ao qual adicionou-se 2/3 do volume de isopropanol gelado. Inverteu-se lenta e cuidadosamente o tubo, para precipitar o DNA. O precipitado foi então pescado com uma pipeta Pasteur de ponta curvada, e transferido para tubo de 1,5mL contendo 1mL de etanol 70%. O tubo foi invertido para lavar o DNA e posteriormente centrifugado por 5 minutos a 10.000g, temperatura ambiente (Centrífuga “Mikro 22 R”, Rotor “1151” – Hettich Zentrifugen). O precipitado foi seco ficando por 10 minutos em rotação na câmara de vácuo (“Centrivac Console” - Labconco), e ressuspendido em 100µL de TE. Os DNAs genômicos foram estocados em congelador a -20°C.

Tampão CTAB – 2% CTAB (cetiltrimetilbrometo de amônio), 1,4M NaCl, 20mM EDTA (ácido etilenodiaminotetracético), 100mM Tris, 1% PVP (Polivinilpirrolidona). Autoclava-se o tampão e depois adiciona-se Betamercaptoetanol para uma concentração final de 10mM.

Tampão TE - 10mM Tris pH 8,0; 1mM EDTA pH 8,0.

III.7.4. Sequenciamento Automático de Nucleotídeos

O princípio do sequenciamento automático de nucleotídeos é a incorporação de nucleotídeos marcados com moléculas fluorescentes aos fragmentos de DNA de interesse. O sequenciador automático identifica a sequência de bases do fragmento de DNA através do reconhecimento das moléculas de fluorescência. Esta incorporação de nucleotídeos marcados é feita através de uma reação de PCR. O sequenciamento dos fragmentos de DNA foi realizado utilizando-se os iniciadores localizados nas extremidades das sequências. Para fragmentos subclonados em vetores, foram utilizados iniciadores localizados nas regiões 5’ ou 3’ flanqueadoras do sítio múltiplo de

clonagem dos respectivos vetores. Para sequenciamento direto de produtos de PCR foi utilizado ou o iniciador direto (“forward”) ou o reverso (“reverse”) da própria PCR. As reações de sequenciamento consistiram de 240 ng de DNA proveniente das minipreparações, ou 50 ng de produto de PCR. Foram adicionados a cada reação, juntamente com o DNA, 3,2 pmoles do respectivo iniciador; 2µL de Big Dye™ v 3.0 (Applied Biosystems) e água deionizada para completar um volume final de 10µL. Os tubos ou placas de sequenciamento foram levados para o termociclador “PTC 100™ Programmable Thermal Controller” (MJ Research, Inc.), passando pelo termociclo que consistiu de 40 vezes a repetição da seguinte sequência: 96°C por 10 segundos, 52°C por 20 segundos, 60°C por 4 minutos. O DNA marcado com as moléculas fluorescentes foi então precipitado. Para a precipitação, adicionou-se a cada reação 100µL de isopropanol 75% e agitou-se os tubos ou placas no “vortex”. Os mesmos foram deixados por 15 minutos em temperatura ambiente, e em seguida centrifugados a 2000g por 50 minutos, temperatura ambiente. O sobrenadante foi descartado vertendo-se os tubos ou placas sobre papel absorvente. Em seguida foram realizadas duas lavagens do precipitado com etanol 70%. Cada lavagem foi feita com a adição de 150µL de etanol 70% e centrifugação a 2000g por 10 minutos em temperatura ambiente. O sobrenadante foi descartado vertendo-se os tubos ou placas sobre papel absorvente. Os mesmos foram levados à estufa a 37°C por 30 minutos, para que o etanol secasse completamente. Os DNAs marcados foram ressuspensos em 10µL de formamida (“Hi-Di Formamide” P/N 4311320 – Perkin Elmer Biosystems), desnaturados, ficando por 4 minutos a 96°C, e deixados no gelo por 10 minutos. Em seguida as amostras foram colocadas para sequenciar no aparelho “3700 Genetic Analyser” ou “3100 Genetic Analyser” - ABI PRISM. Os eletroferogramas das sequências foram visualizadas no programa Sequencing Analysis 3.4 (ABI PRISM).

III.8. Sequenciamento Completo dos Clones de cDNA do SUCEST

III.8.1. Estimativa do Tamanho dos Clones de cDNA

Uma estimativa de tamanho dos clones a serem sequenciados foi realizada, de modo a permitir o delineamento da melhor estratégia a ser empregada no seu sequenciamento. Como os cDNAs foram subclonados no vetor pSPORT1 (figura 5-A), para estimar seus tamanhos foram realizadas reações de PCR utilizando-se os iniciadores “T7” e “SP6”, que se anelam respectivamente às duas bordas do sítio múltiplo de clonagem (“polylinker”) do pSPORT1. As reações consistiram de 4ng de DNA; 5 μ L de PCR Buffer 10X; 200 μ M de dNTPs; 1mM de MgCl₂; 10pmoles do oligonucleotídeo T7; 10pmoles do oligonucleotídeo Sp6; 2,5U da enzima “Taq DNA Polymerase Recombinant” (Invitrogen) e água deionizada para completar um volume final de 50 μ L. Os tubos foram levados ao termociclador “Gene Amp PCR System 2400” - Perkin Elmer, passando pelo seguinte ciclo: desnaturação inicial de 5 minutos a 94°C, seguida de 35 ciclos de 94°C por 30 segundos, 56°C por 30 segundos e 72°C por 90 segundos, e tempo complementar de extensão a 72°C por 10 minutos. As amostras amplificadas foram migradas em gel de agarose 0,7%, ao lado dos marcadores de peso molecular “100pb DNA ladder” e “1kb DNA ladder” (Gibco BRL), de modo a se estimar seus tamanhos.

III.8.2. Sequenciamento de Clones Menores que 1 Kb

Em cada rodada de sequenciamento, como descrito no item III.7.4, pode-se gerar uma sequência (“read”) de aproximadamente 600 pares de base com boa qualidade. Para se obter as sequências completas dos clones menores que 1kb, foi realizado sequenciamento a partir da extremidade 5’ e da extremidade 3’ dos cDNAs, utilizando-se os iniciadores T7 e SP6. A união destes dois “reads”, através de sobreposição, foi suficiente para se obter a sequência completa dos clones (figura 5-B).

Figura 5: **A:** Representação esquemática dos clones de cDNA do SUCEST, subclonados no vetor pSPORT1. É possível sequenciar os insertos a partir de suas extremidades 5' e 3' utilizando os iniciadores T7 e Sp6, respectivamente. **B:** Diagrama esquematizando a montagem dos clones de tamanho menor que 1kb. Os "reads" gerados pelo sequenciamento com os iniciadores T7 e SP6 foram suficientes para a montagem da sequência completa. **C:** Esquematização da montagem dos clones de tamanhos entre 1kb e 3,5kb através da estratégia de "primer walking". A partir do "read" gerado usando-se o iniciador T7, um novo iniciador era desenhado, permitindo sequenciar mais uma porção do clone, onde era desenhado um novo iniciador, e assim por diante de modo a percorrer toda a sequência do clone de cDNA.

III.8.3. Sequenciamento de Clones de Tamanho entre 1 Kb e 3,5 Kb

A estratégia de “Primer Walking” foi utilizada para se obter as sequências completas dos clones de cDNA de tamanho entre 1kb e 3,5kb. Esta estratégia consistiu em realizar-se primeiramente o sequenciamento (como descrito no item III.7.4), com os iniciadores T7 e SP6. Em seguida um novo iniciador foi desenhado na extremidade de um dos “reads” obtidos, e um novo sequenciamento foi realizado a partir deste iniciador. Este procedimento permitiu caminhar pelas sequências dos clones, até se obter um “read” que apresentasse sobreposição com a região localizada na extremidade oposta. Desta forma a sobreposição de três ou mais “reads” formaram as sequências completas dos clones (figura 5-C).

III.8.4. Sequenciamento de Clones de Tamanho Maior que 3,5 Kb

As sequências dos clones de cDNA de tamanho maior que 3,5 Kb foram obtidas através da construção de sub-bibliotecas destes clones (“Shot Gun”). Os cDNAs foram fragmentados por sonicação, seus fragmentos foram subclonados no vetor pMOSBlue e sequenciados. A montagem, por sobreposição, das sequências dos diversos fragmentos permitiu a reconstrução das sequências completas originais dos clones.

III.8.4.1. Sonicação do DNA e Subclonagem dos Fragmentos

O DNA utilizado para fazer as bibliotecas foram quebrados aleatoriamente através de sonicação, realizada no aparelho “Ultrasonic Processor” – Cole Parmer. Aproximadamente 2,5µg das minipreparações de DNA plasmidial dos clones de interesse foram colocados em tubos de 0,2mL, e a sonicação se deu pela aplicação de 4 pulsos de 45 segundos, na potência de 75%. Para purificar os fragmentos de DNA resultantes da sonicação por faixas de tamanho, os produtos sonicados foram separados por eletroforese em gel de agarose “Low Melting” 1%. Dois fragmentos de gel foram excisados, utilizando-se lâminas de bisturi: um contendo os fragmentos de DNA de tamanhos variando entre 0,5kb a 1,5kb, e outro contendo os fragmentos de

DNA de tamanhos entre 1,5kb a 4,0kb. O DNA foi purificado dos fragmentos de gel utilizando-se o “GFX™ PCR DNA and Gel Band Purification Kit” (Amersham), de acordo com as instruções do fabricante. O DNA obtido nas extrações foi quantificado em fluorômetro (DyNA Quant 200, da Hoefer). Para realizar-se a subclonagem, as extremidades dos fragmentos de DNA foram preenchidas, ficando com pontas cegas, e em seguida foram desfosforiladas, utilizando-se enzimas e reagentes do “pMOS Blue Blunt Ended Cloning Kit” (Amersham), de acordo com as instruções do fabricante. Após este tratamento, procedeu-se a ligação dos fragmentos com o vetor pMOS Blue, utilizando-se também o mesmo kit.

III.8.4.2. Sequenciamento das Sub-Bibliotecas e Montagem das Sequências

O sequenciamento dos clones das sub-bibliotecas foi realizado como descrito no item III.7.4. As sequências obtidas foram isoladas do vetor pMOS Blue e montadas por sobreposição utilizando-se o programa “PhredPhrapConsed” (pacote PPC). A união das sequências das sub-bibliotecas permitiu reconstituir as sequências completas dos clones de cDNA.

III.9. Análise das Sequências Completas dos Clones do SUCEST

A análise das sequências foi realizada com o intuito de se determinar a qual família de elementos transponíveis, previamente descritas na literatura, pertencia cada cDNA. Para isto foram feitas comparações de suas sequências com os bancos de dados do Genbank (www.ncbi.nlm.nih.gov), utilizando-se os programas BLASTX e BLASTP (Altschul *et al*, 1990 e Zhang *et al*, 1997). Através destas comparações determinaram-se os elementos transponíveis com os quais os clones apresentaram maior homologia, ou seja, similaridade de sequência. Para determinar a fase de leitura das sequências dos clones, estas foram alinhadas com as sequências de aminoácidos dos elementos homólogos, utilizando-se o programa BLAST-2-Sequences/BLASTX (Tatusova *et al*, 1999). Foram também verificadas as bibliotecas de origem de cada clone, consultando as informações contidas em Vettore *et al* (2001).

III.10. Análise da diversidade entre os cDNAs homólogos à superfamília *hAT*

III.10.1. Alinhamento de Domínios Conservados das Transposases do tipo *hAT* dos cDNAs de cana-de-açúcar

Para analisar a diversidade presente entre os cDNAs homólogos à superfamília *hAT* da cana-de-açúcar, foi feita uma comparação entre suas sequências de aminoácidos. Esta comparação foi realizada através do alinhamento de segmentos relativos a dois domínios protéicos característicos desta superfamília, descritos em Kempken *et al* (2001). O domínio de dimerização tem como número de acesso no banco Pfam PF05699, e o outro domínio, próximo ao domínio de dimerização, ainda não teve sua função descrita. Foram incluídas no alinhamento sequências destes domínios presentes nos elementos originais da superfamília *hAT*, Ac, Tam3 e *hobo*, e em outros organismos como arroz, milho, arabidopsis e humano. Para o alinhamento foi utilizado o programa “Clustal X 1.81” (Thompson *et al*, 1997), com a função múltiplo alinhamento, parâmetros “default”. As sequências de outros organismos incluídas no alinhamento foram baixadas do “Genbank”. Elas foram identificadas no banco por comparação com os elementos Ac e Tam3, através de BLASTX e BLASTP. Os números de acesso destas sequências encontram-se listados na tabela VIII. O alinhamento e a árvore resultantes encontram-se na figura 8.

III.10.2. Alinhamento de um Domínio Conservado das Transposases do tipo *hAT* presente nos cDNAs de cana-de-açúcar e no gene DAYSLEEPER

A descrição do gene DAYSLEEPER de arabidopsis, o primeiro gene derivado de uma transposase descrito em plantas (Bundock and Hooykaas, 2005) levou à pergunta de que se esta transposase domesticada estaria relacionada a algum dos grupos de cDNAs de cana-de-açúcar homólogos a

hAT. Para responder a esta pergunta foi realizado um alinhamento entre as sequências de aminoácidos de DAYSLEEPER e de cDNAs dos grupos I e II. O segmento alinhado localiza-se entre os aminoácidos 679 e 712 da transposase de Ac e 640 a 673 da transposase de Tam3. O alinhamento múltiplo foi realizado utilizando-se o programa Clustal X nos parâmetros “default”. O alinhamento e a árvore resultantes encontram-se na figura 11.

III.10.3. Busca por Genes Homólogos aos cDNAs de Cana-de-Açúcar no Genoma do Arroz

Após a análise da diversidade dos cDNAs de cana, que indicou dividirem-se em dois grupos distintos, foi realizada uma busca por sequências homólogas aos cDNAs do grupo I da cana no genoma do arroz. Para a busca foi utilizada como “sonda eletrônica” a sequência inteira de nucleotídeos do cDNA TE074. A busca foi realizada através de comparação da sequência do TE074 com bancos de dados do genoma do arroz (Genbank - na seção “Plant Genomes Central”).

III.10.4. Alinhamento dos Clones de cDNA do grupo I com Transposases Domesticadas do tipo *hAT*

Após a descrição de DAYSLEEPER, três genes homólogos a esta transposase domesticada foram descritos em trigo e cevada (Muehlbauer et al, 2006). Na cana-de-açúcar, cDNAs do tipo *hAT*, do grupo I, foram caracterizados como homólogos a DAYSLEEPER. No arroz foram identificados homólogos aos cDNAs do grupo I. Com base nestas informações um novo alinhamento foi planejado, de modo a verificar as relações de similaridade entre os cDNAs de cana-de-açúcar do grupo I, seus homólogos em arroz e estas transposases domesticadas. O alinhamento foi realizado no programa ClustalX, incluindo-se um segmento das sequências traduzidas dos cDNAs de cana, das sequências das transposases dos elementos Ac e Tam3 e das transposases domesticadas DAYSLEEPER, w-gary1, w-gary2 e b-gary (tabela IX). Foram alinhados três cDNAs do grupo I, TE048, TE074 e TE096. O segmento de 260 aminoácidos incluído no alinhamento engloba os dois domínios conservados

descritos no item III.10.1, concatenados. O alinhamento e a árvore resultantes encontram-se na figura 12.

III.10.5. Geração de Árvores Filogenéticas a partir de Alinhamentos

As árvores filogenéticas apresentadas neste trabalho foram geradas a partir dos diversos alinhamentos descritos, utilizando-se o programa Mega 3.1 (Kumar *et al*, 2004). As árvores foram geradas através do algoritmo de “neighbor joining” (Saitou e Nei, 1987), com teste de “bootstrap” de 500 réplicas.

III.11. Distribuição dos Elementos Homólogos à Superfamília *hAT* nos Genomas de Cana-de-açúcar e Outras Monocotiledôneas

A análise de diversidade demonstrou haver dois grupos distintos entre os cDNAs homólogos a *hAT*. Para identificar o padrão de distribuição destes dois grupos nos genomas de variedades parentais e híbridas de cana-de-açúcar e de outras monocotiledôneas, foi realizado um ensaio de hibridação molecular. Sondas de clones representantes dos dois grupos foram hibridizadas contra DNAs genômicos de diferentes espécies, com o objetivo de detectar sua presença nestes genomas. Os DNAs genômicos utilizados foram dos parentais da cana-de-açúcar, *Saccharum spontaneum* (Mandalay) e *Saccharum officinarum* (Badila), das variedades de cana-de-açúcar SP-80-3280 e SP-89-1115, das gramíneas *Oryza sativa* (arroz), e *Zea mays* (milho), das outras monocotiledôneas *Catassatum fimbriatum* (orquídea) e *Ananas comosus* (abacaxi), e da eudicotiledônea *Solanum tuberosum* (batata). Estes materiais vegetais estão descritos no item III.1.1. Os DNAs genômicos foram extraídos como descrito no item III.7.3.

III.11.1. Hibridação Molecular

III.11.1.1. Sondas

Foram realizados dois ensaios de hibridação molecular com DNA genômico, utilizando-se diferentes sondas. No primeiro ensaio foram utilizados como sondas cDNAs inteiros, o TE203 do grupo I, com 1005 pb, e o TE001 do grupo II, com 1844 pb. Os cDNAs foram liberados do vetor pSPORT1 através de hidrólise com endonucleases de restrição do sítio poli-linker. O produto da hidrólise foi migrado em gel de agarose e as bandas de interesse foram isoladas do gel e purificadas utilizando-se o “Perfectprep Gel Cleanup Kit” (Eppendorf).

Posteriormente, novas sondas foram desenhadas, internamente às sequências dos cDNAs, com o objetivo de obter um perfil de hibridação mais específico. As regiões correspondentes às sondas não são compartilhadas pelos TEs dos grupos I e II, como forma de garantir uma hibridação estridente. Estas sondas foram obtidas através de amplificação por PCR sobre as minipreparações de DNA plasmidial dos clones de cDNA. Foram utilizados na PCR os seguintes iniciadores: “SHAT-2-F” e “SHAT-2-R”, desenhados sobre a sequência do elemento TE074, representante do grupo I, que amplificaram uma sonda de 182 pb; “SHAT-3-F” e “SHAT-4-R”, desenhados sobre a sequência do TE191, e “SHAT-5-F” e “SHAT-6-R”, desenhados sobre a sequência do TE257, ambos representantes do grupo II, gerando sondas de 187 pb e 146 pb, respectivamente (figura 6-A). As sondas utilizadas na hibridação foram desenhadas em domínios não conservados, de modo a evitar hibridação cruzada. Para as reações de PCR foi utilizada Taq DNA Polimerase Recombinante (Invitrogen), de acordo com as instruções do fabricante. As sondas foram purificadas da reação de PCR com “Perfectprep Gel Cleanup Kit” (Eppendorf), de acordo com as instruções do fabricante. Também foi utilizada na hibridação uma sonda controle das quantidades de DNA nas membranas. Esta sonda compreende uma região interna do elemento Ac original do milho, de 1,6 kb, e foi obtida a partir do vetor pAc, que contém o elemento Ac9 completo. Foi realizada uma hidrólise do vetor com a endonuclease *HindIII*, que liberou o inserto utilizado como sonda (figura 6-B). O produto da reação de hidrólise foi migrado em gel de agarose, e purificado do gel com o “Perfectprep Gel Cleanup Kit” (Eppendorf), de acordo com as instruções do fabricante.

III.11.1.2. Hidrólise do DNA Genômico

Para o primeiro ensaio de hibridação os DNAs genômicos foram hidrolisados com as endonucleases *HindIII* e *XbaI* (New England Biolabs). No segundo ensaio foram utilizadas as endonucleases *DraI* e *XbaI* (New England Biolabs). As reações foram montadas como descrito a seguir: ***HindIII*** - 20µg de cada DNA, endonuclease (10U/µg de DNA), tampão da enzima para 1X e água deionizada para completar um volume final de 60µL. ***DraI*** - 20µg de cada DNA, endonuclease (10U/µg de DNA), tampão da enzima para 1X e água deionizada para completar um volume final de 60µL. ***XbaI*** - 20µg de cada DNA, endonuclease (10U/µg de DNA), tampão da enzima para 1X, BSA (bovinum serum albumine) para 1X e água deionizada para completar um volume final de 60µL. As reações de hidrólise ficaram incubadas a 37°C por um período de 4 horas.

III.11.1.3. Gel de Agarose

As reações de hidrólise foram migradas em gel de agarose 0,8% em TBE 1/2X. Realizou-se corrida por eletroforese a 37V, 15mA, por 16 horas. Após a eletroforese, os géis foram corados em uma solução de brometo de etídeo a 5µg/mL, visualizados no “Image Master®VDS” (Pharmacia Biotech), e fotografados no “Thermal Imaging System FTI-500” (Fuji Film).

A

B

Figura 6: Esquematização das sondas utilizadas na hibridação molecular. **A:** Os iniciadores “SHAT-2-F” e “SHAT-2-R” foram desenhados sobre a sequência do elemento representante do grupo I, TE074, e amplificaram uma sonda de 182 pb. Os iniciadores “SHAT-3-F” e “SHAT-4-R” foram desenhados sobre a sequência do TE191, e os iniciadores “SHAT-5-F” e “SHAT-6-R”, sobre a sequência do TE257, ambos representantes do grupo II, gerando sondas de 187 e 146 pb, respectivamente. Os números entre parênteses ao lado dos nomes dos iniciadores indicam suas posições nas sequências de nucleotídeos dos elementos. **B:** Esquematização do vetor pAc, que contém o Ac9 subclonado no sítio de *Pst*I do plasmídeo pBR322. A sonda do elemento Ac foi obtida pela hidrólise do vetor com a endonuclease *Hind*III, liberando um inserto de 1,6 kb.

III.11.1.4. Southern Blot (Transferência Alcalina)

Após fotografado, o gel foi submerso por 15 minutos em solução de ácido clorídrico 0,25N, passando por água destilada para enxágue e submerso por 20 minutos em NaOH 0,4N. O DNA foi transferido do gel para a membrana de náilon por capilaridade. A transferência foi montada colocando-se o gel sobre papéis de filtro, em contato com o tampão de transferência (NaOH 0,4N). Sobre o gel foi colocada a membrana de náilon (“Gene Screen Plus Hybridization Transfer Membrane” – NENTM Life Science), mais dois papéis de filtro e toalhas de papel. A transferência transcorreu por um período de 16 horas. Após isto o sistema foi desmontado e a membrana foi submersa em uma solução de 2X SSC por 15 minutos, e seca sobre a bancada entre toalhas de papel.

III.11.1.5. Pré-Hibridação

As membranas foram colocadas em garrafas próprias, contendo 15 ml da solução de pré-hibridação e levadas ao forno a 65°C, com rotação por 4 horas.

Solução de pré-hibridação - 10% dextran sulfato, 1% SDS, 6% NaCl, em água deionizada.

III.11.1.6. Marcação das Sondas

As sondas foram marcadas com P α^{32} dCTP, utilizando-se o “Kit Random Primers DNA Labelling System” da Invitrogen, de acordo com as instruções do fabricante. Adicionou-se à sonda 0,5 mg de DNA de esperma de salmão para cada 10 mL da solução de pré-hibridação. A desnaturação desta mistura foi feita adicionando-se 1/3 do volume de NaOH 1N e deixando-se por 5 minutos à temperatura ambiente. Após isto adicionou-se 1X o volume de Tris

1M pH8,0, deixando 5 minutos à temperatura ambiente e finalmente, colocando-se no gelo por 2 minutos.

III.11.1.7. Hibridação

As sondas, devidamente marcadas, foram colocadas nas garrafas com as membranas e a solução de pré-hibridação. Estas retornaram para o forno a 65°C, com rotação, por 16 horas.

III.11.1.8. Lavagens das Membranas

Após o período de hibridação as membranas passaram por duas lavagens, uma com a solução 1 e outra com a solução 2. Foram adicionados 20 mL de solução por garrafa, que passou pela lavagem ficando 30 minutos no forno a 65°C, com rotação. Após as lavagens as membranas foram enxaguadas em cuba contendo a solução 2, e secas com toalhas de papel. Para a desibridização (e reutilização), as membranas foram submersas em uma cuba contendo SDS 0,5% pré-aquecido até o ponto de ebulição, e levadas ao forno a 65°C por 30 minutos. Este procedimento foi repetido. As membranas foram enxaguadas com água destilada e secas entre toalhas de papel sobre a bancada, ficando prontas para uma nova hibridação.

