NILO LUIZ SACCARO JUNIOR

**O sistema *Mutator* em cana-de-açúcar: uma
análise comparativa com arroz**

São Paulo

2007

Resumo

Os elementos transponíveis (TEs) constituem grande parte do material genético de diversos eucariotos, alcançando entre 50-80% do genoma de gramíneas. Os projetos genoma proporcionaram um aumento das informações disponíveis sobre estes elementos, o que evidenciou sua importância e possibilitou o desenvolvimento de novas abordagens para seu estudo. O sistema *Mutator* (*Mu*) de milho é o mais ativo e mutagênico transposon de plantas. Além do elemento autônomo, *MuDR*, o sistema compreende ainda um conjunto de elementos bastante heterogêneo em sua sequência e estrutura, chamados *MuLEs*, que podem conter até mesmo fragmentos de genes do hospedeiro. As sequências de transposons mais abundantemente expressas no transcriptoma de cana-de-açúcar são relacionadas a *MuDR* e se agrupam em quatro clados (nomeados Classes I, II, III e IV), existentes antes da divergência entre Mono e Eudicotiledôneas. O trabalho apresentado aqui teve o objetivo de aprofundar o conhecimento sobre o sistema *Mutator* em cana-de-açúcar a partir da análise comparativa entre sequências dessa planta e de arroz (cujo genoma está totalmente sequenciado). Foi possível avaliar a abundância e diversidade do sistema *Mu* em gramíneas, ficando evidente uma amplificação de elementos clado-específica, tendo a Classe II sofrido uma explosão no número de cópias ao longo da evolução destas plantas. Análises estruturais revelaram que, enquanto as Classes I e II compreendem elementos com características de transposons, as Classes III e IV são, na verdade, transposases domesticadas. Foram completamente sequenciados dois clones de BAC de cana-de-açúcar, um proveniente de cada parental do híbrido (*Saccharum officinarum* e *Saccharum spontaneum*), ambos contendo elementos da Classe III. Estes elementos foram caracterizados e a sequência genômica de cana foi comparada com sua ortóloga em arroz, revelando um acúmulo de TEs nas regiões intergênicas.

Abstract

Transposable elements (TEs) constitute great part of eukaryote genetic material, in grasses, they comprise between 50-80% of the genome. Genome projects have significantly increased the amount of information about these elements, revealing their importance and allowing the development of new approaches for their study. The *Mutator* system (*Mu*) of maize is the most active and mutagenic plant transposon. Beyond the autonomous element, *MuDR*, the system comprises a very heterogeneous, in sequence and structure, set of elements, called *MuLEs*, that can contain even host gene fragments. The most abundant transposon related sequences expressed in sugarcane transcriptoma are the *MuDR*-like. They group into four clades (called Classes I, II, III and IV) that exist prior to the Mono and Eudicot split. The aim of this work is to gain knowledge about the *Mutator* system in sugarcane through the comparative analysis against rice (whose genome is completely sequenced). The results described the abundance and diversity of the *Mu* system in grasses, evidencing a clado-specific amplification with a burst of Class II along the evolution of this plant group. Structural analyses showed that, while Classes I and II comprise elements with transposon characteristics, Classes III and IV are domesticated transposases. One BAC clone from each sugarcane parental genotype (*Saccharum officinarum* and *Saccharum spontaneum*) have been completely sequenced, both containing Class III elements. These elements have been characterized and the sugarcane genomic sequences were compared with their orthologues in rice. The comparative analyses showed an accumulation of TEs in the intergenic regions.

1. Introdução

1.1. Elementos transponíveis

Os elementos transponíveis (TEs) foram originalmente caracterizados em plantas, devido ao seu envolvimento na origem de mutações instáveis, fato observado através de estudos com milho (*Zea mays*) pela citogeneticista Barbara McClintock na década de 40. Foi ela quem primeiro definiu o conceito de transposição e denominou-os “elementos controladores do gene” (McClintock 1956). Desde então, inúmeros TEs têm sido identificados em uma grande variedade de organismos que vai de bactérias a eucariotos superiores (Kidwell e Lisch 2002, Kazazian 2004). A característica fundamental de todos eles é a capacidade de se inserir em diferentes posições no genoma, alterando muitas vezes a função dos genes a que se associam (Grandbastien 1992). Hoje, sabe-se que os TEs constituem grande parte do material genético de diversos eucariotos, alcançando 45% do genoma humano e variando entre 50-80% do genoma de algumas gramíneas (Feschotte *et al.* 2002). Apesar da origem e do papel biológico dessas seqüências móveis permanecerem desconhecidos, acredita-se que elas geram variabilidade genética e que, portanto, possuem grande importância para a plasticidade do genoma necessária à evolução (Flavell *et al.* 1997, Evgen'ev 2007). O efeito deletério da transposição pode ser minimizado pela manutenção de elementos em estado quiescente durante o crescimento e desenvolvimento normais da planta hospedeira. Por outro lado, a ativação desses elementos em situações de estresse aumentaria a taxa de mutação, possibilitando uma reestruturação do genoma, eventualmente aproveitada pela seleção natural de forma a melhorar a adaptação do organismo (McClintock 1956 e 1984, Wessler 1996).

Os TEs de eucariotos dividem-se em duas classes: transposons e retroelementos. Os primeiros mobilizam-se geralmente de maneira não replicativa, de um *locus* para outro, *via* uma molécula de DNA. Os retroelementos transpõem-se de forma replicativa através de um intermediário de RNA, havendo a cada ciclo de transposição, portanto, o aumento no número de cópias no genoma hospedeiro (Kidwell e Lisch 2002). Todos os TEs geram uma duplicação do sítio de inserção (TSD, target site duplication) no genoma (Grandbastien *et al.* 1994). Dentro de cada classe, existe uma grande diversidade de clados de elementos. Por exemplo, em gramíneas foram descritas pelo menos 17 clados de transposons e 43 de retroelementos (Rossi *et al.* 2001).

Os retroelementos estão divididos em dois grupos: retroelementos com LTRs (Longas Terminações Repetidas, na mesma direção) e retroelementos sem LTRs (Fig.1). O grupo com LTRs é composto de retrovírus e retrotransposons, que possuem uma organização similar. A diferença reside no fato dos retrotransposons não apresentarem, diferentemente dos retrovírus, o domínio *env*, que codifica uma glicoproteína capaz de formar o envelope protéico da partícula viral, conferindo assim o caráter infeccioso (Fosket 1994, Schmidt 1999). Os retrotransposons podem ainda ser separados em dois grupos, *Ty1/copia* e *Ty3/gypsy* (segundo a nomenclatura original de levedura e *Drosophila*), de acordo com a posição dos domínios protéicos *INT* e *RT-RNaseH* dentro da poliproteína *pol*. Os retroelementos sem LTR se dividem em LINEs ("Long Interspersed Nuclear Elements") e SINEs ("Short Interspersed Nuclear Elements") (Fig. 1). Os LINEs são elementos autônomos que codificam todas as enzimas necessárias para sua transposição, enquanto os SINEs, contendo apenas poucas centenas de pares de bases, são elementos defectivos que necessitam das enzimas em *trans* para completar seu ciclo replicativo

(Smyth 1993, Schmidt 1999). Existem evidências que indicam que os SINEs parasitam a maquinaria de transposição dos LINEs (Ogiwara *et al.* 1999).

Figura 1. Estrutura e classificação dos elementos de transposição.

As duplicações do sítio de inserção estão indicadas por setas. LTR: longas repetições terminais. TIR: repetições terminais invertidas. *Gag* e *pol* são os dois quadros abertos de leitura dos retroelementos com LTRs. *ORF1*, *EN* e *RT* são domínios protéicos dos retroelementos sem LTRs (Adaptado de

Os transposons de plantas têm repetições terminais invertidas (TIRs) (Fig. 1), onde, em alguns casos, se localiza a região promotora, e movem-se excisando-se e inserindo-se em uma nova localização (Flavell *et al.* 1994). São altamente variáveis em tamanho. Os elementos autônomos apresentam um gene que codifica para uma enzima, a transposase, responsável pela excisão e inserção do elemento. Os transposons não autônomos geralmente derivam de deleções internas dos autônomos, podendo também ser mobilizados através da maquinaria fornecida por estes (Miskey *et al.* 2007).

A atividade dos TEs é regulada através de complexos mecanismos de controle, que garantem a viabilidade do hospedeiro e assim, a perpetuação dos próprios elementos (Grandbastien 1998). Por esse motivo, embora existam milhões de elementos no genoma dos

eucariontes, aparentemente só uns poucos são ativos e a tarefa para determinar se certo elemento está ativo, inativo ou silenciado epigeneticamente torna-se um grande desafio (Feschotte *et al.* 2002). A transcrição é o primeiro ponto de controle da atividade dos TEs (Grandbastien *et al.* 2005). Por outro lado, a transcrição de um TE não se correlaciona necessariamente com a presença de novas inserções já que, para completar o ciclo transposicional, além da transcrição é necessário ocorrer a tradução do mensageiro, transcrição reversa no caso dos retrotransposons, e a integração em um novo local. A regulação em qualquer destes níveis limita a transposição (Slotkin e Martienssen 2007).

1.2. O sistema *Mutator*

O transposon *Mutator* (*Mu*) foi originalmente identificado em milho (Robertson 1978) e utilizado para a identificação de genes através de “transposon tagging” nesta espécie (Walbot 1992). Em milho, o sistema *Mutator* possui um mecanismo de regulação complexo envolvendo silenciamento gênico pós-transcricional (PTGS) e transcricional (TGS) (Rudenko *et al.* 2003). O elemento regulatório autônomo, *MuDR*, apresenta entre 5 a 30 cópias nas linhagens de milho *Mutator*. Este elemento possui dois genes que são transcritos em sentido inverso a partir dos promotores contidos nas TIRs (terminal inverted repeats), *mudrA* e *mudrB*. O produto de *mudrA*, MURA, tem homologia com transposases bacterianas e encontra-se presente em diversas espécies vegetais. Já o outro gene, *mudrB*, codifica para a proteína MURB cuja função ainda não está esclarecida. Embora em milho MURB seja necessária para ocorrer transposição, o gene *mudrB* tem sido identificado apenas em elementos das espécies do gênero *Zea* (Lisch *et al.* 2001, Rossi *et al.* 2004). A figura 2 esquematiza a estrutura de *MuDR* apresentando a

organização dos genes *mudrA* e *mudrB* e as TIRs. Indicam-se também os transcritos com seus íntrons e éxons, os produtos protéicos e seus tamanhos correspondentes.

Figura 2. Estrutura do elemento autônomo *MuDR* de milho.

O sistema *Mutator* compreende, além dos elementos autônomos *MuDR*, um conjunto de elementos bastante heterogêneo em sua seqüência e estrutura (Fig. 3). Estas seqüências, denominadas *MuLEs* (*Mutator-like elements*), podem ou não possuir TIRs em seus extremos, assim como homologia com os genes *mudrA* e *mudrB*. Quando *MuLEs* apresentam homologia com *mudrA* ou *mudrB*, são denominados *hMuDRs*. Embora tais elementos não sejam responsáveis pela atividade de *Mutator*, evidências indicam que estão envolvidos na regulação do sistema, competindo com os elementos autônomos pelas transposases (Lisch 2002). *MuLEs* têm sido identificados em várias espécies de Mono e Eudicotiledôneas, como arroz (*Oryza sativa*), algodão (*Gossypium hirsutum*), soja (*Glycine max*) e *Arabidopsis thaliana* (Yu et al. 2000, Lisch et al. 2001).