Solução 1: 0,5X SSC / 0,1% SDS em água deionizada.

Solução 2: 0,1X SSC / 0,1% SDS em água deionizada.

III.11.1.9. Exposição e Revelação

As membranas foram envolvidas com filme plástico e colocadas em cassetes de exposição, juntamente com um intensificador e uma chapa de

filme de raio X (Kodak), ficando mantidas a -70°C . Esta chapa de filme ficou exposta à membrana pelo tempo necessário para sua impressão. A revelação do filme foi feita em tanque de revelação em sala escura, utilizando-se revelador e fixador Kodak. Para isso, o filme foi mantido no revelador por 3 minutos, na água por 2 minutos para enxágue, no fixador por 5 minutos e novamente na água por dois minutos. O filme foi finalmente enxaguado com água corrente e colocado para secar.

III.12. Estudo da Expressão dos Elementos Homólogos à Superfamília *hAT* de Cana-de-açúcar

III.12.1. Hibridação Molecular de RNA

Para o estudo da expressão dos elementos homólogos a *hAT* do grupo I, foi realizado ensaio de hibridação molecular de RNA. Investigou-se se este grupo apresentaria um padrão de expressão diferenciado em folhas, raízes e calos de cana-de-açúcar. As manipulações de RNA requerem um tratamento diferenciado de todo o material a ser utilizado, para a eliminação das RNases. Toda a vidraria, assim como cadinhos, bastões e espátulas devem ser tratados com clorofórmio, enxaguados com água deionizada e esterilizados por autoclavagem de 20 minutos a 120°C . Cubas de eletroforese, pentes e suportes para o preparo dos géis devem ser tratados por 16 horas com uma solução de NaOH 0,2N e enxaguados com água deionizada. Toda a água deionizada empregada no preparo das soluções utilizadas no procedimento deve ser tratada com DEPC. Foram utilizadas pipetas e ponteiros descartáveis e todo o procedimento foi feito com luvas.

Água DEPC – a água deionizada é acrescida de 0,01% DEPC (Sigma), e fica mantida a 37°C por 16 horas. Após este período é autoclavada por 20 minutos a 120°C . Este tratamento inativa RNases presentes na água.

III.12.1.1. Extração de RNA Total

A extração de RNA dos materiais vegetais foi feita utilizando-se o método do cloreto de lítio. Para cada amostra da extração foram pulverizados 100 mg de material, em cadinhos de porcelana contendo nitrogênio líquido, até obter-se um pó fino. Este pó foi transferido para microtubos de 1,5 mL, aos quais se adicionou 0,5 mL de tampão de extração e 0,5 mL de fenol:clorofórmio 1:1. O material foi homogeneizado por agitação em “vortex” por 30 segundos, e em seguida centrifugado a 13.000g por 5 minutos a 4°C. A fase aquosa contendo o RNA foi transferida para novos tubos, e repetiu-se a extração com tampão de extração e fenol:clorofórmio 1:1. A fase aquosa foi transferida para novos tubos em que se adicionou o mesmo volume de LiCl 6M. As amostras foram mantidas a 4°C por 16 horas. Após este período, foram centrifugadas a 13.000g por 10 minutos a 4°C. O sobrenadante foi descartado e o precipitado ressuscitado em 1 mL de LiCl 3M, seguido de uma nova centrifugação. Novamente o sobrenadante foi descartado e o precipitado ressuscitado em 250µL de água deionizada tratada com DEPC. Uma breve centrifugação de 30 segundos a 13.000g foi realizada, para precipitar eventuais partículas de matéria insolúvel que ainda permanecessem nas amostras. O sobrenadante foi recolhido e passado para novos tubos, aos quais foi adicionado $\frac{1}{10}$ do volume de acetato de sódio 3M, e 2X o volume de etanol 100%. Em seguida os tubos foram mantidos a -80°C por 15 minutos para precipitação do RNA, centrifugados a 13.000g por 15 minutos a 4°C e o sobrenadante foi descartado. O precipitado foi lavado com 0,5 mL de etanol 70% gelado, centrifugado 13.000g por 5 minutos a 4°C e o etanol 70% foi descartado. O RNA precipitado foi ressuscitado em 50 µL de água deionizada tratada com DEPC, ficando por 10 minutos a 55°C para ajudar na solubilização. Os RNAs ficaram estocados em congelador a -80°C.

LiCl 6M – 127,2 g de cloreto de lítio em pó dissolvidos em 500 mL de água deionizada. Esterilizar por autoclavagem a 120°C por 20 minutos.

Tampão de extração – Tris-Cl (0,2M); LiCl (0,1M); EDTA (5mM) e SDS 1%.

III.12.1.2. Gel de Formaldeído

Os RNAs foram migrados em gel desnaturante de agarose, constituído de 1,5% agarose; 6,7% Formaldeído; MOPS 1X e água deionizada. A agarose foi adicionada à água e levada ao microondas para dissolver. Após esfriar para 60°C, o MOPS 10X e o formaldeído 37% foram adicionados ao gel para que ficassem nas concentrações mencionadas acima. O gel ficou polimerizando dentro de uma capela com exaustão, para evitar que vapores do formaldeído saíssem para o ambiente. A eletroforese foi realizada em Tampão MOPS 1X. Foi feita uma pré- corrida do gel antes de aplicar as amostras, de 15 minutos a 80V. As amostras a serem aplicadas no gel constituíram-se de 10 µg de RNA em 10µL, acrescidos de 4,6µL de MOPS 10X; 8,2µL de formaldeído 37% e 22,8µL de formamida. Estas ficaram incubadas por 15 minutos a 60°C e em seguida, adicionadas de 9,0µL de tampão de amostra e 1,0µL de brometo de etídeo a 4µg/mL. As amostras foram então aplicadas no gel e realizou-se uma eletroforese a 80V por 3 a 4 horas. Após a eletroforese, o gel foi lavado três vezes por 10 minutos com água deionizada e uma vez com SSC 10X por 40 minutos.

MOPS 10X – MOPS pH 7,0 (0,2 M); Acetato de sódio pH 7,0 (0,5 M) e EDTA (10mM).

Tampão MOPS 10X – MOPS pH 7,0 (0,4 M); Acetato de sódio pH 7,0 (0,1 M) e EDTA (10mM).

Tampão de Amostra – Azul de bromofenol (0,25%); xileno cianol (0,25%); EDTA (1mM) e glicerol (50%).

III.12.1.3. Transferência (Northern Blot)

A transferência ocorreu em SSC 10X, por capilaridade. Esta foi montada colocando-se o gel sobre papel filtro em contato com o SSC 10X da cuba de transferência. Uma membrana de náilon (“Gene Screen Plus Hybridization Transfer Membrane” – NENTM Life Science), previamente umedecida com água deionizada e mantida por 10 minutos em SSC 10X, foi depositada sobre o gel e coberta com papel filtro umedecido com SSC 10X e toalhas de papel. A

transferência transcorreu por um período de 18 horas. Após este período a membrana foi enxaguada rapidamente em SSC 10X, levada ao forno a 80°C por 2 horas e exposta à luz UV por 1 minuto, com a face contendo os RNAs para cima, para sua fixação.

III.12.1.4. Sondas

Foi utilizada na hibridação de RNA a sonda amplificada a partir do TE074 (grupo I) com os iniciadores “SHAT-2-F” e “SHAT-2-R” (figura 6-A), como descrito no item III.11.1.1.

Como controle constitutivo foi utilizada uma sonda obtida a partir da ubiquitina da cana-de-açúcar, gentilmente cedida pela Profa. Dra. Kátia C. Scortecchi (UFRN).

III.12.1.5. Marcação das Sondas

As sondas foram marcadas com P α^{32} dCTP, utilizando-se o “Kit Random Primers DNA Labelling System” da Invitrogen, de acordo com as instruções do fabricante, seguido dos procedimentos descritos no item III.11.1.6.

III.12.1.6. Hibridação

As membranas foram colocadas na pré-hibridação em garrafas apropriadas contendo 30mL de solução de hibridação, e ficaram no forno com rotação, a 65°C por 4 horas. As sondas marcadas foram então adicionadas às garrafas, que voltaram para o forno com rotação a 65°C por um período de 40 horas, para a hibridação.

Solução de Hibridação – Tampão fosfato pH 7,4 (0,5M); SDS (7%) e EDTA (1mM).

III.12.1.7. Lavagem das Membranas

Após o período de hibridação as membranas passaram pelo processo de lavagem, que consistiu-se de uma lavagem com SSC 2X por 5 minutos à temperatura ambiente, duas lavagens de 30 minutos cada com SSC 2X / SDS 0,5% a 65°C, e mais uma lavagem com SSC 2X por 5 minutos à temperatura ambiente.

SSC 20X - Cloreto de sódio (3M) e citrato de sódio (0,3M), dissolvidos em água deionizada. Esterilizado em autoclave a 120°C por 20 minutos.

SDS 20% - 20% do peso/volume de SDS em pó dissolvido em água deionizada.

III.12.1.8. Exposição e Revelação

A exposição de filmes às membranas marcadas radioativamente, e a revelação dos filmes foi realizada do mesmo modo como nas hibridações moleculares com DNA. Estes procedimentos estão descritos no item III.11.1.9.

III.12.1.9. Análise das Autorradiografias Geradas com a Hibridação

A intensidade das bandas geradas pela hibridação molecular de RNA foi medida por densitometria, através do programa “Kodak Digital Science 1D”. Foram medidas as intensidades das bandas e da marcação de fundo. Os valores de fundo foram subtraídos dos valores medidos para as bandas. Para chegar a um valor relativo de expressão, os valores das bandas no filme marcado pela sonda do TE074 foram divididos pelos respectivos valores das bandas no filme marcado pela sonda constitutiva da ubiquitina. Os valores relativos de expressão são apresentados na forma deste quociente, em gráfico de barras.

III.13. Caracterização da Estrutura de Elementos do Genoma da Cana-de-açúcar Homólogos à Superfamília *hAT*

Tendo em mãos as sequências completas dos clones de cDNA, buscamos caracterizar a estrutura de elementos presentes no genoma da cana-de-açúcar relativos ao grupo II, através da amplificação de sequências genômicas. Um ensaio de PCR-Inversa foi realizado com o intuito de acessar as TIRs da hipotética cópia genômica relativa ao cDNA TE221.

Uma vez obtidas as sequências das TIRs, foram desenhados iniciadores sobre as mesmas, direcionados para dentro de modo a amplificar do genoma da cana-de-açúcar variantes deste elemento que compartilham as mesmas TIRs.

III.13.1. Ensaio de PCR-Inversa sobre o DNA Genômico da Cana-de-Açúcar a partir do Clone de cDNA TE221 (Grupo II)

III.13.1.1. Hidrólise do DNA Genômico com Endonuclease de Restrição

Para o ensaio de PCR-Inversa o DNA genômico da variedade de cana-de-açúcar SP-80-3280 foi hidrolisado com a endonuclease de restrição *EcoRI* da New England Biolabs. Para cada 10µg do DNA foram utilizadas 120U de cada endonuclease, tampão da enzima para 1X e água deionizada para completar um volume final de 100µL. As reações ficaram incubadas a 37°C por 6 horas, sendo que metade da enzima foi adicionada às reações no início, e a outra metade foi adicionada após 3 horas de reação.

III.13.1.2. Precipitação do Produto da Hidrólise

O produto da reação de hidrólise foi purificado através de extração com fenol/clorofórmio. O volume das reações de hidrólise foi elevado para 500µL, adicionando-se 400µL de água deionizada. Em seguida adicionou-se 500µL de fenol. As amostras foram homogeneizadas através de refluxo com pipeta e centrifugadas a 13.000g por 10 minutos à temperatura ambiente. A fase aquosa foi recolhida, transferida para outro tubo e adicionada do mesmo volume de clorofórmio:álcool isoamílico 24:1. As amostras foram

homogeneizadas através de refluxo com pipeta e centrifugadas a 12.000g por 10 minutos a temperatura ambiente. A fase aquosa foi recolhida e transferida para novo tubo. Adicionou-se $\frac{1}{10}$ do volume de acetato de sódio 3M e 2,5X o volume de etanol 100%. Os tubos foram mantidos a -20°C por 3 horas para precipitação do DNA e centrifugados a 13.000g por 30 minutos à temperatura ambiente. O sobrenadante foi descartado e o precipitado foi lavado com 0,5 mL de etanol 70% gelado, centrifugado 13.000g por 10 minutos à temperatura ambiente. O etanol 70% foi descartado, as amostras foram secas por 5 minutos em rotação na câmara de vácuo, e ressuspendidas em 30 µL de água deionizada.

III.13.1.3. Reação de Ligação (Circularização)

Foram montadas quatro reações de ligação, utilizando respectivamente 25ng, 75ng, 150ng e 300ng do DNA genômico hidrolisado e precipitado. Foi utilizada a ligase da marca Fermentas e o protocolo de circularização de DNA indicado no manual da enzima. Adicionou-se ao DNA 15U da T4 DNA Ligase, tampão da ligase para 1X e água deionizada para um volume final de 150µL. A reação ficou a 22°C por 2 horas e em seguida ficou a 65°C por 10 minutos para a inativação da enzima.

III.13.1.4. PCR-Inversa

III.13.1.4.1. Iniciadores

Foram utilizados na PCR-Inversa dois pares de iniciadores, um par externo e um par interno (“nested”) (tabela I), desenhados sobre o clone de cDNA TE221, pertencente ao grupo II. Estes iniciadores se ancoram nas duas porções sub-terminais, 5’ e 3’ do TE221, e são direcionados para fora da sequência do cDNA, de modo a amplificar regiões vizinhas.

Tabela I – Iniciadores utilizados na PCR-Inversa.

tipo de iniciador	nome do iniciador	sequência do iniciador
par externo	IPP-TE221-1F	5' GACCAGAAGGCCTGCAAC 3'
	IPP-TE221-1R	5' GACTAATGACTGTAAC TTATCAACG 3'
par interno ("nested")	IPP-TE221-2F	5' GCTCAAAGGGATTATCATTCC 3'
	IPP-TE221-2R	5' GCATCTTCAGTCTTGGCTCA 3'

III.13.1.4.2. Reações de PCR-Inversa

A PCR-Inversa foi realizada utilizando-se uma estratégia de iniciadores "nested". O ensaio consistiu-se de duas rodadas de amplificações. A primeira rodada teve como molde o DNA circularizado através das ligações descritas no item III.13.1.3. Foram montadas oito reações de PCR-Inversa nesta primeira rodada, quatro utilizando como molde as respectivas quatro reações de ligação e o par de iniciadores "IPP-TE221-1F" e "IPP-TE221-1R", e quatro utilizando as mesmas reações de ligação, com os iniciadores "IPP-TE221-1F" e "IPP-TE221-2R", como descrito na tabela II. Foi utilizado para a PCR o "Elongase Enzyme Mix" (Invitrogen) de acordo com as instruções do fabricante. Adicionou-se às reações 10,0 µL da respectiva ligação, 0,2 mM do dNTP mix, 10pmoles de cada iniciador, tampão A para 1X e tampão B para 1X, 2,0µL do "Elongase Enzyme Mix" e água deionizada para um volume final de 50,0 µL de reação. As reações foram submetidas ao seguinte termociclo: 94°C por 2 minutos, 35 vezes a repetição do ciclo de 94°C por 30 segundos, 52°C por 30 segundos e 68°C por 5 minutos, e finalmente um período de 10 minutos a 68°C.

Tabela II – Informações sobre as ligações utilizadas como molde e os iniciadores empregados nas reações de PCR-Inversa da primeira rodada.

número da reação da primeira rodada de PCR-Inversa	quantidade de DNA na ligação utilizada como molde (ng)	par de iniciadores utilizados na primeira rodada de PCR-Inversa
1	25	IPP-TE221-1F IPP-TE221-1R
2	75	
3	150	
4	300	
5	25	IPP-TE221-1F IPP-TE221-2R
6	75	
7	150	
8	300	

A segunda rodada de amplificações da PCR-Inversa foi realizada utilizando-se os iniciadores internos (“nested”) “IPP-TE221-2F” e “IPP-TE221-2R”. Foram montadas oito reações, utilizando o produto das respectivas oito reações da primeira rodada como molde. As reações, feitas com o “Elongase Enzyme Mix” (Invitrogen), consistiram de: 1µL dos respectivos produtos da primeira rodada diluídos 1:100 em água deionizada como DNA molde, 0,2 mM do dNTP mix, 10pmoles de cada iniciador, tampão A para 1X e tampão B para 1X, 2,0µL do “Elongase Enzyme Mix” e água deionizada para um volume final de 50,0 µL de reação. As reações foram submetidas ao seguinte termociclo: 94°C por 2 minutos, 35 vezes a repetição do ciclo de 94°C por 30 segundos, 52°C por 30 segundos e 68°C por 5 minutos, e finalmente um período de 10 minutos a 68°C.

III.13.1.5. Purificação das bandas de interesse da PCR-Inversa

Os produtos da PCR-Inversa foram migrados em gel de agarose “low melting” 0,8%. Os fragmentos de gel contendo as bandas de interesse foram excisados com bisturi e as bandas foram recuperadas utilizando-se o “GFX™ PCR DNA and Gel Band Purification Kit” (Amersham), de acordo com as

instruções do fabricante. Os produtos da PCR-Inversa purificados foram ressuspensos em 30µL de água deionizada.

III.13.1.6. Subclonagem do Produto de PCR-Inversa

Os produtos de PCR purificados foram subclonados no vetor pMOSBlue (G.E.), de acordo com as instruções do fabricante.

III.13.1.7. Transformação de Bactérias Competentes pelo Método do Choque Térmico

Bactérias competentes foram transformadas com os produtos da PCR-Inversa pelo método do choque térmico, como descrito no item III.6.2.

III.13.1.8. Minipreparação de DNA plasmidial

As minipreparações de DNA plasmidial foram realizadas como descrito no item III.7.1.

III.13.1.9. Sequenciamento do Produto de PCR-Inversa

Os produtos da PCR-Inversa subclonados no pMOSBlue foram sequenciados como descrito no item III.7.4, utilizando-se os iniciadores “T7” e “pmos”, ancorados nas bordas do sítio múltiplo de clonagem do vetor.

III.13.1.10. Montagem das Sequências Consenso dos Clones

As sequências dos clones foram montadas utilizando-se o programa computacional “Phred Phrap Consed”. Este programa une as sequências dos “reads” provenientes do sequenciamento com os iniciadores 5’ e 3’, “T7” e “pmos” respectivamente, gerando uma sequência consenso. Os “reads” que não puderam ser unidos tiveram seus eletroferogramas visualizados no programa “BIOEDIT”.

III.13.1.11. Análise das Sequências Obtidas a Partir de PCR-Inversa

A análise inicial das sequências consistiu da busca pelas seguintes estruturas: iniciador 3' (IPP-TE221-2F), porção final da sequência do clone de cDNA TE221, identificação de potencial TIR 3', sítio duplicado de inserção, região genômica a jusante do elemento, sítio da endonuclease *EcoRI*, região genômica a montante do elemento, sítio duplicado de inserção, identificação de potencial TIR 5', porção inicial da sequência do clone de cDNA TE221 e iniciador 5' (IPP-TE221-2R), como esquematizado na figura 17. Para as análises foram utilizados os programas "Phred Phrap Consed" e "BIOEDIT". Os alinhamentos entre sequências foram realizados utilizando-se os programas "BIOEDIT" e "Clustal X 1.81" bem como o programa "BLAST 2 Sequences" disponível online no sítio do Genbank (www.ncbi.nlm.nih.gov). A ferramenta "ORF Finder" do Genbank foi utilizada para verificar a localização de preditas ORFs ("open reading frames") nas sequências analisadas. O programa "Oligo Analyser" (www.idtdna.com/analyzer/Applications/OligoAnalyzer) foi utilizado para buscar entre as sequências motivos reverso complementares. O sítio da internet "PLACE" (Database of Plant Cis-acting Regulatory DNA Elements – www.dna.affrc.go.jp/PLACE) foi utilizado para verificar a presença de possíveis motivos regulatórios nas sequências a montante do potencial códon de início.

III.13.1.12. Validação do "TE221-RP"

O ensaio de PCR-Inversa permitiu a obtenção de sequências genômicas não transcritas 5' e 3' relativas às extremidades do cDNA TE221. A reconstrução parcial de um elemento foi montada por sobreposição da sequência do cDNA TE221 e das duas regiões genômicas flangeadoras, formando a quimera TE221-RP. O TE221-RP foi validado através de BLASTX de sua sequência de nucleotídeos no banco de dados do genoma do arroz (www.tigr.org) e do "Swissprot" (www.expasy.org).

III.13.2. Ensaio de PCR sobre o DNA Genômico de Cana-de-açúcar com Iniciadores Desenhados Sobre as TIRs Relativas ao TE221

Este ensaio foi planejado com o intuito de amplificar do genoma da cana-de-açúcar variantes da hipotética cópia genômica relativa ao TE221, que compartilham as mesmas TIRs. Para isto foram desenhados iniciadores sobre as sequências das TIRs relativas ao TE221, direcionados para dentro de modo a amplificar outras sequências que também possuam estes iniciadores em suas extremidades.

III.13.2.1. Amplificações por PCR, Subclonagem e Sequenciamento

Foram realizadas dois tipos de PCR, uma com a enzima Taq DNA Polimerase e um tempo de extensão curto, e outra com a enzima Elongase e tempo de extensão longo, de modo a dar preferência para a amplificação de fragmentos curtos e longos, respectivamente. Para ambas as PCRs foram utilizados os iniciadores “MITE-hAT-1F” e “MITE-hAT-1R”, desenhados sobre as sequências das TIRs relativas ao TE221. As sequências destes iniciadores estão descritas na tabela III.

Tabela III – Sequência dos iniciadores utilizados nas PCRs que amplificaram elementos do genoma da cana-de-açúcar que compartilham as mesmas TIRs relativas ao TE221.

iniciador	sequência 5' → 3'
MITE-hAT-1F	ATA AAT GAT AAG TGA TAA CC
MITE-hAT-1R	ACA GAT GAT CAG CAA TAA TC

As reações de PCR com “Taq DNA Polimerase Recombinante” (Invitrogen) consistiram de 80 ng de DNA genômico da variedade de cana SP-80-3280 como molde; PCR Buffer para 1X; 200µM de dNTPs; 1mM de MgCl₂; 10pmoles de cada iniciador; 5U da enzima Taq DNA Polymerase Recombinante e água deionizada para completar um volume final de 50µL. Os

tubos foram levados ao termociclador, passando pelo seguinte ciclo: desnaturação inicial de 2 minutos a 94°C, seguida de 35 ciclos de 94°C por 30 segundos, 55°C por 30 segundos e 72°C por 2 minutos, e tempo complementar de extensão a 72°C por 10 minutos.

As reações de PCR com Elongase (Invitrogen) consistiram de 80 ng de DNA genômico da variedade de cana SP-80-3280 como molde; 0,2 mM do dNTP mix, 10pmoles de cada iniciador, tampão A para 1X e tampão B para 1X, 2,0µL do “Elongase Enzyme Mix” e água deionizada para um volume final de 50,0 µL de reação. As reações foram submetidas ao seguinte termociclo: 94°C por 2 minutos, 35 vezes a repetição do ciclo de 94°C por 30 segundos, 55°C por 30 segundos e 68°C por 4 minutos, e um período complementar de extensão de 10 minutos a 68°C.

Os produtos de ambas as amplificações foram purificados utilizando-se o “GFX™ PCR DNA and Gel Band Purification Kit” (Amersham), de acordo com as instruções do fabricante, e ressuspensos em 30µL de água deionizada. Os produtos purificados da PCR realizada com Taq DNA Polimerase foram subclonados no vetor pGEM-T-Easy (Promega), e os produtos purificados de PCR realizada com Elongase foram subclonados no vetor pMOSBlue (G.E.), ambos de acordo com as instruções do fabricante. Bactérias competentes foram transformadas com estas ligações pelo método do choque térmico, como descrito no item III.6.2. Os clones bacterianos, selecionados com ampicilina, foram colocados para crescer em inóculos, dos quais foram feitas minipreparações de DNA plasmidial, como descrito no item III.7.1. O sequenciamento destas minipreparações foi realizado como descrito no item III.7.4. As sequências foram analisadas através de BLASTN e BLASTX no Genbank e no TIGR (www.tigr.org).

III.13.2.2. Hibridação molecular de SChAT1 e SChAT2 com Sonda do Grupo II

Duas bandas foram amplificadas na reação de PCR descrita no item III.13.2.1, com os iniciadores “MITE-hAT”. Uma apresentou 4,2 kb, sendo denominada SChAT1, e a outra, de 3,5 kb, SChAT2. Estas foram hibridadas

com a sonda do TE191 (grupo II) a fim de verificar se representam realmente a recuperação de um elemento do tipo hAT do grupo II do genoma da cana. As respectivas minipreparações de DNA plasmidial com estas bandas subclonadas no pMOSBlue foram hidrolisadas com dois conjuntos de endonucleases de restrição: *EcoRI* / *PstI* que liberam o fragmento do vetor, e *EcoRI* / *HindIII*, que liberam o fragmento e o cortam em pelo menos um sítio como predito pela análise de restrição da sequência do cDNA TE221. Foram utilizadas endonucleases da “New England Biolabs”, de acordo com as instruções do fabricante. Os produtos da hidrólise foram migrados em gel de agarose e transferidos para membrana de náilon como descrito no item III.11.1.4. Foi utilizada a sonda interna do TE191 (figura 6-A), que também está presente na sequência do TE221. Estes dois cDNAs apresentam alta identidade de sequência de nucleotídeos. O ensaio foi realizado como descrito no item III.11.1.

IV. RESULTADOS

IV.1. Obtenção das Sequências Completas dos Clones de cDNA e Identificação das Famílias de Elementos Transponíveis Expressas em Cana-de-açúcar

Entre os clones de cDNA gerados pelo projeto SUCEST, 276 já haviam sido previamente identificados como homólogos a elementos de transposição (Rossi *et al*, 2001). Porém, as sequências destes clones eram apenas parciais, visto que o objetivo do projeto era sequenciar pequenas porções do cDNA denominadas “etiquetas” (em inglês, “tags”). De um total de 276 clones, foram obtidas 156 sequências completas com qualidade mínima de 20 avaliada no programa computacional Phred, que representam aproximadamente 56% da coleção original. O sequenciamento de nucleotídeos dos clones de cDNA homólogos a elementos de transposição foi dividido por categorias. Clones de tamanho até 1kb foram sequenciados utilizando-se os iniciadores “T7” e “SP6”, ancorados no vetor pSPORT1 em que os cDNAs foram subclonados, enquanto os clones de tamanho entre 1kb e 3,5kb foram sequenciados com uma série de iniciadores desenhados sobre partes de suas sequências, de modo a caminhar por toda sua extensão (item III.8.3). Houve ainda clones maiores que 3,5kb, dos quais foram feitas sub-bibliotecas para o seu sequenciamento (item III.8.4). A maioria dos clones apresentou tamanho até 2kb, enquanto pouco mais de 22% deles apresentaram tamanho entre 2kb e 4kb, sugerindo a existência de transcritos inteiros. A tabela IV ilustra a distribuição dos clones por faixas de tamanho. Estes dados coincidiram com as estimativas de tamanho realizadas antes do início do sequenciamento (item III.8.1).

Tabela IV – Distribuição dos clones de cDNA completamente sequenciados por faixas de tamanho.

tamanho (Kpb)	número de clones	porcentagem do total
0 - 1	30	19,3 %
1 - 2	91	58,3 %
2 - 3	28	17,9 %
3 - 4	07	4,5 %
total	156	100 %

A análise das sequências completas dos cDNAs permitiu confirmar a similaridade destes com elementos descritos na literatura. Por comparação de sequências com o banco de dados do Genbank (BLASTX e BLASTP), foram identificados cDNAs homólogos a 11 famílias de retrotransposons (classe I), e 9 famílias de transposons (classe II), entre as quais estavam representadas as famílias mais estudadas, como Tnt1, Ac e MuDr. Foi estabelecido como critério para adotar o elemento de referência a primeira resposta do BLAST que indicasse um elemento de transposição já descrito na literatura, com uma similaridade de no mínimo 35%, sobre uma cobertura de no mínimo 50% da sequência de aminoácidos do cDNA. A distribuição dos clones de cDNA, de acordo com as famílias de elementos com as quais apresentaram homologia, está apresentada na tabela V. Nota-se que a família de retrotransposons com o maior número de homólogos encontrados em cana é Hopscotch, com 25 cDNAs homólogos. Entre os transposons a família com o maior número de homólogos encontrados foi MuDr, com 43 representantes, seguida das famílias Ac e Tam3 pertencentes à superfamília *hAT*, com 22 e 10 representantes respectivamente, perfazendo um total de 32 homólogos a *hAT* na cana. Nesta tabela é possível observar que alguns grupos de elementos provavelmente são cDNAs completos, por apresentarem tamanho extenso. Entre eles podemos destacar Hopscotch (3857pb), Tnt1 (3823pb) e CentA (3860pb) como representantes de retrotransposons, e Ac (2526pb) e MuDr (2569pb) como representantes dos transposons.

Tabela V – Identificação das famílias de elementos transponíveis expressas em cana-de-açúcar, através da análise da identidade/similaridade das sequências dos clones de cDNA com elementos descritos na literatura.

classe	família (elemento de referência)	número de homólogos de cana-de-açúcar	clone de maior tamanho (pb)	bibliotecas do SUCEST representadas
retrotransposons (classe I)	CentA	01	3860	LR
	Dea-1	01	2616	CL
	Hopscotch	25	3857	AD, AM, CL, FL, HR, LB, RT, SB, ST
	MAGGY	08	2618	CL, FL, ST
	Opie	08	3116	CL, FL, LV
	Retrofit	01	733	FL
	RIRE	03	1976	FL, ST
	Stonor	08	2615	AM, FL, RZ, SD
	Tekay	01	3230	ST
	Tnt1	05	3823	AM, CL, FL, LB, LR, RZ
	Tto	01	1219	AM
transposons (classe II)	Ac	22	2526	AD, AM, CL, FL, LB, LR, SD
	Candystripe	05	1650	FL,HR,LB
	En-1	02	3924	AD, LB
	MuDr	43	2569	AM, CL, FL, HR, LB, LR, SD, SB, ST, RT, RZ
	Soymar	01	1719	AM
	Tag1	07	2515	AM, CL, FL, LB, RT, ST
	Tam3	10	2445	AD, AM, FL, LR, RT, RZ
	Tip100	03	1676	AM, FL, RT
	TNP2	01	1888	AM
total	-	156	-	-

Estão indicados na tabela os elementos de referência e sua classe (I ou II); o número de cDNAs da cana homólogos a cada família do elemento de referência; o clone de maior tamanho de cada família e os tecidos dos quais foram feitas as bibliotecas do SUCEST que originaram os clones pertencentes a cada família. Bibliotecas do SUCEST: **AD**: mistura de tecidos infectados com *G. diazotrophicans*, **AM**: meristema apical, **CL**: calo, **FL**: flor, **HR**: mistura de tecidos infectados com *H. rubrisubalbicans*, **LB**: brotos laterais, **LR**: folha, **RT**: raiz de plantas maduras, **RZ**: raiz de plantas jovens, **SB**: casca do caule, **SD**: sementes, **ST**: caule.

Uma segunda análise foi realizada sobre o conjunto de cDNAs, de modo a identificar aquelas sequências passíveis de representar um elemento ativo transcricionalmente, e eliminar de maneira bastante estrigente possíveis transcritos quiméricos. Os transcritos quiméricos são gerados por “read through”, ou seja, pela transcrição de um gene vizinho que avançou pela

sequência de um TE. Deste modo, o transcrito apresenta parte de sua sequência homóloga a um TE, mas não provém de uma atividade transcricional do TE em si, mas do gene vizinho. Os dados referentes a esta segunda análise foram incluídos como resultados na publicação “Transcriptionally active transposable elements in recent hybrid sugarcane” Plant Journal, dec 2005, 44(5):707-17 (Araújo *et al*, 2005) (apêndice 3). A tabela VI (Plant Journal) lista estes cDNAs, indicando o tecido da biblioteca de origem, a cobertura da sequência sobre o TE de referência, a presença de mudança de quadro de leitura (“frameshift”) ou de códon de parada (“stop códon”) prematuro, e se trata-se de um cDNA potencialmente completo. Pode-se notar que a maioria dos cDNAs apresenta-se incompleta em sua porção 5’ do transcrito, e a grande maioria neste conjunto não apresenta alteração de quadro de leitura, assim como também não apresenta códons de parada prematuros. Os cDNAs incompletos a 5’ podem ser produto de uma transcrição reversa incompleta durante sua síntese. Entre os cDNAs potencialmente completos referentes aos transposons estão os homólogos a Ac TE048 (2148pb), TE096 (1828pb) e TE221 (2078pb); o TE191(2179pb) homólogo a Tam3; o TE109 (2348pb) e o TE148 (2349pb) homólogos a MuDr. Entre os retrotransposons são listados como cDNAs potencialmente completos o TE028 (2915pb) homólogo a Hopscotch e o TE159 (3219pb) homólogo a Tekay. Quanto aos tecidos que originaram as bibliotecas de transcritos, os que apresentaram maior número de elementos expressos foram flores (FL - 14 cDNAs), calos, sementes e raízes de plantas maduras (CL, SD e RT, respectivamente - 9 cDNAs cada). Os tecidos em que se encontrou o menor número de transcritos foram caule, raízes de plantas jovens e mistura de tecidos infectados com *H. rubrisubalbicans* (ST, RZ e HR, respectivamente - 2 cDNAs cada). Estes dados, referentes aos tecidos de origem dos transcritos, corroboram em parte os dados do ensaio de hibridação de RNA sobre arranjo de DNA (“macroarray”) apresentados no mesmo artigo (Araújo *et al*, 2005). Analisando-se em particular os dados referentes aos homólogos à superfamília *hAT* neste ensaio, pode-se notar que os tecidos em que se constatou a expressão do maior número de TEs foram calo (CL) e flores (FL). Dos 10 transcritos investigados, todos foram detectados no RNA de calos, com exceção apenas do TE007. No RNA de flores foram detectados 5 transcritos (figura 9).

Tabela VI – Características dos cDNAs homólogos a TEs sequenciados selecionados a partir do transcriptoma de cana-de-açúcar.

elemento de referência	nome do clone	tecido	tamanho (pb)	“frameshift”/ “stop codon”	similaridade com elemento de referência (aa) (%)	alinhamento com o elemento de referência	transposase ou poliproteína hipoteticamente completa
Mutator (823 aa) (gi 2130141)	TE019	SD	1734	não/não	43	(1nt)230–714(305nt)	cDNA incompleto
	TE029	SD	788	não/não	40	(16nt)468–713(82nt)	cDNA incompleto
	TE037	LR	1128	sim/sim	72	(3nt)364–643(273nt)	cDNA incompleto
	TE047	CL	1679	não/sim	59	(29nt)319–710(507nt)	cDNA incompleto
	TE059	FL	1063	não/não	44	(32nt)444–640(461nt)	cDNA incompleto
	TE076	SD	1701	não/não	43	(8nt)231–714(265nt)	cDNA incompleto
	TE084	SB	1373	não/não	43	(2nt)318–714(198nt)	cDNA incompleto
	TE085	SD	645	não/não	40	(5nt)508–713(67nt)	cDNA incompleto
	TE101	RT	1213	não/não	40	(4nt)395–713(264nt)	cDNA incompleto
	TE109	AM	2348	não/não	42	(29nt)128–704(520nt)	sim (70%)
	TE120	SB	1552	não/não	43	(9nt)289–714(289nt)	cDNA incompleto
	TE121	SD	970	não/não	42	(37nt)456–714(201nt)	cDNA incompleto
	TE138	RZ	1102	não/não	49	(1nt)540–704(633 nt)	cDNA incompleto
	TE142	SB	1596	sim/sim	67	(12nt)343–719(386nt)	cDNA incompleto
	TE148	LR	2349	não/não	42	(154 nt)133–714(48nt)	sim (70%)
	TE161	RT	1361	não/não	39	(53nt)430–714(501nt)	cDNA incompleto
	TE165	FL	1948	sim/sim	65	(11nt)265–668(729nt)	cDNA incompleto
	TE172	FL	1176	não/não	44	(5nt)448–666(538nt)	cDNA incompleto
	TE177	AD	1651	sim/sim	66	(3nt)338–721(428nt)	cDNA incompleto
	TE187	HR	1293	não/não	45	(34nt)435–666(587nt)	cDNA incompleto
	TE194	CL	851	não/sim	60	(12nt)456–715(2nt)	cDNA incompleto
	TE227	SD	703	não/não	41	(36nt)546–714(244nt)	cDNA incompleto
	TE228	SB	1701	não/não	43	(8nt)231–714(265nt)	cDNA incompleto
	TE263	SD	965	não/não	41	(50nt)468–713(201nt)	cDNA incompleto
	TE266	CL	1268	não/não	41	(29nt)385–713(290nt)	cDNA incompleto
	TE274	LB	1213	não/não	42	(1nt)374–714(201nt)	cDNA incompleto
	TE279	ST	689	não/não	41	(56nt)546–714(210nt)	cDNA incompleto
Activator (807 aa) (gi 22113)	TE007	LB	372	sim/não	62	(0nt)159–573(0nt)	cDNA incompleto
	TE017	FL	1229	não/não	46	(24nt)450–754(257nt)	cDNA incompleto
	TE046	CL	1049	não/não	81	(20nt)158–438(183nt)	cDNA incompleto
	TE048	CL	2148	não/não	49	(188 nt)223–762(316 nt)	sim (67%)
	TE096	LR	1828	não/não	48	(1nt)259–762(291nt)	sim (62%)
	TE203	SD	1005	não/não	45	(11nt)512–754(226nt)	cDNA incompleto
	TE207	AD	1435	não/não	51	(1nt)397–762(312nt)	cDNA incompleto
	TE221	FL	2078	sim/sim	48	(9nt)184–774(322nt)	sim (73%)
Tam3 (749aa) (gi 16064)	TE144	AM	1760	sim/não	51	(27nt)300–684(491nt)	cDNA incompleto
	TE191	AD	2179	não/sim	50	(44nt)238–712(689nt)	sim (63%)
Tip100 (808aa) (gi 4063770)	TE153	RT	989	não/não	61	(3nt)538–805(185nt)	cDNA incompleto
Hopscotch (1439 aa) (gi 7444442)	TE010	FL	1871	não/não	68	(154nt)925–1425(220nt)	cDNA incompleto
	TE028	RT	2915	sim/não	66	(4nt)540–1425(259nt)	sim (62%)
	TE043	CL	1689	não/não	69	(32nt)964–1425(277nt)	cDNA incompleto
	TE070	RT	1502	não/não	68	(14nt)1008–1425(240nt)	cDNA incompleto
	TE083	RT	2464	sim/não	55	(2nt)674–1425(217nt)	cDNA incompleto
	TE098	RT	1750	não/não	66	(2nt)927–1425(251nt)	cDNA incompleto
	TE174	HR	1532	sim/não	65	992–1425(231nt)	cDNA incompleto
	TE176	ST	707	não/não	67	(4nt)1264–1425(217nt)	cDNA incompleto

	TE186	FL	958	não/não	60	(133nt)1235–1425(252nt)	cDNA incompleto
	TE197	FL	1763	não/não	60	(18nt)931–1425(233nt)	cDNA incompleto
	TE250	RT	1815	não/não	62	(2nt)600–1167(1nt)	cDNA incompleto
	TE273	LB	1326	não/não	64	(4nt)1069–1425(251nt)	cDNA incompleto
	TE276	RT	1717	não/não	76	(569nt)896–1276(2nt)	cDNA incompleto
MAGGY (pol. 1199 aa) (gi 522302)	TE027	LR	1664	não/não	52	(123nt)747–1168(280nt)	cDNA incompleto
	TE053	CL	2607	não/não	50	(39nt)394–1095(594nt)	cDNA incompleto
	TE219	FL	1598	não/não	50	861–1168(680nt)	cDNA incompleto
	TE269	CL	2605	sim/não	50	(39nt)394–1095(593nt)	cDNA incompleto
Opie (pol. 1068 aa) (gi 7489803)	TE018	LV	2222	não/não	61	(142nt)4–689(4nt)	cDNA incompleto
	TE024	FL	1509	não/não	82	(37nt)768–1066(575nt)	cDNA incompleto
	TE057	FL	1119	não/não	83	(1nt)574–944(2nt)	cDNA incompleto
	TE182	FL	663	não/não	80	(1nt)725–944(2nt)	cDNA incompleto
	TE245	FL	933	não/não	69	(1nt)645–891(191nt)	cDNA incompleto
Stonor4 (1406 aa) (gi 7489738)	TE009	RZ	541	sim/não	53	(28nt)193–395(2nt)	cDNA incompleto
	TE040	AM	1043	sim/não	61	(130nt)58–397(nt)	cDNA incompleto
	TE106	SD	733	sim/não	59	(4nt)135–412(2nt)	cDNA incompleto
	TE258	AM	1444	sim/não	61	(535nt)58–395(2nt)	cDNA incompleto
	TE270	FL	1890	sim/sim	69	(2nt)372–1019(2nt)	cDNA incompleto
Tekay (1470 aa) (gi 18092337)	TE159	ST	3219	não/sim	74	(2nt)560–1468(520nt)	sim (62%)
Tnt1 (1328aa) (gi 130582)	TE049	CL	2718	não/não	55	(2nt)550–1325(307nt)	cDNA incompleto
Tto1 (1338 aa) (gi 7444414)	TE002	AM	1198	sim/sim	72	(2nt)959–1328(91nt)	cDNA incompleto

Estão indicados na tabela os elementos de referência utilizados na classificação dos cDNAs; o nome atribuído a cada clone de cDNA; os tecidos dos quais foram feitas as bibliotecas do SUCEST que originaram os clones; o tamanho dos clones; a presença de mudança de quadro de leitura e de códon de parada prematuro nas sequências dos cDNAs; a similaridade (BLASTX) com o elemento de referência; uma descrição dos alinhamentos (BLASTX) com o respectivo elemento de referência e se o cDNA é considerado hipoteticamente completo, ou seja, se cobrem pelo menos 60% da proteína do elemento de referência e contém códons de início e parada a 5' e 3' respectivamente. Bibliotecas do SUCEST: **AD**: mistura de tecidos infectados com *G. diazotrophicans*, **AM**: meristema apical, **CL**: calo, **FL**: flor, **HR**: mistura de tecidos infectados com *H. rubrisubalbicans*, **LB**: brotos laterais, **LR**: folha, **RT**: raiz de plantas maduras, **RZ**: raiz de plantas jovens, **SB**: casca do caule, **SD**: sementes, **ST**: caule.

A obtenção das sequências completas dos clones de cDNA e a sua anotação abriram perspectivas para diversos projetos, e neste momento tomou-se a decisão de dividir as sequências para estudo mais aprofundado. As sequências homólogas ao transposon MuDr foram utilizadas num estudo que gerou a publicação “Comparative Analysis of Mutator-like Transposases in Sugarcane” Mol. Genet. Genomics, sept 2004, 272(2):194-203 (Rossi *et al*, 2004) (apêndice 2).

IV.2. A superfamília *hAT*

Em função da diversidade de elementos expressos encontrada no banco de TEs da cana-de-açúcar, decidiu-se eleger uma das famílias de transposons para dar prosseguimento aos estudos de caracterização. Elementos similares aos membros da superfamília *hAT* representam o segundo grupo de transcritos de transposons mais abundantes. Os transposons Ac e Tam3, por outro lado, estão entre os TEs melhor estudados em sistemas vegetais. Com base nas observações acima selecionou-se o subconjunto de transcritos similares à superfamília *hAT* como objeto de aprofundamento de estudo. A tabela VII lista os clones que apresentam similaridade com domínios protéicos que caracterizam os elementos *hAT*, por elemento homólogo em ordem crescente. Estão destacados em cinza aqueles cDNAs contidos na tabela VI, anteriormente apresentados. Os demais transcritos são considerados potencialmente quiméricos.

Entre os clones de maior tamanho homólogos ao elemento Ac destacam-se o TE048 de 2148pb e o TE124 de 2526pb. Ambos não apresentam mudança do quadro de leitura ou códons de parada prematuros em suas sequências, e podem representar cDNAs completos. Quanto aos homólogos a Tam3, o TE257 tem 2440pb e também apresenta as características citadas acima, podendo representar um cDNA completo.

Tabela VII – Caracterização de cada um dos cDNAs da cana-de-açúcar homólogos à superfamília *hAT*.

Nome do clone	tamanho (pb)	elemento homólogo (BLASTX)	similaridade BLASTX (%)	biblioteca SUCEST	Mudança de quadro de leitura/códon de parada prematuro
TE001	1844	Ac	51	AM	não/não
TE007	372	Ac	62	LB	sim/não
TE016	1105	Ac	47	FL	não/não
TE017	1229	Ac	46	FL	não/não
TE032	1367	Ac	47	FL	sim/sim
TE035	1163	Ac	45	LB	não/não
TE045	1050	Ac	81	CL	não/não
TE046	1049	Ac	81	CL	não/não
TE048	2148	Ac	49	CL	não/não
TE074	1017	Ac	45	SD	não/não
TE096	1828	Ac	48	LR	não/não
TE097	647	Ac	56	LR	não/não
TE124	2526	Ac	49	AD	não/não
TE132	1421	Ac	49	AM	não/não
TE191	2179	Ac	50	AD	não/sim
TE203	1005	Ac	45	SD	não/não
TE207	1435	Ac	51	AD	não/não
TE217	1269	Ac	48	FL	não/não
TE221	2078	Ac	48	FL	sim/sim
TE223	1398	Ac	47	FL	não/não
TE243	1142	Ac	51	AM	sim/não
TE265	1458	Ac	50	AM	não/não
TE011	650	Tam3	58	FL	sim/não
TE112	659	Tam3	59	FL	não/sim
TE116	1134	Tam3	48	RT	não/não
TE144	1760	Tam3	51	AM	sim/não
TE154	1934	Tam3	50	SB	não/não
TE170	1207	Tam3	43	FL	não/não
TE198	797	Tam3	49	FL	não/não
TE202	710	Tam3	64	RZ	não/não
TE257	2440	Tam3	47	AM	não/não
TE277	673	Tam3	64	RZ	não/não

As informações incluem tamanho das sequências; elementos com os quais apresentaram maior similaridade no BLASTX, assim como a porcentagem desta similaridade; a biblioteca de origem de cada cDNA; a presença de mudança de quadro de leitura e de códon de parada prematuro nas sequências dos cDNAs. Os clones destacados em cinza estão presentes na tabela curada que integrou o artigo do “Plant Journal”. Bibliotecas do SUCEST: **AD**: mistura de tecidos infectados com *G. diazotrophicans*, **AM**: meristema apical, **CL**: calo, **FL**: flor, **HR**: mistura de tecidos infectados com *H. rubrisubalbicans*, **LB**: brotos laterais, **LR**: folha, **RT**: raiz de plantas maduras, **RZ**: raiz de plantas jovens, **SB**: casca do caule, **SD**: sementes, **ST**: caule.

A figura 7 traz um diagrama esquematizando o alinhamento dos cDNAs homólogos a *hAT* com seus respectivos elementos de referência. O diagrama traz como âncoras destes alinhamentos as transposases de Ac e de Tam3, e apresenta as regiões de similaridade. Pode-se notar que a maioria dos cDNAs apresenta-se incompleta em sua porção 5', refletindo o que se observou para o conjunto completo de todos os cDNAs. Consequentemente, a similaridade entre os cDNAs e os elementos de referência encontra-se localizada na porção carboxi-terminal da transposase.

Entre os clones, alguns apresentaram similaridade dentro de um segmento que se estendeu por 60% ou mais da sequência de aminoácidos da transposase de seu elemento de referência, e podem representar cDNAs completos. Entre os homólogos ao elemento Ac (figura 7-A) podemos destacar: o TE048, de 2148pb, que apresentou similaridade de 49% (293 aminoácidos em 551) com a transposase de Ac; o TE124, de 2526pb, que apresentou similaridade de 49% (325 aminoácidos em 669); o TE191, de 2179pb, que apresentou similaridade de 50% (254 aminoácidos em 494); e o TE221, de 2078pb, que apresentou similaridade de 48% (262 aminoácidos em 536). Entre os homólogos ao elemento Tam3 (figura 7-B), podem representar cDNAs completos o TE144, de 1760pb, que apresentou similaridade de 51% (187 aminoácidos em 469) com a transposase de Tam3; TE154, de 1934pb, que apresentou similaridade de 50% (236 aminoácidos em 467); e o TE257, de 2440pb, que apresentou similaridade de 47% (298 aminoácidos em 618).