Um estudo sobre a presença de *mudrA* em gramíneas indicou que a transposase do sistema *Mutator* é ubíqua no genoma das monocotiledôneas e, baseando-se nesta proteína, é

possível classificar os elementos *MuDR/hMuDR* em três clados. Neste estudo foram identificados cDNAs de trigo e arroz homólogos a *mudrA*, sugerindo que nessas espécies o sistema seja transcricionalmente ativo (Lisch *et al.* 2001). Devido ao fato de ter sido analisado um fragmento interno do gene *mudrA*, não se tem informação sobre as TIRs de cada um dos clados descritos.

No genoma de *A. thaliana*, os *MuLEs* foram classificados em 28 grupos, cada um com diferentes níveis de similaridade com *MuDR*. Posteriormente, foi descrito que dentre os 28 grupos, 9 deles tinham homologia com *mudrA*, sendo que 6 apresentavam TIRs e 3 não as possuíam (Yu *et al.* 2000).

Uma análise de 910 Kb de seqüência genômica de arroz realizada por Turcotte *et al.* (2001) revelou que 19,9% da seqüência analisada era constituída por TEs. Transposons superaram em número os retroelementos (166 contra 22), apesar destes perfazerem uma maior porcentagem dos nucleotídeos inspecionados. Dezessete elementos encontrados foram caracterizados como *MuLEs* e classificados em dez grupos. As TIRs mostraram alta similaridade (> 80%) dentro dos grupos, com as seqüências internas mais variáveis.

MuLEs não-autônomos não derivam apenas de deleções de elementos autônomos. Ao invés disso, uma variedade considerável de seqüências têm sido encontradas entre as TIRs, muitas não relacionadas à transposases. Em milho, arroz e *A. thaliana*, diversos desses *MuLEs* foram encontrados carregando fragmentos do genoma hospedeiro, sendo chamados *Pack-MuLEs* por Jiang *et al.* (2004) (Fig.3). No mesmo estudo, diferentes abordagens indicaram que esses fragmentos são ainda expressos e sofrem ação seletiva. Por outro lado, Juretic *et al.* (2005) identificaram 8274 *MuLEs* com TIRs e TSDs intactos no genoma do arroz, dos quais 1337, chamados *Transduplicated MuLEs*, contém fragmentos de genes hospedeiros (Fig. 3).

Estes fragmentos mostram características de pseudogenes, com domínios protéicos conservados incompletos, quebras de quadro de leitura e códon de parada prematuros, além do padrão de substituições sinônimas e não-sinônimas ser igual ao esperado para pseudogenes. Dessa forma, os resultados destes trabalhos são contraditórios no que diz respeito às forças evolucionárias agindo sobre os fragmentos capturados. Entretanto, ambos os estudos concordam sobre um importante papel dessas seqüências na organização e evolução do genoma hospedeiro, por serem uma potencial fonte para a criação de novos genes e/ou relacionarem-se à regulação de seus cognatos genômicos. Apesar do mecanismo de aquisição dos fragmentos não ser ainda compreendido, a presença de íntrons indica que ele envolve a captura de DNA genômico mais do que de cópias de cDNA dos transcritos celulares (Jiang *et al.* 2004).

Embora os transposons tenham sido vistos durante muito tempo como DNA egoísta, estudos recentes descrevem genes derivados de transposons com funções celulares determinadas (Hudson *et al.* 2003, Bundock e Hooykaas 2005). Tais genes mostram homologia com transposases, mas perderam as duas estruturas típicas de transposons: TIRs e TSDs, passando a ser chamados de "transposases domesticadas" (Fig. 3). Recentemente, uma nova família gênica de transposases domesticadas, relacionada a *Mutator*, foi encontrada em arroz e *A. thaliana*, sendo denominada *MUSTANG* (Cowan *et al.* 2005).

Figura 3. Os diferentes tipos de elementos que compõem o sistema *Mutator*.

1.3. A cana-de-açúcar e o Brasil

A cana-de-açúcar é uma planta alógama, pertencente ao gênero *Saccharum*, tribo *Andropogonae*, família *Poaceae*, dentro da classe das monocotiledôneas. À mesma família pertencem também o arroz (*Oryza sativa*), o trigo (*Triticum aestivum*), o milho (*Zea mays*) e o sorgo (*Sorghum bicolor*). É uma planta nativa das regiões tropicais, tendo iniciado sua dispersão pelo Velho Mundo ainda no período Neolítico, a partir da Nova Guiné. Nas Américas a espécie chegou com a segunda expedição de Cristóvão Colombo no final do século XV. No Brasil, Martim Afonso de Souza introduziu oficialmente mudas de cana-de-açúcar provenientes da Ilha da Madeira em 1502. Desde o início de sua domesticação a planta foi empregada na alimentação e, devido a seu elevado teor calórico, serviu como reserva energética em viagens marítimas e migrações (Cesnik e Miocque, 2004).

Atualmente a cana-de-açúcar é uma das culturas mais importantes nas regiões tropicais e subtropicais do planeta, com cultivo intensivo em mais de noventa países. É utilizada principalmente para obtenção de açúcar, servindo como fonte para cerca de 60% do açúcar bruto produzido no mundo. Em alguns países, como o Brasil, também é empregada na produção de álcool. Brasil e Índia são os principais produtores de cana-de-açúcar do mundo (Grivet e Arruda 2002).

O esgotamento das reservas de combustíveis fósseis mundiais, associado à crescente preocupação com o efeito estufa, causador do aquecimento global, têm levado a um grande aumento na demanda por fontes renováveis de energia e combustíveis menos poluentes. Dentre as alternativas, os biocombustíveis mostram maior viabilidade (Hill *et al.* 2006). Dentre todos os biocombustíveis disponíveis atualmente, aquele que apresenta melhor relação custo-benefício, dos pontos de vista econômico, tecnológico e ambiental, é o etanol produzido a partir da cana-de-açúcar. Processos de produção baseados em outras espécies, como milho ou beterraba, mostraram-se menos eficientes (Pimentel e Patzek 2005, Andreoli e Souza 2006). Neste cenário o Brasil tem um papel de destaque, uma vez que, ao lado dos Estados Unidos, cuja produção é baseada no milho, lidera a produção mundial de etanol produzindo sozinho cerca de 50% do álcool comercializado internacionalmente. Em média 55% da cana-de-açúcar cultivada no Brasil ao longo do ano é transformada em álcool e o restante, em açúcar. O bagaço, resíduo desta transformação, pode ainda ser utilizado na produção de biodiesel, biogás, plásticos biodegradáveis e até mesmo energia elétrica, derivada de sua queima (União da Agroindústria Canavieira de São Paulo – UNICA - www.unica.com.br).

1.4. O genoma da cana-de-açúcar

Embora a cana-de-açúcar seja uma cultura economicamente importante, o seu genoma é pouco estudado, principalmente devido a sua complexidade. Os cultivares modernos são altamente poliplóides, apresentando também aneuploidias. Derivadas de uma hibridação interespecífica entre a espécie domesticada *Saccharum officinarum* ($2n=80$) e a selvagem *Saccharum spontaneum* ($2n=40-128$), as atuais variedades comerciais de cana-de-açúcar contêm de 100 a 130 cromossomos, dos quais 15-25% provém de *S. spontaneum* e o restante de *S. officinarum* (Fig. 4) (D'Hont 2005). 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). Considerando genomas monoplóides, o conteúdo de DNA é de 930 Mpb para *S. officinarum*, 750 Mpb para *S. spontaneum* e aproximadamente 1000 Mpb para cana (D'Hont e Glaszman 2001), enquanto as espécies próximas como sorgo ou arroz apresentam 760 Mpb e 430 Mpb respectivamente. O híbrido é propagado vegetativamente através de um segmento do colmo contendo de 2 a 4 gemas, denominado tolete (Cesnik e Miocque, 2004).

O Projeto SUCEST (Sugarcane Expressed Sequence Tags) da Fundação de Amparo à Pesquisa no Estado de São Paulo (FAPESP) foi o responsável pelo seqüenciamento do transcriptoma da cana-de-açúcar. Foram parcialmente seqüenciados mais de 260.000 clones de cDNA, provenientes de 26 bibliotecas de cDNA, geradas a partir de RNA extraído de diferentes tecidos da planta. Com isso foram produzidas 237.954 ESTs (Expressed Sequence Tags) de alta qualidade. Estas ESTs foram montadas (sobrepostas) resultando em 43.141 transcritos consenso. A anotação dessas seqüências associou quase 50% destes transcritos

com metabolismo protéico, comunicação celular e transdução de sinal, bioenergética e resposta a estresse. TEs representaram 2,3% das seqüências anotadas, enquanto que 16,8% das seqüências não apresentaram homologia com nenhuma seqüência de DNA previamente caracterizada. Uma análise comparativa entre os 43.141 transcritos revelou uma redundância de 22%, indicando que foram identificados 33.620 genes em cana de açúcar. Uma vez que em arroz ou tomate existem aproximadamente 35.000 genes, número também estimado para cana, o número obtido equivale a mais de 90% dos genes da planta (Vettore *et al.* 2003).

A situação de choque genômico e a conseqüente instabilidade genômica, proporcionados pelo cruzamento entre espécies, podem levar à ativação de elementos de transposição quiescentes (Chen e Ni, 2006). A ativação tanto de transposons como de retrotransposons foi verificada em híbridos poliplóides de *A. thaliana* (Comai *et al.* 2000 e Madlung *et al.* 2005), assim como de *Triticum* (Kashkush *et al.* 2003). Em *A. thaliana* a ativação foi correlacionada com a redução do grau de metilação das seqüências dos TEs (Madlung *et al.* 2002). Neste sentido, no transcriptoma da cana foram identificadas 276 seqüências que apresentaram alta similaridade com TEs previamente descritos, representando 21 clados diferentes de TEs (Rossi *et al.* 2001). Uma análise detalhada da expressão destes TEs através de macroarranjos demonstrou alta atividade transcricional de diversos elementos em meristema apical, folhas, flores e calo. Também foi possível verificar, através de experimentos de expressão transiente e estável, a funcionalidade das regiões promotoras de vários retrotransposons da família *Hopscotch* (Araújo e Rossi *et al.* 2005). Estes trabalhos demonstraram que no genoma híbrido da cana-de-açúcar também existe uma alta atividade transcricional de TEs.

Figura 4. Contribuição dos parentais ao conteúdo cromossômico de cana-de-açúcar. (Adaptado de D'Hont 2005).

1.5. A cana de açúcar e *Mutator*

Entre as 276 seqüências identificadas no transcriptoma de cana que apresentaram alta similaridade com TEs previamente descritos, 148 apresentaram similaridade com transposons e 128 com retrotransposons, sendo *Mutator* o transposon mais representado (Rossi *et al.* 2001).

A análise de macroarranjos realizada por Araújo e Rossi *et al.* (2005), anteriormente mencionada, demonstrou que as seqüências homólogas a *mudrA* totalizam 38% dos TEs expressos em cana. O tecido que apresentou a maior quantidade de transcritos *mudrA* foi calo (Fig. 5).

Figura 5. Análise da expressão de genes *mudrA* em cana-de-açúcar.