Figura 7 – Diagramas esquematizando o alinhamento dos cDNAs homólogos a *hAT* com a transposase de seus respectivos elementos de referência. As linhas negras representam a extensão dos cDNAs inteiros. Os números ao lado dos nomes dos cDNAs indicam o tamanho de suas sequências de aminoácidos. As caixas azuis representam as regiões dos cDNAs que apresentam similaridade com a transposase do elemento de referência, e os números abaixo das caixas indicam a posição destas regiões dentro da transposase do elemento de referência. **A:** Alinhamento dos cDNAs homólogos ao elemento Ac. **B:** Alinhamento dos cDNAs homólogos ao elemento Tam3. Os segmentos destacados, que compreendem os resíduos 555 a 594 e 682 a 720 da transposase de Ac, e os resíduos 517 a 557 e 643 a 681 da transposase de Tam3, foram utilizados posteriormente para um alinhamento.

IV.2.1. A superfamília *hAT* é Constituída por Dois Grupos de Sequências

As transposases codificadas por elementos do tipo *hAT* apresentam domínios protéicos típicos, ambos localizados na porção carboxi-terminal das transposases (Kempken *et al*, 2001). Um deles está relacionado com a capacidade de dimerização (Pfam PF05699), e compreende os resíduos 682 a 720 da transposase de Ac, e 643 a 681 da transposase de Tam3. O outro domínio conservado, de função ainda não descrita, compreende os resíduos 555 a 594 da transposase de Ac e 517 a 557 da transposase de Tam3. Com o

intuito de analisar as relações de similaridade entre as sequências dos cDNAs nestes dois domínios e o seu padrão de agrupamento, foram realizados um alinhamento múltiplo e uma árvore filogenética, apresentados na figura 8. Os alinhamentos incluíram sequências de arroz, milho, arabidopsis e humano, de modo a melhor avaliar as relações de similaridade. Estas sequências foram selecionadas a partir do banco de dados Genbank através de BLASTN e BLASTX contra os elementos Ac e Tam3. Seus respectivos números de acesso são identificados na tabela VIII, juntamente com a informação sobre o organismo de origem.

Tabela VIII – Identificação das sequências de aminoácidos relativas a dois domínios conservados da superfamília *hAT* de outros organismos, incluídas no alinhamento das sequências de cana-de-açúcar.

nome da sequência	número de acesso no Genbank	organismo de origem
Ri_1	AC134929.2	<i>Oryza sativa</i> (Arroz)
Ri_2	AK067574	
Ri_3	AK102230	
Ri_4	AK105085	
Ri_5	AK121128	
Ri_6	BAC10751	
Ma_1	AY111033	<i>Zea mays</i> (milho)
At_1	NM_106651	<i>Arabidopsis thaliana</i>
At_2	AL137079	
At_3	AC009322	
At_4	AC011717	
At_5	AAD24567	
Hu_1	CAA76545	<i>Homo sapiens</i> (humano)
Ac	CAA29005	<i>Zea mays</i> (milho)
Tam3	CAA38906	<i>Antirrhinum majus</i>
hobo	P12258	<i>Drosophila melanogaster</i>

Os transcritos que apresentam estes domínios foram alinhados usando o programa ClustalX, juntamente com os representantes da família Ac, Tam3 e hobo, e as sequências de arroz (Ri), arabidopsis (At), milho (Mi) e humano (Hu). Os alinhamentos múltiplos, bem como a árvore apresentados na figura 8 revelam a divisão dos clones de cDNA em dois grupos distintos. O grupo I é constituído por sequências de elevada identidade na região dos domínios conservados, enquanto o grupo II, onde agruparam-se os ícones da superfamília *hAT*, os transposons Ac, Tam3 e hobo, apresenta-se menos conservado.

A transposase hAT-like

B

TE124 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 Ma_1 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE096 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE048 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE265 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE207 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE017 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE203 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE074 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE016 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE223 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE035 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE217 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 TE132 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 R1_1 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 R1_2 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 R1_4 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 R1_5 FSSIADDMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 R1_3 FRSTAKIMHERFDKYW-KDCSLVLAIAVVMDF--RPFMK-LVEF
 At_2 VTGIATMOKVVDKYW-RDCSLVLAMAVVMDF--RPFMK-LVEF
 At_1 VSNLANTLKDKFDEYW-RECFLLMMAVAVVMDF--RPFMK-LVEF
 At_3 VSNLANTLKDKFDEYW-RECFLLMMAVAVVMDF--RPFMK-LVEF
 At_4 VSNLANTLKDKFDEYW-RECFLLMMAVAVVMDF--RPFMK-LVEF
 TE001 LNLEAEAMYTKFQKYW-AAPSMVLLIAAVLDF--SMKAD-FVRF
 TE257 LNLEAEAMYTKFQKYW-AAPSMVLLIAAVLDF--SMKAD-FVRF
 TE154 LNLEAEAMYTKFQKYW-AAPSMVLLIAAVLDF--SMKAD-FVRF
 At_5 IREMVEDMTKDYKYW-EDFDILAMAVALDF--RLKFS-ALEY
 Ac IRRMAVAMSEKFEKYW-KVSNIALAVACFLDF--RYKKI-LIEF
 R1_6 VAQVADRMLTKFEGYW-DLRFVMAFASILD--RYKKI-SVEY
 TE191 IAAMYSKMQEKFDKYW-MISYFTNCVPVILDF--QPKFG-FIEF
 TE221 IAAMYSKMQEKFDKYW-MISYFTNCVPVILDF--QPKFG-FIEF
 TE116 IELMSFEMIEKFDKYW-KEIQPMGLATILD--RPFKD-FLVG
 TE243 LMSMAAMVDKFDKYW-EKKNNVMVATILD--RPFKM-YIKW
 Tam3 LKNIIKEMKKVWVYV-KKVPNVCLSSCLDF--RVRLIGTLEL
 TE170 FRNIAVPMKLKLLKYW-EKTPLLYSYAPILDF--RGKLGFLS-
 Hu_1 KEVIAKELSKTQET--FEIDMFLNVAIFLDF--RYKRLPFLSA
 HoBo AALLKERILNIRKIWMANLSIWHKAAPFLYPAAHLEEDILEI
 1.....10.....20.....30.....40.....

Domínio conservado

C

TE096 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE124 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE048 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE265 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE017 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE203 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE074 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE016 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE223 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE035 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE217 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE132 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE207 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_1 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_2 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_4 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_5 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 Ma_1 IQEFDILNWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_3 IQDFEILEWKKLNTIKFPTLSKMDILAIIPMSMVSSGS
 Tam3 DKDFDILKWWKLNITVKFPTLSKMDILAIIPMSMVSSGS
 Ac SCQFDILSWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 At_1 SQDFEVLGWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 At_3 SQDFEVLGWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 At_4 SQDFEVLGWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 HoBo SQNFEVIEWWKKLNITVKFPTLSKMDILAIIPMSMVSSGS
 At_2 VQDFEVLGWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 R1_6 LDNFDILGWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 TE001 TENFNVLAWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE154 TENFNVLAWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE257 TENFNVLAWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE116 TENFKILDWKKVAGTRYPLRRIARDIFAIPTVTVSSGS
 At_5 FRDMVDVIAWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE243 NKHFNLLRWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE191 DESFDILDWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE221 DESFDILDWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 TE170 DESFDILDWKKNSDAYPILSLMARDFLSIPVSTVSSGS
 Hu_1 GLNEDPLKWSKLNITVKFPTLSKMDILAIIPMSMVSSGS
 1.....10.....20.....30.....40.....

Domínio de dimerização
 (acesso Pfam PF05699)

Figura 8 – Alinhamentos múltiplos, através do ClustalX, de dois domínios conservados característicos das transposases homólogas a *hAT*. **A:** Representação esquemática de uma transposase homóloga ao elemento Ac. Em vermelho, próximo à extremidade carboxi-terminal da proteína, estão representados os dois domínios conservados. Os resíduos relativos a estes domínios estão baseados em suas posições na sequência da transposase de Ac (ORFa). **B:** Alinhamento de um domínio conservado de função ainda não descrita, localizado entre os resíduos 555 a 594 da transposase de Ac. **C:** Alinhamento do domínio de dimerização da superfamília *hAT*, localizado entre os resíduos 682 a 720 da transposase de Ac. Além das sequências de cana-de-açúcar foram incluídas no alinhamento sequências de milho (Ma_1), arroz (Ri_1 a Ri_6), *Arabidopsis thaliana* (At_1 a At_5) e humano (Hu_1), baixadas do Genbank e cujos números de acesso estão discriminados na tabela VIII (Materiais e Métodos). **D:** Árvore filogenética gerada a partir do alinhamento das sequências dos domínios apresentados em B e C, concatenados. Foi utilizado algoritmo de “Neighbor Joining”, com teste de “bootstrap” de 500 réplicas. A seta cinza indica a separação das sequências do grupo I com uma sustentação de 70%.

IV.2.2. A perfil de expressão da superfamília *hAT* em membrana de macroarray

A partir do macroarranjo apresentado em Araújo *et al* (2005) realizou-se uma análise pormenorizada do universo de expressão de cDNAs homólogos a *hAT*. A figura 9 apresenta o detalhe da membrana do array onde pode-se verificar claramente que a maior frequência dos transcritos se dá em calo (CL), onde 9 dos 10 transcritos foram detectados, seguido de flores (FL) e meristema apical (AM), em que 5 transcritos foram detectados. Apenas os transcritos TE017 e TE203 foram detectados em todos os quatro tecidos analisados. Ambos, assim como o TE046, o TE048, o TE096 e o TE207, agrupam-se com os homólogos do grupo I, enquanto os demais são homólogos ao grupo II. É interessante notar que os homólogos do grupo II apresentam um padrão de expressão mais restrito a calo com exceção do TE007 (unicamente em AM) e TE144 (expresso em todos os tecidos menos na folha).

Figura 9 – Expressão de elementos homólogos a *hAT* por tecido. Resultados de ensaio de “macroarray” (Araújo *et al*, 2005). Os quadrados em cinza indicam detecção da expressão, enquanto os quadrados em preto indicam a não detecção. Os tecidos dos quais foi extraído RNA para a hibridação estão identificados na linha de cima: **CL**: calo; **AM**: meristema apical; **LR**: folha; **FL**: flor.

IV.2.3. Padrão de Hibridação de Sondas dos Grupos I e II em Genomas de Gramíneas

Uma vez que foi verificada a divisão dos cDNAs da cana-de-açúcar homólogos a *hAT* em dois grupos distintos, de acordo com a comparação de suas sequências, foi investigado se este padrão se refletiria também na distribuição genômica e número de cópias dos dois grupos nos genomas de gramíneas. Para isto foi realizado ensaio de hibridação molecular com sondas representantes dos dois grupos. Foram utilizados como sondas cDNAs inteiros: o TE203, de 1005 pb para o grupo I, e o TE001, de 1844 pb para o grupo II. Na hibridação com a sonda do grupo I pode-se notar que houve um padrão de hibridação equivalente com todos os DNAs, sendo que foi um pouco mais intensa nos parentais e híbridos de cana. Já no perfil de hibridação com a sonda do grupo II nota-se uma hibridação muito mais intensa nos parentais e híbridos de cana que em arroz e milho. Fica bem claro que o padrão de distribuição e o número de cópias dos grupos I e II nestes genomas é bastante distinto (figura 10).

Figura 10 – Resultado de ensaio de hibridação molecular de sondas dos grupos I e II com DNA genômico de gramíneas. **A:** Gel e agarose contendo o DNA genômico de: **A:** arroz, **M:** milho, **So:** *S. officinarum*, **Ss:** *S. spontaneum*, **Sc:** SP-89-1115, **Sc:** SP-91-1049, hidrolisados com as endonucleases de restrição *HindIII* e *XbaI*. O DNA foi transferido para membrana de nylon e hibridado contra sondas das classes I e II. **B:** Hibridação utilizando como sonda do grupo I o cDNA TE203 inteiro, de 1005 pb. **C:** Hibridação utilizando como sonda do grupo II o cDNA TE001 inteiro, de 1844 pb.

Deste modo, com as observações recuperadas a partir da árvore, do macroarranjo e da hibridação molecular, foi iniciada uma caracterização dos homólogos de *hAT* em cana de açúcar usando abordagens distintas se os clones são do grupo I ou do grupo II como poderá ser verificado a seguir.

IV.2.4. Estudo dos homólogos a *hAT* do Grupo I

A partir deste resultado, as relações de similaridade entre as sequências dentro de cada grupo em particular foram investigadas. Para esta comparação foi utilizado um segmento que sobrepõe-se ao domínio de dimerização, compreendido entre os resíduos 679 a 712 da transposase de Ac e 640 a 673 da transposase de Tam3, presente na maioria dos cDNAs. Os alinhamentos múltiplos foram realizados através do ClustalX. A partir destes alinhamentos foram geradas árvores filogenéticas com o algoritmo de “neighbor joining” e teste de suporte do tipo “bootstrap” com 500 réplicas. Estas árvores ilustram o padrão de agrupamento das proteínas codificadas pelos cDNAs.

A figura 11 apresenta o alinhamento e a árvore referentes às sequências do grupo I alinhadas com os elementos Ac e Tam3, sequências de bancos de dados de arroz (Ri), e arabidopsis (At), e sequências de alguns cDNAs do grupo II. Foi incluída também no alinhamento a sequência de uma transposase domesticada do tipo *hAT*, denominada DAYSLEEPER (Bundock & Hooykaas, 2005). O alinhamento (figura 11-A) mostra que as sequências do grupo I são idênticas em nível de aminoácidos, o que evidencia sua alta conservação. Estas sequências apresentam identidade de nucleotídeos variando entre 87% e 100%, e podem representar transcritos provenientes de diferentes alelos de uma mesma cópia gênica do elemento, ou transcritos de diferentes cópias gênicas com alta conservação de sequência. Na árvore (figura 11-B), a seta número 1 indica o ponto onde ocorre a separação das sequências do grupo I de cana das sequências de arroz (Ri_1, Ri_2, Ri_4 e Ri_5), com um suporte de 94%. Isto sugere que, dentro do domínio protéico analisado, as sequências do grupo I divergiram após a separação destas duas gramíneas. A seta número 2 indica a separação entre as sequências de arroz e cana (gramíneas) e as

sequências de arabidopsis, com suporte de 79%, o que sugere uma divergência das sequências após a separação entre gramíneas e eudicotiledôneas.

Figura 11 – Padrão de agrupamento dos homólogos à superfamília *hAT*. **A:** Alinhamento de um domínio conservado de transposases do tipo *hAT*, localizado entre os resíduos 679 a 712 da transposase de Ac e 640 a 673 da transposase de Tam3. Estão alinhadas as sequências dos cDNAs do grupo I e do grupo II, dos elementos Ac e Tam3, da transposase domesticada DAYSLEEPER e sequências do tipo *hAT* obtidas no Genbank, de arroz (Ri) e arabidopsis (At). As barras azuis destacam os cDNAs do grupo I. Os asteriscos indicam as posições em que o aminoácido é conservado entre todas as sequências. **B:** Filogenia gerada a partir do alinhamento, utilizando-se o algoritmo de “neighbor joining”, com “bootstrap” de 500 réplicas. O comprimento dos braços da árvore reflete o número de diferenças entre os aminoácidos das sequências. A barra azul destaca os cDNAs do grupo I. A seta de número 1 indica o ponto de separação entre sequências de cana e arroz. A seta 2 indica o ponto de separação entre sequências de gramíneas e arabidopsis. A seta 3 indica o ponto de separação entre dois subgrupos, o das sequências mais próximas à transposase domesticada DAYSLEEPER e o das sequências mais próximas aos transposons Ac e Tam3.

A seta número 3 indica o ponto de separação entre os cDNAs do grupo I, mais próximos da transposase domesticada DAYSLEEPER, e dos cDNAs do grupo II, mais próximos aos transposons Ac e Tam3. Os dois grandes ramos em que a árvore se divide a partir da seta 3 apresentam sequências de cana-de-açúcar, arroz e arabidopsis, o que sugere a existência de um ancestral comum que passou por um processo de domesticação anterior à separação de monocotiledôneas e eudicotiledôneas. Após a domesticação este ancestral teria seguido um caminho evolutivo como gene domesticado, cujos homólogos em cana-de-açúcar são referentes aos cDNAs do grupo I, e outro caminho evolutivo como transposon, cujos homólogos em cana são referentes aos cDNAs do grupo II.

IV.2.4.1. Os cDNAs do Grupo I de Cana-de-açúcar e Dois Homólogos Encontrados em Arroz Apresentam Alta Similaridade de Sequência com Transposases Domesticadas

Recentemente foi descrita a primeira transposase domesticada em plantas, DAYSLEEPER de *Arabidopsis thaliana* (Bundock & Hooykaas, 2005). Em seguida foram descritos homólogos a este gene em trigo e cevada (Muehlbauer *et al*, 2006). Devido ao impacto destas descobertas, decidiu-se verificar se DAYSLEEPER apresentaria similaridade com os cDNAs de cana. O alinhamento e a árvore apresentados na figura 11 mostraram que os cDNAs do grupo I agruparam-se com a transposase domesticada.

Com o objetivo de identificar homólogos a esta transposase também no genoma do arroz, foi realizada uma busca no genoma do arroz através de BLASTN no “Plant Genomes Central” do Genbank, utilizando a sequência inteira de nucleotídeos do representante do grupo I, TE074 como sonda eletrônica. A busca identificou dois loci que apresentaram alta homologia com o TE074. O primeiro, denominado “sequência 1 de arroz”, é uma sequência de 801 pb encontrada dentro de um BAC de DNA genômico de arroz com número de acesso no Genbank NT_079916.2, localizada na região subtelomérica do cromossomo 2 do arroz. Esta sequência apresentou identidade de nucleotídeos de 83% (677/812) com o TE074. O segundo locus, denominado “sequência 2

de arroz” é uma sequência de 2161 pb encontrada dentro de um BAC de DNA genômico de arroz, com número de acesso no Genbank NT_107222, localizada na região telomérica do cromossomo 5 do arroz. Esta sequência apresentou identidade de 80% (791/977) com o TE074.

Para elucidar melhor as relações entre a transposase domesticada de *arabidopsis* e seus homólogos em trigo, cevada, arroz e cana-de-açúcar, uma segunda análise envolvendo DAYSLEEPER foi realizada. Foi feito um alinhamento múltiplo no ClustalX, englobando um segmento que abrange os resíduos 509 a 751 da transposase de Ac, e os resíduos 473 a 712 da transposase de Tam3. Foram alinhadas as sequências traduzidas para aminoácidos dos cDNAs TE048, TE074 e TE096 (grupo I) e as sequências 1 e 2 de arroz, bem como as sequências das transposases dos elementos Ac e Tam3, da transposase domesticada DAYSLEEPER, e de seus homólogos w-gary1 e w-gary2 de trigo e b-gary de cevada. Estas sequências estão discriminadas na tabela IX.

Tabela IX – Identificação das sequências incluídas no alinhamento com as transposases domesticadas homólogas a *hAT* e os cDNAs de cana-de-açúcar.

nome da sequência	número de acesso no Genbank	tipo de sequência	organismo de origem
Daysleeper	AY728267.1	aminoácidos	<i>Arabidopsis. thaliana</i>
w-gary1	AM087609.1	aminoácidos	<i>Triticum aestivum</i>
w-gary2	AM087610.1	aminoácidos	<i>Triticum aestivum</i>
b-gary	AM087608.1	aminoácidos	<i>Hordeum vulgare</i>
Ac	CAA29005	aminoácidos	<i>Zea mays</i>
Tam3	CAA38906	aminoácidos	<i>Antirrhinum majus</i>
seq1arroz	NT_079916.2	nucleotídeos traduzidos para aminoácidos	<i>Oryza sativa</i> (cultivar japonica)
seq2arroz	NT_107222.1	nucleotídeos traduzidos para aminoácidos	<i>Oryza sativa</i> (cultivar japonica)

Este segundo alinhamento e a árvore gerada a partir do mesmo, apresentados na figura 12, indicam que as sequências de cana e arroz apresentam alta similaridade, agrupando-se na árvore (figura 12-B) em um ramo com suporte de bootstrap de 100%. Estas formam um subgrupo que, por sua vez, é mais relacionado com DAYSLEEPER que as transposases de cevada e trigo. Os transposons Ac e Tam3 ficaram posicionados na árvore como grupo externo. Para o segmento alinhado, o TE074 apresentou 67% de similaridade com DAYSLEEPER, enquanto o TE048 e o TE096 apresentaram 68%. As sequências do grupo I da cana também agruparam-se mais próximas de DAYSLEEPER que os homólogos de trigo e cevada. As sequências de arroz apresentaram maior similaridade com a transposase domesticada e os TEs do grupo I. A sequência 1 de arroz apresentou similaridade de 95% (244/256) com o TE074 e de 69% (177/256) com DAYSLEEPER ao longo do segmento alinhado, enquanto a sequência 2 de arroz apresentou similaridade de 95% (244/256) com o TE074 e de 68% (175/256) com DAYSLEEPER. Estes dados fortalecem resultados anteriores que indicam que os cDNAs do grupo I podem ser relativos a uma transposase domesticada presente no genoma da cana-de-açúcar, e sugerem que o genoma do arroz também pode possuir cópias desta transposase domesticada.