As linhas representam os clones de cDNA de cana de açúcar (TEXXX) homólogos à transposase de *Mutator*, *mudrA*. As colunas representam os diferentes tecidos; C: calo, M: meristema apical, F: folhas, e FL: flores. Cinza e preto indicam presença e ausência do mensageiro, respectivamente (Adaptado de Araújo e Rossi *et al.* 2005).

O primeiro estudo específico sobre o sistema *Mutator* no transcriptoma de cana identificou diversos mensageiros homólogos a *mudrA*. Nenhuma seqüência apresentou semelhança com *mudrB*. Análises filogenéticas mostraram a presença de pelo menos quatro clados de seqüências *mudrA*, chamados Classes I, II, III e IV, nas espécies analisadas: cana, arroz e *A. thaliana*. Dentro de cada Classe a relação filogenética entre as espécies é respeitada (Fig. 6). Estes resultados sugerem que a existência destes clados é anterior à divergência entre mono e dicotiledôneas há mais de 150 milhões de anos (Rossi *et al.* 2004).

Figura 6. Análise filogenética das seqüências de cana-de-açúcar, *A. thaliana* e *O. sativa* homólogas a *mudrA*.

As seqüências de cana de açúcar, *A. thaliana* e *O. sativa* estão indicadas como TExXX, At.X e Os.X respectivamente. As seqüências de cana marcadas com uma caixa são os maiores clones identificados para cada classe. Os valores de bootstrap maiores que 50 % estão indicados nos ramos correspondentes da árvore (Adaptado de Rossi *et al.* 2004).

Dada a sua função, as transposases possuem um domínio de ligação ao DNA conservado, **CX2CX4HX4(ou 6)C** (onde X é qualquer aminoácido), na extremidade C terminal. Em particular, a transposase MURA reconhece um motivo de 32 pb contido nas TIRs. Nas seqüências de cana-de-açúcar foi possível identificar o domínio conservado de ligação ao DNA que, assim como em arroz e *A. thaliana*, apresenta alternativamente a versão com 4 ou 6 X entre os aminoácidos conservados H e C. Este domínio pôde ser agrupado de acordo com as quatro Classes mencionadas acima (Rossi *et al.* 2004).

Com o intuito de conhecer o número de cópias de cada uma das quatro Classes no genoma de cana-de-açúcar, Saccaro Jr *et al.* (2007) (Anexo I) realizaram uma busca em uma biblioteca comercial de BACs (Bacterial Artificial Chromosomes) de cana-de-açúcar do cultivar R570 (Tomkins *et al.* 1999). Os maiores clones de cDNA pertencentes a cada uma das Classes foram utilizados como sondas (indicados com caixas na figura 6). A biblioteca triada contém 103.269 clones (<https://www.genome.clemson.edu/>). Considerando-se que os clones têm em média 130 kb e que o genoma poliplóide da planta possui 10.000 Mb, a biblioteca representa aproximadamente 1,3 vezes o genoma da cana. Desta maneira, a partir do número total de clones identificados, foi possível calcular o número aproximado de cópias por Classe presentes no genoma de cana-de-açúcar. A Classe I apresentou 275 cópias, a Classe II 1720 cópias, a Classe III 5 cópias e a Classe IV 26 cópias por genoma. O resultado mostrou um panorama muito interessante, onde é possível observar variações de uma ordem de magnitude de diferença no número de cópias entre as Classes, sugerindo diferenças no nível de atividade transposicional.

1.6. Genômica comparativa

Genômica Comparativa é o estudo das semelhanças e diferenças, em estrutura e função, da informação hereditária entre os *taxa*, através de ferramentas moleculares e computacionais. Particularmente nas plantas, é possível verificar que a evolução de porções pequenas mas essenciais do genoma ocorre de forma relativamente lenta, possibilitando, entre espécies que divergiram há muito tempo, o reconhecimento tanto de regiões intragênicas comuns quanto de

arranjos similares de genes ao longo dos cromossomos. Os desvios da colinearidade e sintenia se devem a diversos fatores, como duplicações e segmentações cromossômicas, mobilidade de TEs, deleções de genes e rearranjos localizados (Paterson *et al.* 2000).

O advento do seqüenciamento genômico em grande escala possibilitou análises comparativas entre espécies de diversos graus de proximidade, contribuindo para um maior entendimento acerca da dinâmica evolutiva e funcional dos seres vivos. No caso dos transposons, devido a sua capacidade de mobilização nos genomas, a Genômica Comparativa tornou-se essencial para a compreensão do impacto destes elementos na evolução dos genomas.

O genoma de milho é aproximadamente três vezes maior que o de arroz. Em 1998, SanMiguel e Bennetzen mostraram, através da análise da região flanqueadora do gene *adh1*, que a diferença de tamanho entre os genomas destas duas espécies seria resultado de amplificações de TEs ocorridas após a divergência entre elas, há aproximadamente 16 milhões de anos.

Recentemente, Jannoo *et al.* (2007) compararam as seqüências correspondentes à região genômica do gene *adh1* dos parentais da cana-de-açúcar, *S. officinarum* e *S. spontaneum*, e sorgo. Os resultados indicaram perfeita colinearidade entre os haplótipos homólogos do híbrido bem como alta conservação da estrutura dos genes. Excetuando-se a inserção de alguns retrotransposons, uma alta identidade entre os haplótipos foi também observada em regiões não transcritas. Em relação ao sorgo, as seqüências de cana mostraram colinearidade, com exceção de dois genes presentes apenas em sorgo, havendo também notável identidade entre regiões não-codificantes. A maior parte seqüências não-alinhadas corresponderam a TEs, presentes em maior número nas seqüências provenientes de *Saccharum*. Foi também

verificada a ocorrência de colinearidade da região estudada com a região correspondente em arroz.