A		10	20	30	40	50	60	70
Seq1Arroz		CNYLKLLYDS	AHSIMAAANP	TSNLEFFHEAW	KLQLELSNAT	GHEDPVFSSSI	AKDMHERFDK	YWKDCNLVLA
Seq2Arroz		CNYLKLLYDS	AHSIMAAANP	TSNLEFFHEAW	KLQLELSNAT	GHEDPVFSSSI	AKDMHERFDK	YWKDCNLVLA
TE096		CNYLKLLYDS	AHSVMAASNP	TANIFFHEAW	KIQLELANGT	EHESPFFSSI	AKDMHERFDK	YWKDCSLVLA
TE048		CNYLKLLYDS	AHSVMAASNP	TANIFFHEAW	KIQLELANGA	EHESPFFSSI	AKDMHERFDK	YWKDCSLVLA
TE074		CNYLKLLYDS	AHSIMAAANP	TANIFFHEAW	KLQLELANGT	GHEDPVFSSSI	AKDMHERFDK	YWKDCSLVLA
DAYSLEEPER		CTFLKPLFEA	VSTLQSTGMP	SAVTEFFHEVW	KTQSDLSRAI	AGEDPFVTGI	AKTMQEKVDK	YWRDCSLVLA
wgary1		CKNLRTFYRV	IEVISCPSSV	TANIFYNEVW	KVRTVLQEEA	SNDNGEFSRM	VMEHQETFHG	YWNQSYLWLS
wgary2		CKILRMFYRV	IEVISCPSSV	TANIFYNEVW	KVRTVLHEEA	PNGNGEFSRM	VMEHQETFHG	YWNQSYLWLS
bgary		CKILRTFYRV	IEVISCPSSV	TANIFYNEVW	KVRTVLQEEA	SNDNGEYSRL	VMQMQESFHE	YWNQSYLWLS
Activator		FKCLKKFFDL	TELLSGTQYS	TANLFYKGFEC	EIKDLIDQWC	VHEKTVIRRM	AVAMSEKFEK	YWKVSNLALA
Tam3		VPYLVEFEF	TKIMSSCYTP	TSNIMLLYMW	SVVRLFHQHR	NHAT--LKN	IKEMEKKWVK	YKKVPHVCI
		80	90	100	110	120	130	140
Seq1Arroz		IADVMDPRFK	MKLVEFSYSK	IYGVAAKY-	-VKVDDAVH	ELYKEYVAQP	LPLTPAYVEQ	GDGNNAPASE
Seq2Arroz		IADVMDPRFK	MKLVEFSYSK	IYGVAAKY-	-VKVDDAVH	ELYKEYVAQP	LPLTPAYVEQ	GDGNNAPASE
TE096		IADVMDPRFK	MKLVEFSYSK	IYGAEEAKY-	-VKVINDSLH	ELYKEYVAQP	LPLTPAYVEA	---NNVAANA
TE048		IADVMDPRFK	MKLVEFSYSK	IYGAEEAKY-	-VKVINDSLH	ELYKEYVAQP	LPLTPAYVEA	---NNVAANA
TE074		IADVMDPRFK	MKLVEFSYSK	IYGAEEAKY-	-VKVINDAVH	ELYKEYVAQP	LPLTPAYVEQ	GDGNNAPANA
DAYSLEEPER		MAVMDPRFK	MKLVEFSYSK	IFGEDAGKN-	-IKTVDDGIH	ELFTEYMALP	SPQN-----	-----
wgary1		IPVVDPRFK	IGFIEFRLKR	AFGTHSESY-	-LSDIRETVR	ELFNEXCNPT	DQPS-----	-----AA
wgary2		IPVVDPRFK	IGFIEFRLKR	AFGTHSESY-	-LSDIRETVR	ELFNEXCNPT	DQPS-----	-----AA
bgary		IPVVDPRFK	ISFIEFRLKR	AFGTHSESY-	-LSEIRDIVR	ELFNEXCNPT	DQPS-----	-----TA
Activator		VACFLDPRYK	KILIEFYMKK	FHGDSYKVK-	-VDDFVRVIR	KLYQFYSSCS	PSAP-----	-----KTKTT
Tam3		LSSCLDPRVR	LIGTLELLEK	YHSALNNVYN	GNEERNRILQ	LLYSLYDMYA	PSTD-----	-----
		150	160	170	180	190	200	210
Seq1Arroz		NGTQATAPST	GDGLVDFDMY	LSEIATSQPT	K---SELEQY	LDESILTPRIQ	EPDILNWWKL	NTLKPEPTLSR
Seq2Arroz		NSTQTAPST	GDGLVDFDMY	LSEIATSQPT	K---SELEQY	LDESILTPRIQ	EPDILNWWKL	NTLKPEPTLSK
TE096		NVNQGNPPST	SDGLLDVDMY	LSEIQSSQPA	K---CELEQY	LEESILTPRIQ	EPDILNWWKL	NTVKPEPTLSK
TE048		NVNQGNPPST	SDGLLDVDMY	LSEIQSSQPA	K---CELEQY	LEESILTPRIQ	EPDILNWWKL	NTVKPEPTLSK
TE074		NNSQGPAST	GDGLLDVDMY	LSEIQSSQPS	K---SELEQY	LDESILTPRIQ	EPDILNWWKL	NTVKPEPTLSK
DAYSLEEPER		TTSEG---GK	ADGLSDFDY	IMET-TGQNL	K---SELDQY	LDETLLPRVQ	EPDVLDDWKQ	NKLKYPITLSK
wgary1		SNIAALDVD-	NDLLEDWDQH	LNEQASCQIS	---TELDQY	LEEGVPRKD	EPDILNWWMN	HTAKYPTLAA
wgary2		SNIAALDVD-	NDLLEDWDQH	LNEQASCQIS	---TELDQY	LEEGVPRKD	EPDILNWWMN	HTAKYPTLAA
bgary		SNIAALGADY	NDLLEDWDQH	LNEQAGSQIS	---TELDQY	LEEGVPRKD	EPDILNWWMN	HTAKYPTLAA
Activator		TNDSMDTLM	ENEDDEFQNY	LHELKDYDQV	ES--NELDKY	MSEPLLKHS	QFDILSWWRG	RVAEXPILTQ
Tam3		MDSEPTNASR	GSQFSIFDEL	LSNQSQSQPS	VGNYTEHMLF	VQKPPQKFDK	EPDILKWRWQ	NESLTPVLAR
		220	230	240	250	260		
Seq1Arroz		MARDILAIPH	SMVSSGNSIF	SAGTGTRMLD	DYRSSLRPEI	VEALVCAKDW	L	
Seq2Arroz		MARDILAIPH	SMVSSGNSIF	SAGTGTRMLD	DYRSSLRPEI	VEALVCAKDW	L	
TE096		MARDILAIPH	SMVSSGSSIF	SAGTGNRMLD	DYRSSLRPEI	VEALVCAKDW	L	
TE048		MARDILAIPH	SMVSSGSSIF	SAGTGNRMLD	DYRSSLRPEI	VEALVCAKDW	L	
TE074		MARDILAIPH	SMVSSGSSIF	SAGTGSHMLD	DYRSSLRPEI	VEALVCAKDW	L	
DAYSLEEPER		MARDILSIPV	SAAAFD-YVF	DMEP--REMD	EYKTSLRPET	VEALICAREW	L	
wgary1		MAQDILAMPA	SAVQSEAAFS	SSGP---VIP	KHYSTLSIKT	IEALVCARDW	M	
wgary2		MAQDILAMPA	SAVQSEAAFS	SSGP---VIP	KHYSTLSIKT	IEALVCARDW	M	
bgary		MAQDILAMPA	SAVQSEAAFS	SSGP---VIP	KHYSTLSIKT	IEALVCARDW	M	
Activator		IARDVLAIQV	STVASESAFS	AGGR---VVD	PYRNRLGSEI	VEALICTKDW	V	
Tam3		IARDLLSSQM	STVASESAFS	AGHR---VLT	DARNRLKPGS	VKECHIKWDM	L

Figura 12 – Alinhamento múltiplo, através do ClustalX, de um segmento de proteína homólogo a transposases *hAT*. O alinhamento traz as sequências dos cDNAs do grupo I, TE048, TE074 e TE096, traduzidas para aminoácidos, bem como duas sequências genômicas de arroz homólogas ao TE074 encontradas no Genbank através de BLASTN, traduzidas para aminoácidos. Estão incluídos no alinhamento segmentos das proteínas dos elementos Ac e Tam3, e das seguintes transposases domesticadas homólogas a *hAT*: DAYSLEEPER, w-gary1, w-gary2 e b-gary. Os asteriscos indicam posições em que todas as sequências apresentam o mesmo aminoácido.

IV.2.4.2. Os Elementos do Grupo I estão Representados com Baixo Número de Cópias nos Genomas das Gramíneas

Um ensaio de hibridação molecular de DNA foi realizado com o objetivo de verificar padrões polimórficos entre as linhagens parentais de *Saccharum* e cultivares híbridos, assim como entre diferentes gramíneas e outras angiospermas. Os perfis obtidos na hibridação com a sonda do grupo I caracterizaram sua presença no genoma de todas as gramíneas analisadas (híbridos e parentais da cana-de-açúcar, milho e arroz), com um baixo número de cópias. Não foi detectada hibridação da sonda do grupo I com os genomas das monocotiledôneas que não pertencem à família *poacea*, orquídea e abacaxi, nem com o genoma da batata, uma eudicotiledônea (figura 13-A).

Figura 13 - Autorradiografias de hibridação molecular com sondas radioativas. DNAs genômicos de **S**: *Solanum tuberosum*; **C**: *Catassetum fimbriatum*; **A**: *Ananas comosus*; **O**: *Oryza sativa*; **Z**: *Zea mays*; **Ss**: *Saccharum spontaneum*; **So**: *Saccharum officinarum*; **Sc**: híbridos da cana, foram hidrolisados com endonucleases, corridos em gel de agarose, transferidos para membrana de náilon e hibridados com sondas radioativas. **A**: DNAs hidrolisados com as endonucleases *XbaI* e *DraI*, hibridizados com a sonda do TE074 do grupo I. **B**: DNAs hidrolisados com as endonucleases *XbaI* e *DraI*, hibridizados com a sonda interna ao transposon Ac original do milho. A sonda representante do grupo I, desenhada sobre a sequência do TE074, teve sua presença detectada nos genomas de *Oryza sativa*, *Zea mays*, dos parentais *Saccharum spontaneum* e *Saccharum officinarum*, e dos híbridos da cana-de-açúcar, como indicado com a barra vermelha. A sonda interna do transposon Ac original do milho apresentou padrão de hibridação com os DNAs genômicos do milho e do arroz, e serviu como um controle da quantidades de DNA e da qualidade das membranas.

Os perfis sugerem um padrão de distribuição distinto dos elementos do grupo I nos genomas do arroz e do milho, assim como do parental da cana *Saccharum spontaneum*. Já o parental *Saccharum officinarum* apresentou um padrão de distribuição dos elementos do grupo I em seu genoma semelhante ao padrão dos híbridos. A figura 13-B apresenta o padrão de hibridação da sonda interna ao transposon Ac original do milho. Esta sonda hibridizou com os DNAs genômicos do milho e do arroz, e não apresentou hibridação com os demais. Este resultado sugere que, dentro das gramíneas, apenas o arroz possui homólogos ao elemento Ac do milho conservados o suficiente para serem detectados pela sonda. Já dentro do gênero *Saccharum* esta

conservação não foi mantida. A utilização desta sonda serviu também como um controle das quantidades de DNA e da qualidade das membranas.

IV.2.4.3. O Grupo I Apresenta Diferentes Padrões de Expressão em Tecidos de Cana-de-açúcar

Os transposons apresentam um efeito mutagênico quando transposicionalmente ativos. Devido a esta característica, eles encontram-se geralmente silenciados em seus genomas hospedeiros, de modo a manter a viabilidade dos mesmos. Já para uma transposase domesticada acredita-se que os níveis de expressão devem ser mais próximos de um gene endógeno, uma vez que os seus domínios protéicos estão dentro de um gene com função celular. Com base nestas informações e na hipótese de que os elementos homólogos a *hAT* do grupo I da cana-de-açúcar podem ser relativos a transposases domesticadas, foi realizado um ensaio de hibridação molecular de RNA com o intuito de investigar a expressão destes elementos do grupo I.

O RNA total utilizado no ensaio foi extraído de folhas e raízes de plantas da variedade de cana-de-açúcar R570 crescidas em vasos com terra e vermiculita (1:1) e transferidas para hidroponia, e de calos da variedade de cana-de-açúcar SP-89-1115. Todo o material vegetal do qual foi extraído RNA está descrito detalhadamente no item III.1.2. Os RNAs foram migrados em gel de agarose e transferidos para membrana de náilon, que foi hibridada com a sonda do grupo I, TE074. Os resultados do ensaio podem ser visualizados na figura 14. A sonda do TE074 apresentou hibridação com o RNA dos três tecidos analisados: folhas, raízes e calos, demonstrando haver expressão do elemento relativo ao cDNA TE074 em todos os tecidos. A expressão em calos, no entanto, se mostrou significativamente maior que nos outros tecidos.

Figura 14 – Resultado de ensaio de hibridação molecular com RNA total de folhas, raízes e calos de cana-de-açúcar. Os RNAs foram migrados em gel de agarose e transferidos para membrana de náilon, que foi posteriormente hibridada com as sondas de interesse. **A:** Hibridação com sonda do grupo I, TE074. **B:** Hibridação com sonda de gene constitutivo, ubiquitina. **C:** Resultado da análise da expressão relativa do TE074 nos diferentes tecidos. Os valores de expressão foram obtidos através da medida da intensidade das bandas no filme exposto às membranas hibridadas, por densitometria. A expressão relativa foi calculada pelo quociente entre os valores de expressão da sonda do TE074 e os respectivos valores de expressão da sonda da ubiquitina.

IV.2.5. Estudo dos Homólogos a *hAT* do Grupo II

O grupo II de clones homólogos a superfamília *hAT* foi analisado quanto a sua diversidade de sequências, padrão de inserção e verificação da estrutura de um transposon. A figura 15 apresenta o alinhamento e a árvore referentes às sequências do grupo II, alinhadas com os elementos Ac e Tam3. O longo comprimento dos braços da árvore reflete a alta divergência entre as

sequências, que sugere tratarem-se de diferentes elementos. Os transposons Ac e Tam3 ficaram próximos, compartilhando com o TE032 um ramo com suporte de bootstrap de 62%. O TE191 e o TE221 apresentaram-se fortemente relacionados, formando um ramo com suporte de 100%. Os cDNAs TE257 e TE191 estão indicados com as setas 1 e 2 respectivamente. Pode-se notar uma clara distância entre estas duas sequências, e por esta razão as mesmas foram escolhidas como sondas para subsequente ensaio de hibridação molecular, de modo a verificarmos o padrão de distribuição de dois representantes distintos dentro do grupo II.

Figura 15 – Padrão de agrupamento dos cDNAs do grupo II. **A:** Alinhamento de um domínio conservado de transposases do tipo *hAT*, localizado entre os resíduos 679 a 712 da transposase de Ac e 640 a 673 da transposase de Tam3. Estão alinhadas as sequências dos cDNAs do grupo II e dos elementos Ac e Tam3. **B:** Filogenia gerada a partir do alinhamento, utilizando-se o algoritmo de “neighbor joining”, com “bootstrap” de 500 réplicas. O comprimento dos braços da árvore reflete o número de diferenças entre os aminoácidos das sequências. A barra verde destaca os transposons Ac e Tam3. As setas de número 1 e 2 indicam, respectivamente, o TE257 e o TE191.

IV.2.5.1. Análise da distribuição de dois elementos do grupo II nos genomas de mono e eudicotiledôneas

As sondas do grupo II hibridaram somente com os genomas do gênero *Saccharum*.

Figura 16 – Autorradiografias de hibridação molecular com sondas radioativas. DNAs genômicos de **S**: *Solanum tuberosum*; **C**: *Catassatum fimbriatum*; **A**: *Ananas comosus*; **O**: *Oryza sativa*; **Z**: *Zea mays*; **Ss**: *Saccharum spontaneum*; **So**: *Saccharum officinarum*; **Sc**: híbridos da cana. Os DNAs foram hidrolisados com endonucleases, corridos em gel de agarose, transferidos para membrana de náilon e hibridados com sondas radioativas. **A**: DNAs hidrolisados com a endonuclease *XbaI*, hibridados com as sondas do grupo II, amplificadas a partir do TE257 e do TE191, e com sonda interna do transposon Ac original do milho. **B**: DNAs hidrolisados com a endonuclease *DraI*, hibridados com as mesmas sondas citadas em A. Em ambos os perfis a sonda do TE257 teve sua presença detectada nos genomas do gênero *Saccharum*, os parentais *Saccharum spontaneum* e *Saccharum officinarum* e os híbridos da cana-de-açúcar. A sonda do TE191 apresentou hibridação com os dois híbridos analisados, e com somente um dos parentais, *Saccharum officinarum*. A sonda interna do transposon Ac original do milho apresentou padrão de hibridação com os DNAs genômicos do milho e do arroz, e serviu como um controle da quantidades de DNA e da qualidade das membranas.

A sonda do TE257 hibridou com os genomas dos parentais e híbridos da cana, enquanto a sonda do TE191 hibridou com os genomas dos híbridos da cana e somente um dos parentais, *S. officinarum* (figura 16). Ambas as sondas não apresentaram hibridação com os genomas das monocotiledôneas que não pertencem à família *poacea*, orquídea e abacaxi, nem com o genoma da batata, uma eudicotiledônea. Estes resultados indicam que o grupo II têm sequências específicas do genoma de *Saccharum*.

IV.2.5.2. Estruturas de Elementos Homólogos à Superfamília *hAT* Amplificadas do Genoma da Cana-de-Açúcar

IV.2.5.2.1. Identificação de Potenciais TIRs para os Elementos *hAT* do Grupo II pela Técnica de PCR-Inversa

As Repetições Invertidas Terminais (TIRs) são estruturas importantes a serem descritas para uma família de transposons, uma vez que determinam o sítio de reconhecimento e de ligação da transposase que pode efetuar sua transposição. Com o objetivo de obter as sequências das TIRs da cópia genômica relativa ao cDNA TE221 do grupo II, desenhou-se um ensaio de PCR-Inversa que permitiu acessar sequências flanqueadoras à sequência do cDNA, de ambos os lados. O ensaio de PCR-Inversa está descrito detalhadamente no item III.13.1, e uma esquematização de todo o procedimento encontra-se na figura 17-A. Nesta mesma figura (17-B) é ilustrado o resultado obtido após as duas reações de amplificação (iniciadores externos e “nested”), foi possível observar bandas discretas no gel. Foram selecionadas 2 bandas provenientes das reações 3 e 7 da PCR-Inversa, de aproximadamente 1100pb e 750pb, respectivamente. Estas bandas foram purificadas do gel de agarose e subclonadas em pMOSBlue. Os clones foram posteriormente sequenciados com iniciadores 5’ e 3’ (“T7” e “pmos”). O sequenciamento mostrou existirem três sequências distintas contidas na banda 3, e uma sequência apenas na banda 7, que foram denominadas **3_pl2_G02**, **3_pl2_D02**, **3_pl1_A02** e **7.2_pl2_F11**, respectivamente. Das sequências

geradas, quatro clones foram analisados em detalhe por apresentarem qualidade nos cromatogramas gerados, e por apresentarem pelo menos dois, dos três requisitos básicos. **a)** presença dos iniciadores da PCR-Inversa; **b)** sequência do cDNA do clone TE221 interna aos iniciadores e **c)** sítio da endonuclease.

Figura 17 – Resultados esperados e obtidos do ensaio de PCR-Inversa. **A:** Diagrama esquematizando os passos seguidos na realização da PCR-Inversa e as estruturas esperadas dentro das sequências amplificadas. **B:** Bandas amplificadas através de ensaio de PCR-Inversa. As setas amarelas indicam as bandas 3 e 7.2, respectivamente. Estas bandas foram isoladas, subclonadas e sequenciadas. Após o sequenciamento a banda 3 apresentou três sequências distintas com 98% de identidade nucleotídica entre si, e a banda 7.2 apenas uma sequência.

A análise inicial das sequências consistiu-se da busca pelas seguintes estruturas: iniciador 3' (IPP-TE221-2F), porção final da sequência do clone de cDNA TE221, identificação de potencial TIR 3', sítio duplicado de inserção, região genômica a jusante do elemento, sítio da endonuclease utilizada para hidrolisar o DNA genômico, região genômica a montante do elemento, sítio duplicado de inserção, identificação de potencial TIR 5', porção inicial da sequência do clone de cDNA TE221 e iniciador 5' (IPP-TE221-2R). Uma vez caracterizada a presença dos iniciadores nos insertos, a homologia com o cDNA foi checada usando o algoritmo BLAST através de uma busca no banco de dados de nucleotídeos no NCBI, assim como um alinhamento duplo (BLAST-2-sequences) no banco local do laboratório. Posteriormente, um alinhamento múltiplo dos 4 insertos foi feito localmente de modo a identificar possíveis regiões de similaridade entre os insertos. A figura 18 ilustra, de modo esquemático, o resultado desta análise. Nota-se que três dos quatro insertos correspondem à mesma região genômica, apresentando 98% de identidade em nucleotídeos, com alguns polimorfismos pontuais (SNPs-“single nucleotide polymorphisms”). Cabe ressaltar que estes três insertos foram clonados a partir da banda **3** isolada da PCR-Inversa. A sequência relativa ao fragmento **7.2** da PCR-Inversa, de 755pb, também incluída no alinhamento, apresenta identidade de 98% até a posição 491 com os outros três insertos. A partir desta posição a sequência **7.2** mostrou-se bastante divergente das demais. A identidade de nucleotídeos entre as sequências **3** e a sequência **7.2** apenas retorna a partir da posição 700 desta última, em que aparece a sequência do iniciador IPP-TE221-2R repetida duas vezes, das posições 700 a 720 e 735 a 755 (figura 18). Esta identidade, porém, restringe-se à região do iniciador e não há identidade com a porção terminal do cDNA.

Figura 18 – Diagrama esquematizando o alinhamento das sequências 3_pl2_G02, 3_pl2_D02, 3_pl1_A02 (provenientes do fragmento 3 da PCR-Inversa) e 7.2_pl2_F11 (proveniente do fragmento 7.2 da PCR-Inversa). As setas coloridas representam os iniciadores, “Forward” em vermelho e “Reverse” em verde. As caixas em branco representam a sequência dos clones, e as porcentagens referem-se às identidades em nucleotídeos entre os fragmentos. A caixa hachurada indica a posição relativa das TIRs e no alto a sequência correspondente está destacada.

O fato da sequência 7.2 apresentar estas características pode ser devido a um rearranjo do DNA genômico do qual foi amplificada, ou a um artefato ocorrido durante a reação de PCR-Inversa. Independente deste fato, a região da posição 1 a 491 foi analisada em detalhe, e assumiu-se como premissa que a TIR estaria contida na porção “extra-cDNA” destes insertos. Tomamos 50 nucleotídeos considerados terminais da inserção genômica, da posição 442 à 491, e buscou-se através do programa “Oligo Analyser” a presença de regiões terminais inversamente repetidas nos insertos. Este programa reconhece motivos reverso-complementares e localizou uma região de 21 nucleotídeos, entre as posições 637 e 657 dos três clones. A figura 19 apresenta a sequência do clone “3_pl2_G02” onde estão indicadas as regiões especificadas no diagrama apresentado na figura 18. Os detalhes do

alinhamento entre as quatro sequências, que é esquematizado na figura 18, podem ser observados na figura 20.

```
>3_p12_G02
ATGCTCAAAGGGATTATCATTCCATATTGATGCTTGCAGCATCGTATTGATG
CTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCATAGCCTAGCAC
TTAAAGAGTATTTCTGTGTATTAGATAATACTAGAACTATACTCTCTAGTA
ATTTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGAATGACAAT
ACTGCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCTGTGGTCGCTA
CTGAAGAAATGTACTACTCTTTTGCCGTGTTTCAGTTTGAGGAAGTGAATGTG
GTGGGTAAAATGTGATGTGTCAATCAATCATTGTTTCTAGATTAACTTAAT
ACTGCAGTTCCAATGTTTATTCTGTGTTTGACAAAAAGTGGATTTCAGTGCTG
ATCATCCAATGTTTCGGCCTCACTGTGTGCAATGGTTCGAGCAAAATGCTTT
GAAGATTATTGCTGATCATCTGTactgctgctgtagtggttgctgggtggtct
ggctgtggcagtggtgcctgatggctgatgccttttgtatatttatggagcta
ctgcataatgcattattgtccttgggtctattgcactctggtgataataactga
tatattccattgATAAATGATAAGTGATAACCTGCTGTGATCATGTAACCTC
TGGGAAAAGCTTTTGCTTATTAACATAATTGTCTCTGGCCTTGCTGTCTACTA
ATATTGTACCTGTGAACCTAGAACAAGACATCATTTCTTTAATTTTTATGTA
CATGTACAGTTATAAGAAATTGAGAACTAGTCACTTGATGTAGTATTTTCATT
TTTTAGAGGCCTTGATTCTTAAATCATGTCTTAATTCTTTTATATTTAATAA
TGACAGGTAAAGCCGAGAAAGCTGAGATCTACTATTTGGAAGGACATGGATCC
CATCTATCAAGATGGTAAAGTCATACAAAGCCGATTTAAGCACTGCTATGAG
GTATTTGCTGTTGCCCGTACTTCAGGGACTAGTCATATGCGTAGACATTTGG
AAATTTGTGAGCCAAGACTGAAGATGC
```

Figura 19 – Sequência do clone “3_p12_G02”, amplificado na PCR-Inversa. Em vermelho e sublinhada está a região do iniciador IPP-TE221-2F. Em verde sublinhada está a região do iniciador IPP-TE221-2R. Em azul estão indicadas as regiões 3’ e 5’ terminais do clone de cDNA TE221. Destacada em amarelo está a potencial TIR 3’ do elemento, e em destaque azul está a potencial TIR 5’ do elemento. As letras minúsculas correspondem à região genômica, a jusante e a montante do elemento.

	10	20	30	40	50
					
3_p12_G02	AT	GCTCAAAGGGATTATCATTCC	ATATTGATGCTTGCAGCATCGTATTGA		
3_p12_D02	AT	GCTCAAAGGGATTATCATTCC	ATATTGATGCTTGCAGCATCGTATTGA		
3_p11_A02	AT	GCTCAAAGGGATTATCATTCC	ATATTGATGCTTGCAGCATCGTATTGA		
7.2_p11_C10	AT	GCTCAAAGGGATTATCATTCC	ATATTGATGCTTGCAGCATCGTATTGA		

	60	70	80	90	100
					
3_p12_G02	TGCTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCATAGCCTA				
3_p12_D02	TGCTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCATAGCCTA				
3_p11_A02	TGCTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCATAGCCTA				
7.2_p11_c10	TGCTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCATAGCCTA				

	110	120	130	140	150
					
3_p12_G02	GCACTTAAAGAGTATTTCTGTGTATTAGATAAATACTAGAAACTATACTCT				
3_p12_D02	GCACTTAAAGAGTATTTCTGTGTATTAGATAAATACTAGAAACTATACTCT				
3_p11_A02	GCACTTAAAGAGTATTTCTGTGTATTAGATAAATACTAGAAACTATACTCT				
7.2_p11_c10	GCACTTAAAGAGTATTTCTGTGTATTAGATGATACTAGAAACTATACTCT				

	160	170	180	190	200
					
3_p12_G02	CTAGTAATTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGA				
3_p12_D02	CTAGTAATTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGA				
3_p11_A02	CTAGTAATTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGA				
7.2_p11_c10	CTAGTAATTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGA				

	210	220	230	240	250
					
3_p12_G02	ATGACAATACTGCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCT				
3_p12_D02	ATGACAATACTGCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCT				
3_p11_A02	ATGACAATACTGCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCT				
7.2_p11_c10	ATGACAATACTGCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCT				

	260	270	280	290	300
					
3_p12_G02	GTGGTCGCTACTGAAGAAATGTACTACTCTTTTGCCGTGTTGAGTTGAG				
3_p12_D02	GTGGTCGCTACTGAAGAAATGTACTACTCTTTTGCCGTGTTGAGTTGAG				
3_p11_A02	GTGGTCGCTACTGAAGAAATGTACTACTCTTTTGCCGTGTTGAGTTGAG				
7.2_p11_c10	GTGGTCGCTACTGAAGAAATGTACTACTCTTTTGCCGTGTTGAGTTGAG				

	310	320	330	340	350
					
3_p12_G02	GAACTGAATGTGGTGGGTAAAAATGTGATGTGTCAATCAATCATTGTTTCT				
3_p12_D02	GAACTGAATGTGGTGGGTAAAAATGTGATGTGTCAATCAATCATTGTTTCT				
3_p11_A02	GAACTGAATGTGGTGGGTAAAAATGTGATGTGTCAATCAATCATTGTTTCT				
7.2_p11_c10	GAACTGAATGTGGTGGGTAAAAATGTGATGTGTCAATCAATCATTGTTTCT				

	360	370	380	390	400
					
3_p12_G02	AGATTAAACTTAATACTGCAGTTCCAATGTTTATTCTGTGTTGACAAAAA				
3_p12_D02	AGATTAAACTTAATACTGCAGTTCCAATGTTTATTCTGTGTTGACAAAAA				
3_p11_A02	AGATTAAACTTAATACTGCAGTTCCAATGTTTATTCTGTGTTGACAAAAA				
7.2_p11_c10	AGATTAAACTTAATACTGCAGTTCCAATGTTTATTCTGTGTTGACAAAAA

Figura 20 – Detalhe do alinhamento dos primeiros 500 pb das sequências provenientes da PCR-Inversa. A barra numérica acima das sequências indica as posições em nucleotídeos. Em vermelho e sublinhada está a região do iniciador IPP-TE221-2F. Em azul está indicada a região 3' terminal do clone de cDNA TE221. A potencial TIR 3' do elemento está destacada em amarelo. Os asteriscos indicam posições em que todas as sequências apresentam os mesmos nucleotídeos.

IV.2.5.2.2. Obtenção das regiões Genômicas 5' e 3' não transcritas do TE221 - reconstrução parcial de um elemento *hAT*-like

O fragmento **3** amplificado na PCR-Inversa permitiu o acesso a uma inserção genômica e seus possíveis variantes alélicos, de um elemento homólogo ao clone de cDNA TE221. Uma vez que estas sequências genômicas contém os iniciadores IPP-TE221-2F e IPP-TE221-2R e as respectivas regiões terminais do clone de cDNA, admitiu-se que as sequências que se seguem correspondem a ambas as direções para fora do clone de cDNA. Com o intuito de iniciar uma reconstrução genômica parcial correspondente ao locus amplificado, foi gerada uma sequência quimérica entre o cDNA TE221 e as respectivas porções terminais contidas nos insertos **3**, conforme apresentado na figura 21. Esta sequência montada foi denominada “TE221-RP”, referente a “TE221 reconstrução parcial”. O primeiro estudo feito sobre esta reconstrução foi para descobrir se a sequência incorporada estenderia o mRNA em sua região 5'. Neste sentido foi realizado BLASTX-sequences, primeiramente entre a sequência do cDNA TE221 e a proteína da ORFa do elemento Ac do milho (número de acesso CAA29005 no Genbank). A

similaridade de aminoácidos entre as duas sequências está compreendida entre as posições 21 e 1766 nucleotídeos do clone de cDNA, e coincidiu com as posições 184 a 774 aminoácidos da ORFa de Ac. Quando a sequência colocada no BLASTX-2-sequences foi a do TE221-RP, a região de similaridade com a ORFa de Ac se estendeu em 49 aminoácidos em sua porção N-terminal. Alinharam-se os nucleotídeos das posições 259 a 2124 de TE221-RP, com as posições 135 a 774 aminoácidos da ORFa de Ac. Estes dados sugerem que o clone de cDNA TE221 é incompleto na sua porção 5', provavelmente devido à ineficiência da transcriptase-reversa no momento de síntese deste cDNA. Esta informação, no entanto, somente poderá ser confirmada com a realização de um experimento de "5' Race", que delimita o início de transcrição.

Figura 21 – Diagrama esquematizando a montagem da sequência "TE221-RP". Foram sobrepostas a sequência do cDNA TE221 e as sequências amplificadas na PCR-Inversa. A putativa TIR 5' encontra-se entre as posições 1 e 18. O putativo TATA box localiza-se entre as posições 241 e 247, seguido do putativo códon de início, no nucleotídeo 250 da sequência do TE221-RP. O hipotético códon de parada fica na posição 2145, e a putativa TIR 3' na extremidade, entre as posições 2861 e 2878.

Uma análise da sequência TE221-RP com a ferramenta ORF Finder do Genbank indicou um provável códon de início de tradução na posição 250, ou seja, 9 nucleotídeos antes do início da similaridade de aminoácidos com a ORFa de Ac. Para avaliar a consistência destas observações, uma análise

comparativa com elementos do tipo *hAT* de outras espécies foi realizada. A tabela X ressalta que assim como para a reconstrução TE221-RP, a similaridade com a transposase de Ac (ORFa) se inicia na posição 134/135 da mesma. No entanto, a região anterior ao início da similaridade com a transposase de Ac é maior nas proteínas homólogas de outras espécies do que na proteína predita pelo ORF Finder para a a sequência TE221-RP (tabela X). Esta diferença pode decorrer de uma particularidade da proteína predita ou mesmo ser consequência de uma deleção interna na sua porção N-terminal.

O ORF Finder indicou também a presença de códons de parada prematuros no clone TE221, o que é sugestivo de que pode tratar-se de uma inserção antiga, e portanto, com mais mutações acumuladas. Na análise por BLASTX-2-sequences entre a sequência de nucleotídeos do TE221-RP e a sequência de aminoácidos da transposase de Ac, a similaridade entre as duas sequências se estende até o nucleotídeo 2124 do TE221-RP. Logo, o códon de parada identificado pelo ORF Finder na posição 2145 pode ser o real códon de parada da transposase íntegra.

Tabela X – Resultado do alinhamento de sequências de proteínas homólogas ao TE221 e da sequência TE221-RP com a transposase de Ac do milho, produto da ORFa (gi 22113).

espécie de origem da sequência	acession number (Genbank)	posição na sequência de nucleotídeos em que começa a similaridade com a transposase de Ac	posição na sequência de aminoácidos da transposase de Ac em que começa a similaridade com a sequência nucleotídica homóloga de outra espécie
TE221-RP	-	9	135
<i>Oryza sativa</i>	BAF19024.1	216	134
<i>Oryza sativa</i>	ABF94985.1	216	134
<i>Medicago trunculata</i>	ABE94495.1	68	135
<i>Pennisetum glaucum</i>	2021344A	130	134
<i>Arabidopsis thaliana</i>	CAB81093.1	33	134
<i>Silene latifolia</i>	AAP59878.1	52	135

De posse de uma região 5' passível de conter as regiões regulatórias, foi realizada uma busca no PLACE ("Database of Plant Cis-acting Regulatory DNA Elements") por elementos regulatórios na porção a montante do putativo códon de início da sequência TE221-RP. Entre os nucleotídeos das posições 1 a 259 foram encontrados sítios de ligação de diversos fatores de transcrição, incluindo-se uma putativa sequência de TATA box. Esta sequência localiza-se entre as posições 241 e 247 (TATTTAA) da sequência TE221-RP e está muito próxima do códon de início de tradução (posição 250). De acordo com Snustad *et al* (1997), a distância média entre o TATA box e o início de transcrição em eucariotos é de aproximadamente 25 nucleotídeos. Somada a esta distância ainda há a distância entre o início de transcrição e o códon de início de tradução, que segundo Kunze *et al* (1987), para o elemento Ac do milho é de 653 nucleotídeos. A distância encontrada entre o TATA box e o "ATG" na sequência TE221-RP é muito pequena se comparada com os dados da literatura citados. Esta distância somente poderia ser explicada por uma deleção entre o TATA box e o "ATG". As distâncias observadas não são compatíveis com um locus contendo uma cópia ativa de um elemento do tipo *hAT*. Este fato é reforçado quando, a partir do estudo de alinhamento apresentado na tabela X, verificou-se que elementos descritos para outras espécies apresentariam uma proteína mais extensa na porção N-terminal. Sugere-se que este locus possa corresponder a um elemento que tem uma deleção que se estende da posição após o TATA box até a porção inicial da região transcrita. As características descritas para a sequência TE221-RP estão detalhadas na figura 22.

>TE221-RP

```

TGATAAAATGATAAGTGATAACCTGCTGTGATCATGTAACCTCGGGAAAAGCTTTTGCT
TATTAACATAATTTGTCTCTGGCCTTGCTGTCACTAATATTGTACCTGTGAACCTAGAAC
AAGACATCATTTCTTTAATTTTATGTACATGTACAGTTATAAGAAATTGAGAACTAGT
CACTTGATGTAGTATTTTATTTTATAGAGCCTTGATTCTTAAATCATGTCTTAATTCT
TTTATATTTTAAATGACAGGTAAGCCGAGAAAGCTGAGATCTACTATTTGGAAGGACA
TGGATCCCATCTATCAAGATGGTAAAGTCATACAAAGCCGATTTAAGCACTGCTATGAG
GTATTTGCTGTTGCCCGTACTTCAGGGACTAGTCATATGCGTAGACATTTGGAATTTG

```


TGAGCCAAGACTGAAGATGCATGATTTCGTTGATAAGTTACAGTCATTAGTCACAACCG
AGTCAGCTATTCTCACTAATTGGAGGTTTGATCCTAAATTGACTCGATGTGAGCTTGTT
CGTCTGATTGTACTACATGAGCTCCCATTCTCTTTTGTGGAGTATGAAGGATTTTCATAG
CTATTGAGCTAGCTTAAATCCACTCGCAGAAACCATATCTAGGACAACAATTAAGGAAA
ACTGCTTAGAAGCCTATAAGAACCACATAACAGCGCTGAGAGAAATGTTTGAGAACTGT
AATTTGAGATTCTCATTGACAGCTGATTTGTGGACTTCCAATAAGAACATAGGCTACAT
GTGCGTCACTTGCCACTATATAGATGATGATTGGAAAGTTCAGAAAAGGATAATCAGAT
TTTGTGTAGTAAAAACACCACATGATGGTTTTAATCTATATACTTCAATGTTGACGACT
ATTAAGTTTTTACAATATTGAAGACAAGTTATTTAGCATTACATTAGACAATGCAGCTCC
AAACAATACAATGATGGATATCTTGAAATCAAATTTGTTAAAGGTGGATATGTTGCATT
GTGATAGTGATCTATTTTCATATTAGATGTGCTGCCCATGTGATTAACCTTATTGTTAAA
GACGGTCTTCAAGCAATTGATGGTGTATCAATAACATAAGAGAGAGTGTAAGTATGT
TCGAGCTTCTCAATCAAGAAAAGAGACATTTGAGGATATAATTGTGGAAC TAGGCATTC
GTTTTCGGAGTGTGCCCAAAATTGATGTAGAAAATCATTGGAAC TCCACGTGTGATATG
ATAGAGTCTGCAATGCCGTTCAAAGAGGCATTTCTTGAGTTAGAAGTGAAAGATAGCTA
CTATACTTACTGCCCCCTATCTCAGGACTGGGAAAGGGCTAATGCAGTATGCAAGTTAT
TAAAGGTATTCAAGAAAGCAATAGAAGTGGTCTCTGGTACAAGCTACCCAACGTTGAAC
CTCTATTTTCATGAAATCTGGAGTATCAAACATGTTTTAGAAAGAGGAAGGATTTAGTTT
AAATGAAACTATTGCAGCCATGGTTTCCAAAATGCAAGAAAAATTTGATAAGTATTGGA
TGATTTTCATACTTCACAAACTGTGTGCCAGTTTATACCTTGATCTGTGGTTCAAATTTGGG
TTCATTGAGTTTCGCTTGAAGCAAGCATTTGGAGAGTATGGGAAGTGTCTTCACCTTG
ATAAAGTGGATCAAGCTATTAGAAGATTGTTCAATGCCTACTCTGCTCAAATGGGAGGT
TCTTCTCAAGTGGAGACCCATGGAGATGATGTGACAAC TGTGGTAAAGGTCACCTTG
GTCTGATTGGATGAACACACAAGTGCCAAAAGAAAAGAAGAAAACAACGAGTATGATAG
GTATCTACGTGATGATTTGTTCCCATGCGATGATGAGAGTTTTGACATTC TAGATTGGT
GGAAAAATGCACGCATCGAAATATCCAACATTGGCTGCCATAGCATGTGATATCTTGGCT
GTTACTGCTTCCACGGTACCATCTGAGTCCGCTTTCAGCACCAGTGGTCATATTATCAA
TGATCACAGGACCAGACTTGCAAGGCAGCACCGTTGAAGCTTATTGTGTTTTCAAGATTG
GCTACACAGCTGGTTCTTCATATTTAGACGTTATATCAATTGACAAC TGGATAGCATC
TCCAATGACTTATATCAGTAGCACATGTCTGGTCATGGCATGTGTTGCGGTGGAAGTTA
ATTTAGTCATGGCAGGGGGCTGAGGGAAC TTGTTTGGATGGCCTCAGACAGAGAGAAAG
GGTTGCTGAGTTTCACCATGTTGTAACTTGAGAGTTTCATTTATTTGTCAAGCAAGGT
GGTGCTTGTCATTATCTTCTCTGCGTAAAAATTGTGTTGTATTTTTTTTTATCATGTTGA
CCAGAAGGCCTGCAACTTCCAAACAAA GCTCAAAGGGATTATCATTCCATATTGATGCT
TGCAGCATCGTATTGATGCTTGCATTTGTTGCTGTTAGCACTTAGTTGTTAGACCTTCA
TAGCCTAGCACTTAAAGAGTATTTCTGTGTATTAGATAATACTAGAACTATACTCTCT
AGTAATTTTTCTTTCTGTGGTGTGTCTCGGTAACAAATTATTGTTAGAATGACAATACT
GCAAGCGGTTCTATTGTGCCAAGCTTGAATAGGCGTGCTGTGGTCGCTACTGAAGAAAT
GTACTACTCTTTTGCCGTGTTCAGTTTGAGGAACTGAATGTGGTGGGTAAAATGTGATG
TGTCATCAATCATTGTTTCTAGATTAAACTTAATACTGCAGTTCCAATGTTTATTCTG

```
TGTTTGACAAAAAGTGGATTCAGTGCTGATCATCCAATGTTTCGGCCTCACTGTGTGCA  
ATGGTTCGAGCAAAATGCTTTGAAGATTATTGCTGATCATCTGTAC
```

Figura 22 – Características encontradas na sequência TE221-RP. Destacada em azul a potencial TIR 5'. Destacado em verde o potencial TATA box. Destacado em cinza e em negrito o predito códon de início. Em azul a região 5' terminal do cDNA. Em verde sublinhado o iniciador IPP-TE221-2R. Destacado em cinza e em negrito o predito códon de parada. Em vermelho sublinhado o iniciador IPP-TE221-2F. Em azul a região 3' terminal do cDNA e destacada em amarelo a potencial TIR 3'.

Tendo em vista validar a montagem TE221-RP, a mesma foi analisada através de TBLASTX no banco de dados do genoma do arroz (www.tigr.org). A melhor correspondência retornou com um locus do cromossomo 7 do arroz (LOC_Os07g43120). O cDNA predito para este locus tem 2784 pb, enquanto o cDNA referente ao TE221-RP é o próprio TE221, de 2078 pb. Os dois cDNAs se alinharam entre as posições 265 a 2046 do TE221-RP e 600 a 2387 do locus de arroz (figura 23-A). Quando a sequência de nucleotídeos do TE221-RP e do locus de arroz são analisadas no banco de dados do “Swissprot” através de BLASTX, ambas recuperam como melhor correspondência a proteína transposase do elemento Ac do milho. Em ambos os casos com alinhamento superior a 80% da proteína de Ac e com similaridade em aminoácidos superior a 45% (figura 23-B).

Oryza sativa
Cromossomo 7
Locus g43120

A

Score = 517 bits (1124), Expect(3) = 3e-172
Identities = 214/394 (54%), Positives = 278/394 (70%), Gaps = 0/394 (0%)
Frame = +1/+3

Query 457 SLVTTE SAILTNAREDPKLRCELVRLIULHELPEFVEYEGHSYSASLNPLAETI SRT 636
+L++ S L NW FDP++R EL+R+IULHELPE FVEY+GF S++ASLNP + I SRT
Sbjct 795 TLMSPNSURLKNVSEFDFEVSRLKELHRCIVLHELPEQFVEYDGFERSFASLNPFYKII SRT 974

Query 637 TIKENCLEARYKJUTALREDFENCNFRSLTADIMT SMCNIGYIC VICHYI DDVAVQKR 816
TI+ +C+ R+K R++H+ N RESLTAD +MT SN+ +GYIC VICH+ID DVA+VQKR
Sbjct 975 TIRNDCIAREKEQKLAKQDFEGANCRSLTADIMT SMCNIGYIC VICHYI DDVAVQKR 1154

Query 817 IIRFCVVKTPHDCENLYT SMLTTIKFYNIEDKLSITLONARPNTDQILKSNLLKVID 996
II+ E VKTPH G ++ +HL+ I+ +NI DK+FS+TLB A+ N++H +LK NL
Sbjct 1155 IIRFCVVKTPHDCENLYT SMLTTIKFYNIEDKLSITLONARPNTDQILKSNLLKVID 1334

Query 997 LMCDSLEHRCARVINLIVKGLQRIQGVNNIRSVKVRASQSRKETFEEDLIVELG 1176
+ L H RCA HVINLI KQGL+ ID ++ NIRESVKY S SRKE FE+II + G
Sbjct 1335 IPAGKLLHRCARVINLIVKGLQRIQGVNNIRSVKVRASQSRKETFEEDLIVELG 1514

Query 1177 IIRFSVPKIDVENMNSTCMIESAMPKEARELEKVD SYTYCPSSQWERRANAVCKI 1356
I P+DV VNST M+ +R PE A+ L V+D Y Y PS WER+ V +
Sbjct 1515 IIRFSVPKIDVENMNSTCMIESAMPKEARELEKVD SYTYCPSSQWERRANAVCKI 1694

Query 1357 LKVEKKALEVVSGETSYPTLNLYEKEIWSIKVLEEGESLNETIARW SMOKEKEDKYM 1536
LKV R VVSG+ YPT NLYEKE+W IK VL++E + + +R+HV KH++KEDKYM
Sbjct 1695 LKVEKKALEVVSGETSYPTLNLYEKEIWSIKVLEEGESLNETIARW SMOKEKEDKYM 1874

Query 1537 ISYFTNCUPVILDLWFKFGTEERLQAFGEYCK 1638
SY C+PVI D EKE F+KEEL QAFGE K
Sbjct 1875 ISYFTNCUPVILDLWFKFGTEERLQAFGEYCK 1976

TE221-flancos
(2878 bases; cDNA tem 2078 bases)

J033024N05 cDNA (2784 bases)

B

```
>sp|P08770|TRAI_MAIZE Putative AC transposase (ORFA)
Length=807

Score = 294 bits (752), Expect = 1e-78
Identities = 203/651 (31%), Positives = 315/651 (48%), Gaps = 40/651 (6%)
Frame = +1

Query 259 KPRKLRSTIW----KMDPIYQDGKVIQ----SRFKHCYEVFAVARTSGTSHMRRLHLEI 411
          + +K S +W K+++ K +Q F +C + GTS R HL
Sbjct 135 RAKKCTSDVWQHFTKKEIEVEVDGKKYVQVWGHCFNPNCKAKYRAEGHGTSGFRNHLRT 194

Query 412 CEPRLKMHDFVDKLSLVTTESAILTNWRFDPKLTCELVRLIVLHELFPFVEYEGFHS 591
          +K K + + + ++D ++ +L I++HE PF+ VE+E F
Sbjct 195 SHSLVK-GQLCLKSEKDHGKDINLIEPKYDEVVSLKKHLAIIMHEYPFNI VEHEYFVE 253

Query 592 YSASLNPLAETISRTTIKENCLAYKNHITALREMFENCNFRFSLTADLWTSNKNIGYMC 771
          + SL P SR T ++ ++ Y L ++ RFS T D+WTS +N YMC
Sbjct 254 FVKSLRPHFPFKSRVTARKYIMDLYLEEKEKLYGKLKDQVSRFSTTMDMWTSQCQNSYMC 313

Query 772 VTCHYIDDDWKVQKRIIRFCVVKTPHDGPNLYTSMLTTFIKFYNIEDKLSITLDNAAPNN 951
          VT H+IDDDW +QKRI+ F V+ H G L + + +NIE KLF++LDNA+ N
Sbjct 314 VTIHWIDDDWCLQKRIVGFFHVEGRHTGQRLSQTFTTAIMVKWNIKKLFLALSDNASANE 373

Query 952 TMDILKSNLLKVDL-LHCDSDLFHIRCAAHVINLVKDLQAIDGVINNIESRVKYVRA 1128
          + + +L D L CD FH+RCA H++NL+ KDGL I G I I+ V V++
Sbjct 374 VAVHDIIEDLQDTSNLVCDGAFFHVRCAHILNLVAKDGLAVIAGTIEKIKAILVAVKS 433

Query 1129 SQSRKETFFDIIIVELGIRFSVPKIDVENHWNSTCDMIESAMPFKEAFLEVEKDD--SY 1302
          S + E E + + DV WNST M+ A+ +K A + L+ D Y
Sbjct 434 SPLQWELMKCASECDLQSKGISYDVSTWNSYMLRLDALYKPAIRLRLKTSDFPRYV 493

Query 1303 TYCPSSQDWERANAVCKLLKVFKAIEVVGSTSYPTLNLYFHEIWSIKHVLEEGFSLNE 1482
          CP +++W+ A + K LK F E++SGT Y T NL++ IK +++
Sbjct 494 AICPKAEWKMALTLFKCLKKFFDLTELLSGTQYSTANLFYKGFCEIKOLDIQWCVHEKF 553

Query 1483 TIAAMVSMQEKFKYWMISYFTNCVPVILDLWFKFGFIEFLKQAFGEYCKCpsp**sg 1662
          I M M EKFKRW +S V LD +K IEF +K+ G+ K
Sbjct 554 VIRRMVAMSEKFEKYWVSNIALAVACFLDPYKILIEFYMKKFGHDSYK----VHV 608

Query 1663 ssY*KIVQCLLCSNGRFFSSGDP-----WR*CDNCW*RSLLV*LDEHTSAKRKE----- 1809
          + +++ L +F+SS P ++ +L+ D+ E
Sbjct 609 DDFVRVIRKLY----QFYSSCSFSAFKTKTTTNSMDUTLMENEDDEFQNYLHELKDYDQ 664

Query 1810 -ENNEYDRYLDDLPCCDDESFDILDNWKMASKYPTLAAIACDILAVTASTVPSESASF 1986
          E+NE D+Y+ + L + FDIL WW+ ++YP L IA D+LA+ STV SESAFS
Sbjct 665 VESNELDKYMSFELLKHSQG-FDILSNWRGRVAEYPILTQIARDVLAIQVSTVASESASF 723

Query 1987 TSGHIINDHRTLAGSTVEAYCVFKIGYTA---GSSYLDVI--SIDNLSI 2124
          G +++ +R RL VEA K A G++Y + ++ LDS+
Sbjct 724 AGGRVVDYPYRNLGSEIVEALICTKDNVAASRGATYFPTMIGDLEVLDSV 774
```

```
>sp|P03010|TRAC9_MAIZE Putative AC9 transposase
Length=839

Score = 270 bits (689), Expect = 3e-71
Identities = 195/646 (30%), Positives = 297/646 (45%), Gaps = 63/646 (9%)
Frame = +1

Query 259 KPRKLRSTIW----KMDPIYQDGKVIQ----SRFKHCYEVFAVARTSGTSHMRRLHLEI 411
          + +K S +W K+++ K +Q F +C + GTS R HL
Sbjct 80 RAKKCTSDVWQHFTKKEIEVEVDGKKYVQVWGHCFNPNCKAKYRAEGHGTSGFRNHLRT 139

Query 412 CEPRLKMHDFVDKLSLVTTESAILTNWRFDPKLTCELVRLIVLHELFPFVEYEGFHS 591
          +K K + + + ++D ++ +L I++HE PF+ VE+E F
Sbjct 140 SHSLVK-GQLCLKSEKDHGKDINLIEPKYDEVVSLKKHLAIIMHEYPFNI VEHEYFVE 198

Query 592 YSASLNPLAETISRTTIKENCLAYKNHITALREMFENCNFRFSLTADLWTSNKNIGYMC 771
          + SL P SR T ++ ++ Y L ++ RFS T D+WTS +N YMC
Sbjct 199 FVKSLRPHFPFKSRVTARKYIMDLYLEEKEKLYGKLKDQVSRFSTTMDMWTSQCQNSYMC 258

Query 772 VTCHYIDDDWKVQKRIIRFCVVKTPHDGPNLYTSMLTTFIKFYNIEDKLSITLDNAAPNN 951
          VT H+IDDDW +QKRI+ F V+ H G L + + +NIE KLF++LDNA+ N
Sbjct 259 VTIHWIDDDWCLQKRIVGFFHVEGRHTGQRLSQTFTTAIMVKWNIKKLFLALSDNASANE 318

Query 952 TMDILKSNLLKVDL-LHCDSDLFHIRCAAHVINLVKDLQAIDGVINNIESRVKYVRA 1128
          + + +L D L CD FH+RCA H++NL+ KDGL I G I I+ V V++
Sbjct 319 VAVHDIIEDLQDTSNLVCDGAFFHVRCAHILNLVAKDGLAVIAGTIEKIKAILVAVKS 378

Query 1129 SQSRKETFFDIIIVELGIRFSVPKIDVENHWNSTCDMIESAMPFKEAFLEVEKDDSYTY 1308
          S + E E + + DV WNST M+ A+ +K A + L+ D
Sbjct 379 SPLQWELMKCASECDLQSKGISYDVSTWNSYMLRLDALYKPAIRLRLKTSDFPRYV 438

Query 1309 C-----PSSQDWERANAVCKLLKVFKAIEVVGSTSY 1404
          C P S W + K LK F E++SGT Y
Sbjct 439 CLNCCCTCHHYKFSINQMSIIVGTQVFLKFRSGRHL--TLFKCLKKFPDTELLSGTQY 496

Query 1405 PTLNLYFHEIWSIKHVLEEGFSLNETIAMVSMQEKFKYWMISYFTNCVPVILDLWF 1584
          T NL++ IK +++ I M M EKFKYW +S V LD +
Sbjct 497 STANLFYKGFCEIKOLDIQWCHEKFVIRRMVAMSEKFEKYWVSNIALAVACFLDPY 556

Query 1585 KFGFIEFLKQAFGEYCKCpsp**sgssY*KIVQCLLCSNGRFFSSGDP-----WR*CDN 1749
          K IEF +K+ G+ K + +++ L +F+SS P ++
Sbjct 557 KKILIEFYMKKFGHDSYK----VDDDFVRVIRKLY----QFYSSCSFSAFKTKTTTND 607

Query 1750 CW*RSLLV*LDEHTSAKRKE-----ENNEYDRYLDDLPCCDDESFDILDNWKMASK 1908
          +L+ D+ E E+NE D+Y+ + L + FDIL WW+ ++
Sbjct 608 SMDUTLMENEDDEFQNYLHELKDYDQVESNELDKYMSFELLKHSQG-FDILSNWRGRVAE 666

Query 1909 YPTLAAIACDILAVTASTVPSESASFSTSGHIINDHRTLAGSTVEA 2046
          YP L IA D+LA+ STV SESAFS G +++ +R RL VEA
Sbjct 667 YPILTQIARDVLAIQVSTVASESAGRGRVVDYPYRNLGSEIVEA 712
```