O arroz, por ser diplóide, economicamente importante e possuir um genoma pequeno, de aproximadamente 430 Mpb, é a planta modelo não só para o estudo das gramíneas como também para o das monocotiledôneas em geral. Seu genoma encontra-se totalmente seqüenciado (Internacional Rice Genome Sequencing Project, <http://rgp.dna.affrc.go.jp/>). Uma comparação entre o transcriptoma de cana e o genoma de arroz mostrou que 81,6% das seqüências de ESTs de cana possuem homologias no genoma de arroz. Os 18,4% restantes dividiram-se em dois grupos: seqüências que mostraram homologia com seqüências de outras angiospermas, mas não com arroz (4,9%), indicando uma possível perda gênica nesta espécie; e seqüências que não mostraram homologia com nenhuma outra espécie analisada (13.5%), podendo representar novidades evolutivas ou seqüências de evolução muito rápida (Vincentz *et al.* 2004).

A maior parte da diferença entre o tamanho dos genomas de arroz e de cana se deve à poliploidia da última, porém mesmo o genoma monoplóide de cana é mais de duas vezes maior que o do arroz. Uma das hipóteses para explicar a diferença seria o aumento do número de TEs no genoma da cana-de-açúcar. A proximidade filogenética com a cana-de-açúcar, aliada ao fato de possuir um genoma mais simples, torna o arroz ideal para um estudo comparativo do sistema *Mutator* nos dois genomas.

2. Conclusões

Os resultados apresentados nesta dissertação suportam o papel dos TEs como protagonistas da evolução do genoma. Foi possível avaliar a abundância e diversidade do sistema *Mu* em gramíneas, através de uma análise comparativa entre cana-de-açúcar e arroz, integrando a informação proveniente dos diversos trabalhos pré-existentes. Neste contexto, ficou evidente uma explosão do número de cópias dos elementos *mudrA* de Classes II ocorrendo ao longo da evolução das gramíneas, antes e depois da divergência entre *Oryza* e *Saccharum*. A partir da análise estrutural dessas seqüências foi possível propor que o sistema *Mutator* em gramíneas é composto por: A) seqüências *mudrA* pertencentes às Classes I e II, que conservam características de transposons e por tanto compreendem transposons propriamente ditos; e seus correspondentes *Pack-MuLEs* e *Transduplicated MuLEs*; e B) pelos elementos de Classes III e IV, que são transposases domesticadas, recentemente descrita como *MUSTANGs*. Como mostrado pelas análises filogenéticas realizadas, o evento de domesticação ocorreu relativamente cedo na evolução das Angiospermas, antes da divergência entre Eudicotiledôneas e Monocotiledôneas.

Foi obtida a seqüência completa para dois clones de BAC de cana de açúcar, um proveniente de cada parental do híbrido (*S. officinarum* e *S. spontaneum*), totalizando aproximadamente 247 mil pares de base de seqüência genômica. A existência de colinearidade entre estes BACs mostrou que ambos os fragmentos genômicos são haplótipos de *S. officinarum* e *S. spontaneum*. Os *MUSTANGs* encontrados em cada um deles revelaram-se ortólogos ao mesmo *locus* de arroz. A obtenção destas seqüências permitirá ainda que, no futuro, estudos funcionais sejam realizados esclarecendo a função destas transposases domesticadas.

Os genes completos, presentes em ambos BACs, possibilitaram identificar a região ortóloga de arroz. A colinearidade entre *Saccharum* e arroz está alterada na região do *MUSTANG* devido à ocorrência de um rearranjo.

A comparação das regiões intergênicas entre cana e arroz permitiu a conclusão de que amplificação de TEs é, sem dúvida, um dos motivos para a diferença de tamanho dos genomas entre as duas espécies. Os resultados sugerem ainda que a atividade dos retrotransposons seria responsável também pela presença de fragmentos gênicos distintos nas regiões intergênicas de cana.

Desde sua descoberta na década de 40, os TEs têm suscitado inúmeras questões, por vezes polêmicas. Porém, sejam parasitas, elementos essenciais ou um pouco de ambos, é indiscutível hoje sua importância para a evolução da vida na Terra. As informações apresentadas nesta dissertação possibilitam alguns esclarecimentos e, principalmente, novas questões sobre essas seqüências de DNA que continuam tão (ou mais) intrigantes hoje quanto o foram para Bárbara McClintock mais de sessenta anos atrás.