Figura 23 – Validação do TE221-RP através da comparação de sua sequência com banco de dados do arroz e “Swissprot” **A:** Identificação de um locus do cromossomo 7 do arroz com similaridade de 70% dentro de uma cobertura de 394 nucleotídeos. **B:** Comparação da sequência de nucleotídeos do TE221-RP e do locus do arroz no banco de dados do Swissprot. Para ambos a maior correspondência foi a transposase de Ac.

IV.2.5.2.3. Amplificação genômica de elemento SChAT do Grupo II a partir de suas TIRs

As TIRs (Repetições Invertidas Terminais) relativas a uma hipotética cópia genômica relacionada ao cDNA TE221 foram descritas no item IV.2.5.2.1. Elas localizam-se nas extremidades dos elementos e determinam o sítio de reconhecimento e de ligação da transposase. Admitindo-se que estas TIRs possam ser compartilhadas por outras variantes do elemento relacionado com o TE221, foram desenhados iniciadores sobre suas sequências, voltados para dentro, de modo a amplificar outras sequências do genoma da cana que

compartilhassem estas TIRs em suas extremidades. A amplificação foi realizada com os iniciadores MITE-hAT-1F e MITE-hAT-1R sobre DNA genômico da cana-de-açúcar, como descrito no item III.13.2. Como resultado foram obtidas três bandas, de 215 pb, 3,5 kpb e 4,2 kpb (figura 24-A). A banda de 215 pb foi sequenciada e comparada com o banco de dados de arroz no TIGR (“TIGR Rice Database”) através de BLASTN (figura 24-B). Esta sequência apresentou 62% de identidade de nucleotídeos com uma sequência de arroz anotada como similar ao domínio de dimerização *hAT* PF05699 do Pfam (figura 24-B), o que demonstrou que além de compartilhar as TIRs relativas ao TE221, ela apresenta em sua sequência de nucleotídeos um segmento defectivo, porém que ainda mantém a homologia com a superfamília *hAT*. Estas características permitem classificar tal sequência como um MITE (“miniature inverted-repeat transposable element”).

Figura 24 – Bandas amplificadas com os iniciadores MITE-hAT-1F e MITE-hAT-1R, desenhados sobre as TIRs descritas para o TE221. **A:** A PCR 1 amplificou uma banda de 250 pb, com Taq polimerase e um ciclo curto de extensão. A PCR 2 amplificou duas bandas, uma de 3,5 kb e outra de 4,2 kb, aproximadamente. **B:** A banda de 250 pb foi sequenciada e um BLASTN no genoma do arroz retornou uma sequência do tipo *hAT*.

As bandas de 4,2 kb e 3,5 kb foram sequenciadas em suas extremidades. Uma análise destas sequências parciais por BLASTN e BLASTX no Genbank revelou apresentarem homologia com cDNAs de cana-de-açúcar do grupo II. Este resultado corrobora a recuperação de elementos do genoma a partir de amplificação com os iniciadores desenhados sobre as TIRs. Estes elementos recuperados, de 4,2 kb e 3,5 kb foram denominados SChAT1 e SChAT2, respectivamente. Como estratégia para confirmar a homologia de SChAT1 e SChAT2 com o cDNA TE221, foi realizado um ensaio de hibridação molecular dos mesmos com a sonda interna do TE191. Esta sonda, descrita III.11.1.1 é compartilhada pelo TE221.

A figura 25 traz o resultado deste ensaio. No gel de agarose foram migradas as minipreparações de DNA plasmidial de SChAT1 e SChAT2, hidrolisadas com dois conjuntos de endonucleases. A hidrólise *EcoRI/PstI* libera os fragmentos do vetor de clonagem (raias 3 a 6). O cDNA TE221 apresenta um sítio de restrição para a endonuclease *HindIII*, na posição 1684. A hidrólise com *EcoRI/HindIII*, portanto, libera e cliva em pelo menos um sítio os fragmentos (raias 7 a 10). As setas vermelhas na autorradiografia indicam os fragmentos com os quais a sonda hibridou. Nas raias 3 e 7 encontram-se os fragmentos relativos a SChAT1, sendo observado na raia 3 o fragmento de 4,2 kb. Nas raias 4 e 8 encontram-se os fragmentos relativos a SChAT2, sendo observado na raia 4 o fragmento de 3,5 kb.

Figura 25 – Ensaio de hibridação molecular de sonda do grupo II (TE191) contra as minipreparações de DNA plasmidial de SChAT1 e SChAT2. As setas vermelhas no gel de agarose e na autorradiografia indicam os DNAs que apresentaram hibridação.

V. DISCUSSÃO

O impacto dos elementos de transposição na estruturação dos genomas pode ser melhor avaliado quando do sequenciamento de diversos genomas. Um exemplo foi a detecção de que a diferença entre o tamanho dos genomas de duas espécies de arroz, *Oriza australiensis* e *Oriza sativa*, é devida a um aumento no número de cópias de três famílias de retrotransposons com LTRs, ocorrido após a separação das duas espécies (Piegu *et al*, 2006). Utilizando estratégias como a análise de sequências de pontas de BACs no banco “Oryza Map Alignment Project” através de BLASTN, e o posterior sequenciamento de BACs inteiros de *O. australiensis*, foi possível reconstruir e validar a sequência putativa de três retrotransposons com LTRs do tipo Ty3/gypsy. Estes teriam tido um incremento quantitativo dentro do genoma de *Oriza australiensis* após a separação de *Oriza sativa*.

O projeto SUCEST (“Sugarcane Expressed Sequence Tag”) foi o primeiro projeto voltado para o sequenciamento de genes da cana-de-açúcar em larga escala. A estratégia de sequenciar parcialmente clones de cDNA de diferentes bibliotecas, gerando “tags” (sequências curtas de DNA), permitiu a identificação de sequências expressas e sua anotação. Estes aspectos foram importantes na facilitação do acesso à informação genética da cana-de-açúcar, um híbrido que apresenta um genoma complexo (Vettore *et al*, 2001). A partir do universo de “etiquetas” gerado pelo SUCEST foi possível realizar pela primeira vez uma análise global dos elementos de transposição presentes no genoma desta importante cultura tropical que é a cana-de-açúcar (Rossi *et al*, 2001).

V.1. Obtenção das Sequências Completas dos Clones de cDNA e Identificação das Famílias de Elementos Transponíveis Expressas em Cana-de-Açúcar

De um total de 291.689 clones de cDNA, 276 clones foram identificados como homólogos a elementos de transposição. O sequenciamento completo de 156 destes cDNAs, realizado como objetivo inicial deste trabalho, possibilitou uma análise detalhada destes elementos expressos no genoma da cana-de-

açúcar. Comparando-se a anotação de Rossi e colaboradores (2001) com a anotação das sequências inteiras, praticamente não houve mudanças neste panorama, o que dá suporte ao critério estrigente utilizado na seleção inicial. As famílias mais representadas de retrotransposons foram Hopscotch, Opie2, MAGGY, Stonor e Retrofit. Para os transposons, as famílias mais representadas encontradas foram MuDr, Ac e Tam3, que juntos compõem a superfamília *hAT*, *En/Spm*, *Tag1*, *Soymar* e *Tip100*, com ênfase para as três primeiras famílias, que juntas representaram mais de 80% dos cDNAs analisados em ambos os casos.

Os dados referentes à diversidade dos cDNAs foram incluídos como resultados na forma de tabela na publicação “Transcriptionally active transposable elements in recent hybrid sugarcane” *Plant Journal*, dec 2005, 44(5):707-17 (Araújo *et al*, 2004) (apêndice 2). Pode-se notar que o número de clones incluídos na tabela do artigo (tabela VI) é menor que o número de clones presentes na tabela V. Isto deve-se ao fato da tabela do artigo ter sido curada, de modo a retirar da análise os clones que pudessem corresponder a transcritos híbridos. Esta seleção foi bastante estrigente e visou evitar a descrição de sequências com expressão devido a um evento de leitura ininterrupta (“read through”). No entanto, as sequências abandonadas no artigo foram mantidas neste estudo, pois sabe-se que a transcrição de um gene por “read through” pode estar envolvida na regulação deste pelo mecanismo de interferência de RNA. Borsani e colaboradores (2005) encontraram dois genes que apresentam sobreposição de sequência antisenso, P5CDH relacionado a resposta a estresse, e SRO5, induzido pela presença de sal. Quando ambos os transcritos estão presentes ocorre um silenciamento pós transcricional de P5CDH, induzido pela formação de siRNAs que levam à clivagem dos transcritos de P5CDH. A expressão de SRO5 é necessária para iniciar o silenciamento. Os autores sugerem que os dois genes estão relacionados funcionalmente, e que esta forma de regulação por interferência de RNA pode se aplicar a outros genes que naturalmente apresentem sobreposição antisenso, nos genomas de eucariotos. Esta pode ser uma abordagem bastante interessante no sentido de estudar a regulação dos elementos de transposição da cana-de-açúcar. No entanto, depende do mapeamento e

sequenciamento das regiões genômicas que contém os TEs nos genomas de interesse.

V.2. Análise da Diversidade Entre os cDNAs Homólogos à Superfamília *hAT*

A caracterização da diversidade de famílias de TEs encontradas entre os clones de cDNA da cana-de-açúcar constituiu a primeira parte dos objetivos realizados neste trabalho. Após esta etapa, o foco das investigações passou a ser o conjunto de cDNAs homólogo à superfamília de transposons *hAT*, devido à sua importância histórica e por ser a segunda família mais representada entre os cDNAs, constituída por 32 homólogos.

Uma análise comparativa foi realizada com as sequências dos cDNAs de cana-de-açúcar homólogos a *hAT*, com o intuito de verificar a diversidade presente entre estas sequências. O alinhamento de dois domínios protéicos conservados característicos desta superfamília revelou um padrão de agrupamento destas sequências em dois grupos distintos (figura 8). O denominado grupo I é composto por sequências bastante conservadas entre si. As sequências do grupo I apresentaram 100% de similaridade de aminoácidos e uma identidade de nucleotídeos variando de 87% a 100% dentro do domínio alinhado. Este fato pode sugerir que os cDNAs deste grupo são um conjunto de transcritos provenientes de um número pequeno de cópias genômicas altamente conservadas. O resultado da hibridação molecular da sonda do grupo I (TE074) com DNA genômico dos parentais e híbridos da cana corrobora esta hipótese. A sonda detectou um baixo número de cópias dos genes relacionados com esta sonda (figura 13-A). O grupo II englobou sequências com maior diversidade de sequências, e agrupou-se com os elementos originais que compõem a superfamília *hAT*.

Do mesmo modo como foram caracterizados dois grupos de homólogos a *hAT* no genoma de cana, Rossi e colaboradores, em 2004, caracterizaram a existência de quatro classes de homólogos à proteína MURA (transposase) do transposon Mutator. Uma análise filogenética demonstrou que a separação destas classes teria ocorrido anteriormente à separação das monocotiledôneas

e eudicotiledôneas, uma vez que tanto sequências de cana como de arroz e *arabidopsis* agruparam-se em três das quatro diferentes classes observadas.

O padrão de hibridação molecular visualizado no genoma de cana e de outras espécies dá suporte à separação de dois grupos. Nota-se que enquanto o grupo I está presente em todas as gramíneas analisadas, incluindo arroz, milho e os parentais e híbridos de cana-de-açúcar, o grupo II parece ser detectado apenas no genoma de *Saccharum*. É plausível assumir que as transposases do tipo *hAT*, amplamente distribuídas entre plantas, animais e fungos, evoluíram verticalmente e possuem um ancestral comum (Rubin *et al*, 2001). Neste caso, as sequências das transposases do grupo II da cana-de-açúcar podem ter divergido rapidamente após a separação das gramíneas, ou seja, dentro de um intervalo de 60 milhões de anos (Kellog, 2001). Considerando-se que a cana-de-açúcar emergiu dentro de uma linhagem específica de *Saccharum* (Grivet *et al*, 2004), pode-se formular a hipótese de que a transposase do TE257 do grupo II, presente tanto nos parentais como nos híbridos da cana-de-açúcar, surgiu em um cultivar ancestral, anterior à diferenciação entre *S. officinarum* e *S. spontaneum*. De acordo com D'Hont (1996), a contribuição de *Saccharum spontaneum* para o número de cromossomos do híbrido é de apenas 15 a 25%. Sendo assim, é provável que haja sequências com taxas de evolução rápidas que encontrem-se presentes somente nos genomas de *S. officinarum* e dos híbridos, como foi o caso da transposase do TE191.

Os elementos de transposição foram considerados como DNA lixo, ou “genes egoístas” (Doolittle *et al*, 1980) por muito tempo. Porém, recentemente, eles têm sido considerados cada vez mais como peças chave para a manutenção e evolução dos genomas, e a interação destes elementos com seus genomas hospedeiros pode ser considerada uma coevolução. (Kidwell & Lisch, 2000). Uma importante função tem sido designada para os elementos de transposição: a de doadores de domínios protéicos na formação de novos genes. Um exemplo é o gene quimérico SETMAR, derivado da fusão do gene SET, uma metiltransferase que promove a metilação das histonas H3, com o transposon Hsmar1, ambos presentes em primatas. Esta fusão formou um novo gene presente unicamente em hominóides (Cordaux *et al*, 2006). Um ensaio de alteração de mobilidade eletroforética (“gel shift”) demonstrou haver

uma forte ligação entre a transposase MAR (região da transposase de Hsmar1 incorporada a SETMAR) de humanos e oligonucleotídeos referentes à sequência consenso das TIRs do transposon Hsmar1. Estas evidências sugerem que o domínio funcional da transposase de Hsmar1 capturado na formação do gene SETMAR foi o de ligação ao DNA. Uma vez que uma busca no genoma humano revelou haver aproximadamente 760 sequências homólogas ao sítio de ligação de 19pb de MAR, o recrutamento do domínio de ligação ao DNA pode ter fornecido um meio de direcionar a metilação das histonas H3 a sítios particulares no genoma.

Outro exemplo é o gene DAYSLEEPER, que é uma transposase domesticada descrita em plantas (Bundock & Hooykaas, 2005). Um ensaio de mono-híbrido demonstrou que a proteína codificada pelo mesmo é um fator de ligação ao motivo Kubox1, presente na região a montante do gene de reparo de DNA de arabidopsis Ku70. O motivo Kubox1 também está presente a montante de vários outros genes de plantas, e ensaios de mono-híbrido utilizando o promotor de outros genes demonstraram que a ligação da proteína DAYSLEEPER não é restrita ao promotor de Ku70, e pode estar envolvida na regulação de múltiplos genes. A análise da sequência do gene DAYSLEEPER revelou que ele possui domínios conservados característicos das transposases da superfamília *hAT* (Rubin *et al*, 2001), porém, estão ausentes aminoácidos considerados essenciais para a transposição do elemento Ac, assim como o sítio duplicado de inserção e as TIRs. Estes aspectos caracterizam um elemento “fóssil”, incapaz de se transpor e frequentemente silenciado transcricionalmente. No entanto, foram detectadas ESTs relativas a DAYSLEEPER, o que é indicativo da transcrição do gene. Frente às informações acima citadas os autores a propuseram que DAYSLEEPER pode ter adquirido uma função celular alternativa. Um ensaio de “gel shift” demonstrou que a proteína DAYSLEEPER se liga a dímeros, trímeros e tetrâmeros da sequência de Kubox1, e esta ligação é dependente da presença de íons zinco. Esta dependência se explicou pela subsequente identificação de um “BED-zinc finger” na proteína, que provê a capacidade de ligação a DNA na presença de zinco (Aravind 2000). Mutantes de arabidopsis homozigóticos e heterozigóticos para o gene DAYSLEEPER foram isolados, com o intuito de verificar a importância do mesmo para a planta. O gene foi interrompido pela

inserção de uma estrutura de repetidos T-DNAs. As plântulas homozigóticas para a inserção dos T-DNAs perderam completamente a expressão de DAYSLEEPER, e apresentaram um severo fenótipo recessivo, com crescimento lento, sem expansão dos cotilédones ou desenvolvimento de folhas ou órgãos florais normais. As raízes apresentaram-se curtas, com um excesso de pêlos radiculares e uma desorganização do meristema lateral. Após um período prolongado nas placas de crescimento apenas 1% das plântulas apresentaram o desenvolvimento de algum tecido fotossintetizante. Plântulas heterozigóticas apresentaram fenótipo normal, e a introdução de DAYSLEEPER sob o controle do promotor 35S em plântulas recessivas reconstituiu o fenótipo selvagem. Estes experimentos acumularam evidências suficientes para sugerir que DAYSLEEPER é o primeiro gene derivado de um transposon que desempenha uma função essencial para o crescimento e desenvolvimento de arabidopsis, e provavelmente atua como fator de transcrição sobre vários genes (Bundock & Hooykaas, 2005). Em seguida foram descritos homólogos a DAYSLEEPER no trigo e na cevada, w-gary1, w-gary2 e b-gary, que apresentam como características a ausência das TIRs, do sítio duplicado de inserção e de aminoácidos necessários para a transposição em suas sequências. Há também evidências de que, comparando-se a localização cromossômica de gary na cevada, no trigo e no arroz, este parece residir no mesmo locus há pelo menos 60 milhões de anos, ou seja, o mesmo estaria imóvel desde um período anterior à separação destas gramíneas. Com base nestas informações os autores sugerem que w-gary1, w-gary2 e b-gary são genes derivados de uma transposase *hAT* domesticada (Muehlbauer *et al*, 2006). A atualidade e impacto destas descobertas denotam a importância da identificação de homólogos a estes genes também na cana-de-açúcar e no arroz, e a necessidade de se comprovar se são mesmo transposases domesticadas através de ensaios *in vitro* e *in vivo*.

O alinhamento de sequências de cDNAs do grupos I e II de cana, de homólogos em arroz e arabidopsis, de DAYSLEEPER e dos transposons originais Ac e Tam3, bem como a árvore gerada a partir deste alinhamento, apresentada na figura 11, indicam a separação destas sequências em dois grandes ramos. Um ramo contendo os cDNAs do grupo I, os homólogos de arroz e arabidopsis e a transposase domesticada DAYSLEEPER, e o outro

ramo contendo os cDNAs do grupo II, os homólogos de arroz e arabidopsis e os transposons Ac e Tam3. Este padrão de agrupamento sugere a existência de um ancestral comum que divergiu em duas vertentes: uma seguiu um caminho evolutivo como transposon, que está representado em cana pelos cDNAs do grupo II, e outra passou por um processo de domesticação anterior à separação das classes Monocotiledonea e Eudicotiledonea, visto que há homólogos de arroz e arabidopsis agrupados neste ramo, juntamente com DAYSLEEPER. Esta transposase domesticada estaria representada na cana pelos cDNAs do grupo I.

A identificação do grupo I de cDNAs da cana-de-açúcar como homólogo ao gene derivado da transposase domesticada DAYSLEEPER levou a uma investigação com o objetivo de identificar homólogos a esta transposase também no genoma do arroz. Uma busca identificou dois loci que apresentaram alta homologia com o TE074. Um novo alinhamento foi então realizado, incluindo sequências representantes dos grupos I e II de cana, sequências dos dois homólogos ao TE074 no arroz, DAYSLEEPER, w-gary1, w-gary2, b-gary, e os transposons Ac e Tam3, apresentado na figura 12. Este alinhamento revelou que, dentro do segmento alinhado, DAYSLEEPER é mais próxima dos dois homólogos do arroz e dos cDNAs do grupo I de cana TE096, TE048 e TE074, chegando a compartilhar com estes últimos 68% de similaridade de sequência de aminoácidos. Os genes gary apresentaram similaridade com DAYSLEEPER, mas menos evidente que com os homólogos do arroz e os cDNAs do grupo I. A hipótese de que os cDNAs do grupo I podem tratar-se de transposases domesticadas pode explicar o padrão de agrupamento no alinhamento e o diferente padrão de distribuição dos grupos I e II observado no ensaio de hibridação molecular. Enquanto os cDNAs do grupo II estariam relacionados com elementos de transposição, os do grupo I seriam referentes a um gene derivado de uma transposase domesticada. Por exercerem uma função na célula e consequentemente estarem mais sujeitos a pressões seletivas, suas sequências se manteriam mais conservadas. Os transposons, por sofrerem menor pressão seletiva e portanto serem mais tolerantes a mutações, teriam suas sequências menos conservadas, como foi observado na hibridação molecular, em que a sonda do TE257 foi detectada

somente nos genomas de *Saccharum*, e a sonda do TE191 foi detectada somente nos genomas dos híbridos da cana e da variedade parental *Saccharum officinarum*.

A atividade transcricional, tanto de TEs do grupo I da cana-de-açúcar quanto do grupo II foi confirmada em resultados de macroarranjos de DNA (“macroarrays”). Os transcritos do grupo I apresentaram-se mais abundantes que os do grupo II nos diferentes tecidos. Apenas os transcritos TE017 e o TE203, ambos do grupo I, foram detectados em todos os quatro tecidos analisados. A identidade entre as sequências de nucleotídeos de TE017 e TE203 é de 99% dentro de uma cobertura de aproximadamente 1kb, contendo apenas SNPs de diferença. Esta identidade alta e o fato de se encontrarem expressos nos mesmos tecidos sugerem que podem tratar-se de dois alelos de um mesmo locus. Já os cDNAs TE203, expresso nos quatro tecidos analisados, e o TE207, expresso em todos os tecidos com exceção de folhas, apresentam 88% de identidade sobre parte de suas sequências de nucleotídeos. Este nível de identidade e o padrão distinto de expressão sugerem que provavelmente tratam-se de dois loci distintos, e não alelos do mesmo locus. A maior frequência de transcrição para os TEs do grupo I corrobora a hipótese da domesticação. Uma vez que os TEs encontram-se sob regulação transcricional em seus genomas hospedeiros, pode-se supor que uma transposase capturada para um gene com função celular apresentaria níveis mais altos de transcrição. Por outro lado, o tecido que apresentou a maior frequência de transcritos foi calo (CL), onde 9 dos 10 transcritos do tipo *hAT* foram detectados. Os resultados do experimento de Northern Blot revelaram haver uma maior expressão do grupo I em calos, em detrimento de folhas e raízes, o que corrobora o padrão observado no macroarray. É interessante notar que os homólogos do grupo II apresentam um padrão de expressão mais restrito a calo com exceção do TE007 (unicamente em AM) e TE144 (expresso em todos os tecidos menos na folha).

Os resultados citados acima sugerem que o genoma da cana-de-açúcar ancora dois tipos de transposases do tipo *hAT*, um relacionado com transposons legítimos e outro relacionado com transposases domesticadas. Saccaro Jr. e colaboradores (2007) descreveram recentemente resultados semelhantes aos aqui apresentados. Estudos realizados por Rossi e

colaboradores (2004) revelaram o agrupamento dos cDNAs de cana-de-açúcar homólogos ao transposon Mutator em quatro classes distintas, presentes também em arroz e arabidopsis, que divergiram antes da separação entre as monocotiledôneas e as eudicotiledôneas. O prosseguimento destes estudos visou obter uma melhor compreensão da distribuição, número de cópias e diversidade de sequências do tipo MuDr-A através de análise comparativa entre sequências de cana-de-açúcar e arroz (Saccaro Jr. *et al*, 2007). *Oryza sativa* foi a espécie escolhida por ter o seu genoma já inteiramente sequenciado, e pelo fato de ser um modelo vegetal em que elementos de transposição já foram extensivamente caracterizados, inclusive o sistema MuDr. Ensaio de hibridação molecular de DNA genômico de cana foram realizados, utilizando com sondas cDNAs inteiros representantes das quatro classes descritas em Rossi *et al* (2004). Foi também realizada uma busca em biblioteca comercial de BACs de cana, utilizando as mesmas sondas da hibridação molecular. Além disso, uma busca *in silico* foi realizada no banco de dados do arroz no “TIGR”, utilizando como sondas eletrônicas todas as sequências homólogas a Mutator utilizadas nos estudos filogenéticos descritos em Rossi *et al* (2004). Como resultado destes experimentos foi verificado que o número de cópias é classe-específico e conservado entre cana e arroz. A classe II é a mais representada, seguida da classe I, e as classes III e IV apresentam um baixo número de cópias. Subsequentes análises foram realizadas no banco de dados do arroz do “TIGR”, visando avaliar a presença de TIRs nos elementos das classes I e II. De 50 sequências da classe I analisadas, apenas 9 não apresentaram TIRs. Para a classe II, 20 das 386 sequências analisadas não apresentaram TIRs. As classes III e IV não apresentaram TIRs. A associação das TIRs com as sequências da proteína MURA (transposase de Mutator) encontradas, permitiu a identificação de 17 transposons *bona fide* de classe I, e 202 de classe II. Buscas por homologia de sequência revelaram que as classes III e IV são homólogas à transposase domesticada MUSTANG, enquanto nenhuma das sequências das classes I e II apresentou identidade significativa com MUSTANG. Estes dados sugerem que as classes I e II, tanto de cana-de-açúcar quanto de arroz, correspondem a transposons *bona fide*, enquanto as classes III e IV correspondem a genes contendo transposases domesticadas do tipo MUSTANG.

Esta análise realizada por Saccaro Jr. e colaboradores (2007) indica que o sistema Mutator em gramíneas é composto por transposons *bona fide* e transposases domesticadas, que seguiram caminhos evolutivos distintos antes da separação de monocotiledôneas e eudicotiledôneas. Com base nestes dados é plausível sugerir que a domesticação de transposases, ou seja, o recrutamento de domínios protéicos de transposases na formação de novos genes é um fenômeno anterior à divergência entre mono e eudicotiledôneas. Eventos de domesticação devem ter ocorrido também com outras transposases, como as do tipo *hAT*, como corroborado pelas evidências obtidas por este trabalho. Cabe agora identificar a função potencial destas transposases a fim de consolidar a hipótese da domesticação.

V.3. Identificação de Potenciais TIRs e consequente isolamento de uma cópia de SChAT a partir do genoma de cana de açúcar

Através do ensaio de PCR-Inversa foi possível identificar as TIRs referentes à potencial cópia genômica relacionada ao cDNA TE221. Estas TIRs se estendem por 21 pb, e apresentam 15 pb reverso-complementares.

Dentro das sequências amplificadas no ensaio de PCR-Inversa era esperada a presença de determinadas estruturas, como esquematizado na figura 17-A. A partir da sequência dos iniciadores “forward” e “reverse” localizados nas extremidades das sequências dos amplicons, deveriam estar presentes as regiões subterminais não transcritas dos cDNAs, seguidas das TIRs, dos sítios duplicados de inserção e das regiões genômicas externas ao elemento. O sítio da endonuclease de restrição utilizada para hidrolisar o DNA genômico deveria ser encontrado na junção entre as porções referentes às regiões externas ao elemento a 3’ e a 5’. Nas sequências dos amplicons obtidos, cujo alinhamento está esquematizado na figura 17-B, foram localizados os iniciadores da PCR-Inversa, bem como as regiões subterminais do cDNA. Porém, não foi localizado o sítio da endonuclease de restrição utilizada para clivar o DNA genômico, que posteriormente foi circularizado formando assim anéis de DNA. Esta ausência do sítio da endonuclease de restrição pode ser explicada pela hipótese de que as sequências foram amplificadas a partir do ancoramento dos iniciadores “forward” e “reverse” nas

regiões subterminais de duas cópias vizinhas do mesmo elemento, contidas dentro do mesmo anel de DNA genômico. Neste caso a porção amplificada seria a localizada entre as duas cópias do elemento, e não a união das porções a montante e a jusante de uma única cópia do elemento contida no anel de DNA genômico, como esperado inicialmente (figura 26).

Figura 26 – Possíveis formas de amplificação de bandas no ensaio de PCR-Inversa. **A:** Amplificação a partir de uma única cópia do elemento localizada no anel de DNA genômico utilizado como molde. Neste caso o sítio da endonuclease de restrição pode ser identificado no amplicon. **B:** Amplificação a partir de duas cópias vizinhas do elemento localizadas no mesmo anel de DNA genômico. Neste caso o sítio da endonuclease de restrição não está presente no amplicon. Em ambos os casos as sequências dos iniciadores, as sequências subterminais 5' e 3' do cDNA, bem como as TIRs e as regiões genômicas externas às TIRs estão presentes no amplicon.

Com base nos resultados apresentados, elaboramos um modelo sobre a derivação dos transposons do tipo *hAT*, em que propomos que a domesticação ocorreu anteriormente à separação das classes Monocotiledônea e Eudicotiledonea, há pelo menos 60 milhões de anos atrás (figura 27).

Figura 27 – Modelo sugerido para a derivação dos transposons do tipo *hAT*.

V.4. Conclusões

1. A partir do sequenciamento dos clones de cDNA foi confirmada a similaridade dos mesmos com elementos de transposição de fato.
2. Os clones de cDNA com similaridade à família *hAT* estão organizados em dois grupos, em função da similaridade interna das proteínas preditas.
3. Os membros do grupo I estão presentes em baixo número de cópias e estão conservados na família *Poaceae*. Sugere-se que estes sejam transposons domesticados uma vez que além da conservação e baixo número de cópias genômicas, são expressos em diferentes tecidos, em níveis facilmente detectáveis por diferentes métodos.
4. Os membros do grupo II estão presentes em maior número de cópias no gênero *Saccharum* que os membros do grupo I. Apresentam baixo nível de transcrição.
5. Foi possível recuperar a partir de TIRs isoladas, três versões genômicas que constituem membros da família SChAT. Uma delas representa elementos do tipo MITE, e as outras duas, versões com tamanho compatível a um elemento completo.

VI. BIBLIOGRAFIA

- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. J. Mol. Biol., 215:403-410. 1990.
- ARAUJO, P.G.; ROSSI, M.; de JESUS, E.M.; SACCARO, N.L.Jr.; KAJIHARA, D.; MASSA, R.; de FELIX, J.M.; DRUMMOND, R.D.; FALCO, M.C.; CHABREGAS, S.M.; ULIAN, E.C.; MENOSSI, M.; VAN SLUYS, M.A. Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J., 44(5): 707-717. 2005.
- ARAVIND, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem. Sci., 25(9): 421-423. 2000.
- BAKER, B.; SCHELL, J.; LORZ, H.; FEDOROFF, N. Transposition of the maize controlling element "Activator" in tobacco. Proc. Natl. Acad. Sci. U S A. Jul;83(13):4844-4848. 1986.
- BAZIN, C.; DENIS, B.; CAPY, P.; BONNIVARD, E.; HIQUET, D. Characterization of permissivity for hobo-mediated gonadal dysgenesis in *Drosophila melanogaster*. Mol. Gen. Genet., 261(3):480-486. 1999.
- BECKER, A.H.; KUNZE, R. Maize Activator transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. Mol.Gen.Genet., 254: 219-230. 1997.
- BENNETZEN, J.L. The contributions of retroelements to plant genome organization, function and evolution. Trends in Microbiology, 4(9):347-353. 1996.
- BENNETZEN, J.L. Transposable element contributions to plant gene and genome evolution. Plant Molecular Biology, 42:251-269. 2000.
- BLACKMAN, R.; GELBART, W.M. The transposable element *hobo* of *Drosophila melanogaster*. In BERG, D.E. and HOWE, M.M. (ed) Mobile DNA. American Society of Microbiology. Washington DC. 1989.
- BORSANI, O.; ZHU, J.; VERSLUES, P.E.; SUNKAR, R.; ZHU, J.K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell, 123(7): 1279-1291. 2005.
- BROOKFIELD, J.F.Y. The ecology of the genome – mobile DNA elements and their hosts. Nature Reviews Genetics, february 6: 128-136. 2005.
- BUNDOCK, P.; HOOYKAAS, P. An Arabidopsis hAT-like transposase is essential for plant development. Nature, 436(7048):282-284. 2005.

- CARPENTER, R.; MARTIN, C.; COEN, E. Comparison of genetic behaviour of the transposable element Tam3 at two unlinked pigment loci in *Antirrhinum majus*. Mol. Gen. Genet., 207:82-89. 1987.
- CASACUBERTA, J.M.; SANTIAGO, N. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene, 311:1-11. 2003.
- CESNIK, R.; MIOCQUE, J. Melhoramento da cana-de-açúcar. Embrapa Informação Tecnológica, Brasília, DF. 2004.
- CHEN, Z.J.; NI, Z. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays, 28: 240-252. 2006.
- COHEN, E.S.; ROBBINS, T.P.; ALMEIDA, J.; HUDSON, A.; CARPENTER, R. Consequences and mechanisms of transposition in *Antirrhinum majus*. In BERG, D.E. and HOWE, M.M. (ed) Mobile DNA. American Society of Microbiology. Washington DC. 1989.
- COWAN, K.R.; HOEN, D.R.; SCHOEN, D.J.; BUREAU, T.E. MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol. Biol. Evol., 22(10): 2084-2089. 2005.
- CORDAUX, R.; UDIT, S.; BATZER, M.A.; FESCHOTTE, C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. PNAS, 103(21):8001-8006. 2006.
- DEININGER, L.P. SINEs: short interspersed repeated DNA elements in higher eucaryotes. In Mobile DNA Berg D.E, Howe M.M. eds. p 619-636. American Society for Microbiology, Washington D.C. USA. 1989.
- D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol. Gen. Genet., 250:405-413. 1996.
- ESSERS, L.; ADOLPHS, R.T.; KUNZE, R. A highly conserved domain of the maize Activator transposase is involved in dimerization. The Plant Cell, 12:211-223. 2000.
- FEDOROFF, N.V.; WESSLER, S.; SHURE, M. Isolation of the transposable maize controlling elements Ac and Ds. Cell, 35(1):235-242. 1983.
- FEDOROFF, N.V. Maize transposable elements. In BERG, D.E. and HOWE, M.M. (ed) Mobile DNA. American Society of Microbiology. Washington DC. 1989.
- FESCHOTTE, C.; JIANG, N.; WESSLER, S.R. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*, may 3(5):329-341. 2002.

- FESCHOTTE, C.; ZHANG, X.; WESSLER, S.R. Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In Craig, N.L.; Craigie, R.; Gellert, M. e Lambowitz, A.M. (eds), Mobile DNA II. American Society for Microbiology Press, Washington DC, pp. 1147-1158. 2002.
- FONG, A.M.V.A.; JUDELSON, H.S. The hAT-like DNA transposon DodoPi resides in a cluster of retro and DNA transposons in the stramenopile *Phytophthora infestans*. Mol. Gen. Genomics, 271: 577-585. 2004.
- FUJINO, K.; SEKIGUCHI, H.; KIGUCHI, T. Identification of an active transposon in intact rice plants. Mol. Gen. Genomics, 273: 150-157. 2005.
- GRANDBASTIEN, M.A. Stress activation and genomic impact of plant retrotransposons. J. Soc. Biol. 198:425-32. 2004.
- GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J.C. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics, 142(3):987-1000. 1996.
- GRIVET, L.; DANIELS, C.; GLASZMANN, J.C.; D'HONT, A. A review of recent molecular genetics evidence for sugarcane evolution and domestication. Ethnobotany Research & Applications. 2:9-17. 2004.
- HAN, J.S.; BOEKE, J.D. A highly active synthetic mammalian retrotransposon. Nature, 429(6989): 314-318. 2004.
- HANAHAN, D. Studies on transformation of *E. coli* with plasmids. J. Mol. Biol., 156:557-580. 1983.
- HARRISON, B.J.; FINCHAM, J.R.S. Instability at the pal locus in *Antirrhinum majus*: I. Effects of environment on frequencies of somatic and germinal mutation. Heredity, 19:237-258. 1964.
- HASHIDA, S.; KITAMURA, K.; MIKAMI, T.; KISHIMA, Y. Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*. Plant Physiology, 132:1207-1216. 2003.
- HEHL, R.; NACKEN, W.K.; KRAUSE, A.; SAEDLER, H.; SOMMER, H. Structural analysis of Tam-3, a transposable element from *Antirrhinum majus*, reveals homologies to the Ac element from Maize. Plant Molecular Biology, 16(2):369-371. 1991.
- KALENDAR, R.; TANSKANEN, J.; IMMONEN, S.; NEVO, E.; SCHULMAN, A.H. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. U. S. A. 97:6603-7. 2000.

- KASHKUSH, K.; FELDMAN, M.; LEVY, A.A. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics, 160: 1651-1659. 2002.
- KELLOG, E. Evolutionary history of the grasses. Plant Physiology. 125:1198-1205. 2001.
- KEMPKEN, F.; WINDHOFER, F. The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma, 110(1):1-9. 2001.
- KIDWELL, M.; LISCH, D. Transposable elements and host genome evolution. Trends Ecol. Evol., 15(3): 95-99. 2000.
- KUMAR, A, BENNETZEN, J.L. Plant Retrotransposons. Ann. Rev. Genet. 33: 479-532. 1999.
- KUMAR, S.; TAMURA, K.; NEI, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform., 5(2):150-163. 2004.
- KUNZE, R.; STOCHAJ, U.; LAUFS, J.; STARLINGER, P. Transcription of transposable element Activator (Ac) of *Zea mays* L. EMBO J. Jun;6(6):1555-1563. 1987.
- LEWIN, B. Genes VIII. Oxford University Press. New York. 2004.
- LINDER-BASSO, D.; FOGLIA, R.; ZHU, P.; HILLMAN, B.I. Crypt1, na active Ac-like transposon from the chesnut blight fungus, *Cryphonectria parasitica*. Mol Genet Genomics, July, 265: 730-738. 2001.
- LONNIG, W.E., SAEDLER, H. Plant transposons: contributors to evolution? Gene, 205:245-53. 1997.
- MADLUNG, A.; MASUELLI, R.W.; WATSON, B.; REYNOLDS, S.H.; DAVISON, J.; COMAI, L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol, 129: 733-746. 2002.
- MADLUNG, A.; TYAGI, A.P.; WATSON, B.; JIANG, H.; KAGOCHI, T.; DOERGE, R.W.; MARTIENSSEN, R.; COMAI, L. Genomic changes in synthetic *Arabidopsis* polyploids. The Plant Journal, 41(2): 221-230. 2005.
- McCLINTOCK, B. Maize genetics. Year Book Carnegie Inst. Washington. 45:176-186. 1946.
- McCLINTOCK, B. Mutable loci in maize. Carnegie Inst. Washington Yearbook 47:155-169. 1948.

- McCLINTOCK, B. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol., 21:197-216. 1956.
- McCLINTOCK, B. The significance of responses of the genome to challenge. Science, 226:792-801. 1984.
- McGINNIS, W.; SHERMOEN, A.W.; BECKENDORF, S.K. A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. Cell, 34:75-84. 1983.
- MING, R.; LIU, S.-C.; LIN, Y.-R.; da SILVA, J.; WILSON, W.; BRAGA, D.; van DEYNZE, A.; WENSLAFF, T.F.; WU, K.K.; MOORE, P.H.; BURNQUIST, W.; SORRELLS, M.E.; IRVINE, J.E.; PATERSON, A.H. Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics, 150: 1663-1682. 1998.
- MUEHLBAUER, G.J., BHAT, B.S., SYED, N.H., HEINEN, S., CHO, S., MARSHALL, D., PATEYRON, S., BUISINE, N., CHALHOUB, B., FLAVELL, A.J. A hAT superfamily transposase recruited by the cereal grass genome. Mol Genet Genomics, Jun;275(6):553-63. 2006.
- ORGEL, L.F. CRICK, F.H. Selfish DNA: the ultimate parasite. Nature, 393: 68-72. 1980.
- PEREPELITSA-BELANCIO, V.; DEININGER, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat. Genet., 35(4): 363-366. 2003.
- PIEGU, B.; GUYOT, R.; PICAULT, N.; ROULIN, A.; SANIYAL, A.; KIM, H.; COLLURA, K.; BRAR, D.S.; JACKSON, S.; WING, R.A.; PANAUD, O. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. Oct;16(10):1262-9. 2006.
- POHLMAN, R.F.; FEDOROFF, N.V.; MESSING, J. The nucleotide sequence of the maize controlling element activator. Cell, 37:635-643. 1984.
- PRITHAM, E.J.; ZHANG, Y.H.; FESCHOTTE, C.; KESSELI, R.V. An Ac-like transposable element family with transcriptionally active Y-linked copies in the white campion, *Silene latifolia*. Genetics, 165: 799-807. 2003.
- ROSSI, M.; ARAÚJO, P.G.; VAN SLUYS, M.A. Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Genet. and Mol. Biology, 24(1-4):147-154. 2001.
- RUBIN, E.; LITHWICK, G.; LEVY, A.A. Structure and evolution of the *hAT* transposon superfamily. Genetics, 158(3):949-57. 2001.

- SACCARO JR, N.L.; VAN SLUYS, M.A.; VARANI, A.M.; ROSSI, M. Mudr-A-like sequences from rice and sugarcane cluster as two *bona fide* transposon clades and two domesticated transposases. Gene, 392(1-2):117-125. 2007.
- SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., 4(4):406-425. 1987.
- SANMIGUEL, P., GAUT, B.S., TIKHONOV, A., NAKAJIMA, Y., BENNETZEN, J.L. The paleontology of intergene retrotransposons of maize. Nature Genetics, :20: 43-5. 1998.
- SCORTECCI, K.C.; DESSAUX, Y.; PETIT, A.; VAN SLUYS, M.A. Somatic excision of the Ac transposable element in transgenic *Arabidopsis thaliana* after 5-azacytidine treatment. Plant Cell Physiol., 38(3):336-343. 1997.
- SLOTKIN, R.K.; FREELING, M.; LISCH, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nature Genetics, 37(6): 641-644. 2005.
- SMYTH, D.R. Plant Retrotransposons. In Control of Gene Expression. Cap. 1. Verna, P. (ed.). CRC Press. New York. 1993.
- SNUSTAD, P.D.; SIMMONS, M. Principles of Genetics. Ed. John Wiley & Sons. Nova Iorque. 1997.
- STRECK, R.D.; MACGAFFEY, J.E.; BECKENDORF, S.K. The structure of *hobo* transposable element and their insertion sites. EMBO J., 5:3615-3623. 1986.
- TATUSOVA, T.A.; MADDEN, T.L. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS., 174(2):247-250. 1999.
- TELLES, G.P.; DA SILVA, F.R. Trimming and clustering sugarcane ESTs. Genet. Mol. Biol. 24:17-23. 2001.
- TELLES, G. P.; BRAGA, M. D. V.; Dias, Z.; Lin, T-L .; Quitzau, J.A.A. ; da Silva, F. R. ; Meidanis, J. Bioinformatics of the sugarcane EST project. Genet. Mol. Biol. 24:9 – 15. 2001.
- THOMAS, C.A. The genetic organization of chromosomes. Ann.Rev.Genet., 5: 237-256. 1971.
- THOMPSON, J.D.; GIBSON, T.J.; PLEWNIAK, F.; JEANMOUGIN, F.; HIGGINS, D.G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24:4876-4882. 1997.
- VAN SLUYS, M.A., TEMPÉ, J., FEDOROFF, N. Studies on the introduction and mobility of the maize Activator element in *Arabidopsis thaliana* and *Daucus carota*. EMBO J. 6:3881–3889. 1987.

- VAN SLUYS, M.A.; SCORTECCI, K.C.; TEMPÉ, J. DNA methylation associated to Ac element in Arabidopsis. Plant Physiol. Biochem., 31(6): 805-813. 1993.
- VETTORE, ANDRÉ L.; SILVA, F.R.; KEMPER, E.L.; ARRUDA, P. The libraries that made SUCEST. Genet. Mol. Biol., 24(1-4):1-7. 2001.
- VICIENT, C.M., SCHULMAN, A. H. Copia-like retrotransposons in the rice genome: few and assorted. Genome Lett., 1: 35-47. 2002.
- VINCENTZ, M. *et al.*. Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. Plant Physiol., 134: 951-959. 2004.
- VOYTAS, D.F., NAYLOR, G.J. Rapid flux in plant genomes. Nat. Genet. 20:6-7. 1998.
- WARREN, W.D.; ATKINSON, P.W.; O'BROCHTA, D.A. The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, Ac, and Tam3 (*hAT*) element family. Genet. Res. Camb., 64:87-97. 1994.
- WRIGHT, D.A; VOYTAS D.F. Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. Genome Res. 12:122-31. 2002.
- XIONG Y, EICKBUSH TH Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J., 9: 3353-62. 1990.
- XU, Z.; DOONER, H.K. Mx-rMx, a family of interacting transposons in the growing hAT superfamily of Maize. The Plant Cell, 17: 375-388. 2005.
- ZHANG, J.; MADDEN, T.L. Power BLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res., 7(6):649-656. 1997.