3. Bibliografia

- Almeida LM, Silva IT, Silva WA Jr, Castro JP, Riggs PK, Carareto CM, Amaral ME (2007) **The contribution of transposable elements to Bos taurus gene structure.** Gene 390(1-2): 180-189.
- Araújo PG, Rossi MM, de Jesus EM, Saccaro-Junior NL, Kajihara D, Massa R, Felix JM, Drummond RD, Falco MC, Chabregas SM, Ulian EC, Menossi M, Van Sluys MS (2005) **Transcriptionally active transposable elements in recent hybrid sugarcane.** Plant Journal 44(5): 707-717.
- Andreoli C, de Souza SP (2006) **Cana de açúcar: a melhor alternativa para a conversão da energia solar e fóssil em etanol.** Revista Economia e Energia 59(6): 27-33.
- Bundock P, Hooykaas P (2005) **An arabidopsis hAT-like transposase is essential for plant development.** Nature 436: 282-284.
- Cesnik R, Miocque J (2004) **Melhoramento da cana-de-açúcar.** Embrapa Informação Tecnológica, DF.
- Chen ZJ, Ni Z (2006) **Mechanisms of genomic rearrangements and gene expression changes in plant polyploids.** BioEssays, 28: 240-252.
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B (2000) **Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids.** Plant Cell 12(9):1551-1568.
- Cowan RK, Hoen DR, Shoen DJ, Bureau TE (2005) **MUSTANG is a novel family of domesticated transposase genes found in diverse Angiosperms.** Mol. Biol. Evol. 22(10): 2084-2089.
- Dawkins R (1986) **The blind watchmaker.** Harlow, Longman, Reino Unido.
- Devos KM, Brown JK, Bennetzen JL (2002) **Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis.** Genome Res. 12(7): 1075-1079
- D'Hont A (2005) **Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana.** Cytogenet. Genome Res. 109(1-3): 27-33
- D'Hont A, Glaszman JC (2001) **Sugarcane genome analysis with molecular markers: a first decade of research.** Proc. Int. Soc. Sugarcane Technol. 24: 556-559.
- Diao X, Freeling M, Lisch D (2006) **Horizontal transfer of a plant transposon.** PLoS Biol. 4(1): 5.
- Evgen'ev MB (2007) **Mobile elements and genome evolution.** Molecular Biology 41(2): 203-213.
- Feschotte C, Jiang N, Wessler SR (2002) **Plant transposable elements: where genetics meets genomics.** Nature Genetics 3: 329-341.
- Flavell AJ, Pearce SR, Heslop-Harrison P, Kumar A (1997) **The evolution of Ty1-copia group retrotransposons in eukaryote genomes.** Genetica 100: 185-195.
- Flavell AJ, Pearce SR, Kumar A (1994) **Plant transposable elements and the genome.** Current Opinion in Genetics and Development 4: 834-844.
- Fosket DE (1994) **The size and complexity of plants genomes.** Plant Growth and Development – a molecular approach. Cap. 3. 1ª. Ed. Academic Press. New York.

- Fulton TM, Chunwongse J, Tanksley SD (1995) **Microprep protocol for extraction of DNA from tomato and other herbaceous plants**. Plant Molecular Biology Reporter 13: 207-209.
- Gordon D, Abajian C, Green P (1998) **Consed: a graphical tool for sequence finishing**. Genome Research 8: 195-202.
- Grandbastien M (1998) **Activation of plant RTNs under stress conditions**. Trends in Plant Sci. 3: 181-189.
- Grandbastien MA (1992) **Retroelements in higher plants**. Trend Gen. 8: 103-108.
- Grandbastien MA, Audeon C, Bonnivarde E, Casacuberta JM, Chalhoub B, Costa AP, Lê QH, Melayah D, Petit M, Poncet C, Tam SM, Van Sluys MA, Mhiri C (2005) **Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae**. Cytogenet Genome Res. 110(1-4): 229-41.
- Grandbastien MA, Audeon C, Casacuberta JM, Grappin P, Lucas H, Moreau MC e Pouteau S. (1994) **Functional analysis of the tobacco Tnt1 retrotransposon**. Genetica 93:181-189.
- Grivet L, Arruda P (2002) **Sugarcane genomics: depicting the complex genome of a important tropical crop**. Current Opinion in Plant Biology 5(2): 122-127.
- Grivet L, D'Hont A, Roques D, Feldman P, Lanaud C, Glaszmann JC (1996) **RFLP mapping in cultivated sugarcane (Saccharum spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid**. Genetics 142: 987-1000.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) **Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. Nucleic Acids Res. 22: 4673-4680.
- Hill J, Nelson E, Tilman D, Polasky S, Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. **Proc Natl Acad Sci 103(30): 11206-10**.
- Hudson ME, Lisch DR, Quail PH (2003) **The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway**. The Plant Journal 34: 453-471.
- Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) **Pack-MULE transposable elements mediate gene evolution in plants**. Nature 431: 569-573.
- Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) **The evolutionary fate of MULE-mediated duplications of host gene fragments in rice**. Genome Res. 15:1292-1297.
- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) **Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence**. Proc. Natl. Acad. Sci. 97(12): 6603-6607.
- Kashkush K., Feldman M, Levy AA (2002) **Gene loss, silencing and activation in a newly synthesized wheat allotetraploid**. Genetics 160: 1651-1659.
- Kazazian Jr HH (2004) **Mobile elements: drivers of genome evolution**. Science 303(5664): 1626-1632.
- Kidwell MG, Lisch DR (2000) **Transposable elements and host genome evolution**. Trends Ecol. Evol. 15(3): 95-99.

- Leprinc AS, Grandbastien MA, Christian M (2001) **Retrotransposons of the Tnt1B family are mobile in *Nicotiana plumbaginifolia* and can induce alternative splicing of the host gene upon insertion.** Plant Mol. Biol. 47(4): 533-541.
- Lisch DR (2002) ***Mutator* transposons.** Trends in Plant Science 7:498-504.
- Lisch DR, Freeling M, Langham RJ, Choy MY (2001) ***Mutator* transposase is widespread in the grasses.** Plant Physiol. 125:1293-1303.
- Long M (2001) **Evolution of novel genes.** Curr. Opin. Genet. Dev. 11(6): 673-680
- Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) **Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids.** Plant Physiol, 129: 733-746.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) **Genomic changes in synthetic *Arabidopsis* polyploids.** The Plant Journal 41(2): 221-230.
- McClintock B (1956) **Controlling Elements and the gene.** Cold Spring Harbor Symp. Quant Biol. 21:197-216.
- McClintock B (1984) **The significance of responses of the genome to challenge.** Science 226:792-801.
- Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrels ME, Irvie JE, Paterson AH (1998) **Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes.** Genetics 150: 1663-1682.
- Miskey C, Papp B, Mátés L, Sinzelle L, Keller H, Izsvák Z, Ivics Z (2007) **The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends.** Mol. Cell Biol. 27(12):4589-4600.
- Monteiro-Vitorello CB, de Oliveira MC, Zerillo MM, Varani AM, Civerolo E, Van Sluys MA. (2005) ***Xylella* and *Xanthomonas* Mobil'omics.** OMICS 9(2): 146-59
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) **Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize.** Nat Genet. 37(9): 997-1002.
- Ogiwara I, Miya M, Ohshima K, e Okada N (1999) **Retropositional parasitism of SINEs on LINEs: Identification of SINEs and LINEs in elasmobranchs.** Mol. Biol. Evol. 16:1238-1250.
- Paterson AH, Bowers JE, Burow MD, Draye X, Elisk CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, Wright RJ (2000) **Comparative genomics of plant chromosomes.** Plant Cell 12(9):1523-1540.
- Pennisi E (2007) **Jump genes hop into the evolutionary limelight.** Science 317: 894-895.
- Pimentel D, Patek TW (2005) **Ethanol Production Using Corn, Switchgrass, and Wood; Biodiesel Production Using Soybean and Sunflower.** Natural Resources Res. 14(1): 65-76.
- Robertson DS (1978) **Characterization of a *Mutator* system in maize.** Mutat. Res. 51:21-28.

- Rossi M, Araujo PG, de Jesus EM, Varani AM e Van Sluys MA(2004) **Molecular Genetics and Genomics** 272:194-203.
- Rossi MM, Araujo PG e Van Sluys M-A (2001) **Survey of transposable elements in sugarcane expressed sequence tags ESTs**. *Genetics and Molecular Biology* 24:147-154.
- Rudenko GN, Ono A, Walbot V (2003) **Initiation of silencing of maize MuDR/Mu transposable elements**. *The Plant Journal* 33:1013-1025.
- Sabot F, Schulman AH (2006) **Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome**. *Heredity* 97(6): 381-388.
- Saccaro Jr NL, Van Sluys MA, Varani AM, Rossi M. (2007) **MudrA-like sequences from rice and sugarcane cluster as two bona fide transposon clades and two domesticated transposases**. *Gene* 392(1-2):117-125.
- Saitou N, Nei M (1987) **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol. Biol. Evol.* 4(4):406-25
- Sambrook J, Fritsch EF, Maniatis T (1989) **Molecular Cloning, a laboratory manual**. Second Edition, Cold Spring Harbor Laboratory Press.
- San Miguel P e Bennetzen JL (1998) **Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons**. *Ann. Bot.* 81:37-44.
- Schmidt T (1999) **LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes**. *Plant Mol. Biol.* 40(6): 903-10.
- Slotkin RK, Martienssen R (2007) **Transposable elements and the epigenetic regulation of the genome**. *Nat. Rev. Genet.* 8(4): 272-285.
- Smyth DR (1993) **Plant retrotransposons**. In VERMA, D. P. S. ed. *Control of plant gene expression*. Boca Raton, USA, CRC Press, p.1-15.
- Siguier P, Filée J, Chandler M (2006) **Insertion sequences in prokaryotic genomes**. *Curr. Opin. Microbiol.* 9(5):526-531.
- Sironi M, Menozzi G, Comi GP, Bresolin N, Cagliani R, Pozzoli U (2005) **Fixation of conserved sequences shapes human intron size and influences transposon-insertion dynamics**. *Trends Genet.* 21(9): 484-488
- Swofford DL (1993) **PAUP: phylogenetic analysis using parsimony**. Illinois Natural History Survey, Urbana-Champaign
- Tatusova TA Madden TL (1999) **BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences**. *FEMS Microbiol. Lett.* 174(2): 247-50.
- Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo S, Wing RA (1999) **A bacterial artificial chromosome library for sugarcane**. *Theoretical and Applied Genetics* 99:419-424.
- Turcotte K, Srinivasan S e Bureau T (2001) **Survey of transposable elements from rice genomic sequences**. *The Plant Journal* 25(2), 169-179.
- Van Sluys MA, Tempé J, Fedoroff N (1987) **Studies on the introduction and mobility of the maize Activator element in *Arabidopsis thaliana* and *Daucus carota***. *EMBO J.* 6(13): 3881-3889.
- Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci Junior M, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA,

- Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon ML, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Junior VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P (2003) **Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane** (2003) *Genome Res.* 13(12): 2725-35.
- Vicient CM, Suoniemi A, Ananthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) **Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus Hordeum.** *Plant Cell* 11(9): 1769-1784.
- Vincentz M, Cara FAA, Okura VK, da Silva FR, Pedrosa GL, Hemerly AS, Capella NA, Marins M, Ferreira PC, França SC, Grivet L, Vettore AL, Kemper EL, Burnquist WL, Targon MLP, Siqueira WJ, Kuramae EE, Marino CL, Camargo LEA, Carrer H, Coutinho LL, Furlan LR, Lemos MVF, Nunes LR, Gomes SL, Santelli RV, Goldman MH, Bacci Jr M, Giglioti EA, Thiemann OH, Silva FH, Van Sluys MA, Nobrega FG, Arruda P e Menck CFM (2004) **Evaluation of monocot and eudicot divergence using the sugarcane transcriptome.** *Plant Physiology* 134: 951-959.
- Vitte C, Bennetzen JL (2006) **Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution.** *Proc. Natl. Acad. Sci.* 103(47): 17638-17643.
- Walbot V (1992) **Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 43: 49-82.
- Wessler RS (1996) **Plant retrotransposons: turned on by stress.** *Current Biology.* 6: 959-961.
- Yu Z, Wright SI e Bureau TE (2000) **Mutator-like elements in Arabidopsis thaliana: structure, diversity and evolution.** *Genetics* 156: 2019-2031.
- Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, Lu Y, Denoeud F, Antonarakis SE, Snyder M, Ruan Y, Wei CL, Gingeras TR, Guigó R, Harrow J, Gerstein MB (2007) **Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution.** *Genome Res.* 17(6):839-851.