UNIVERSIDADE ESTADUAL DE CAMPINAS

Karine Miranda Oliveira

"DESENVOLVIMENTO DE MARCADORES MOLECULARES EST-SSRs E MAPEAMENTO FUNCIONAL EM CANA-DE-AÇÚCAR (*Saccharum* spp.)"

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) Karine Miranda Oliveira e aprovada pela Comissão Julgadora.

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

Orientadora: Profa. Dra. Anete Pereira de Souza

Co-orientador: Prof. Dr. Antônio Augusto Franco Garcia

Campinas
2006

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

Ol42d

Oliveira, Karine Miranda
Desenvolvimento de marcadores moleculares EST-SSRs e mapeamento funcional em cana-de-açúcar (*Saccharum* spp.) / Karine Miranda Oliveira. -- Campinas, SP: [s.n.], 2006.

Orientador: Anete Pereira de Souza.
Co-orientador: Antônio Augusto Franco Garcia.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Cana-de-açúcar. 2. Poliplóide. 3. SUCEST. 4. Marcadores EST-SSR. 5. Mapeamento funcional. I. Souza, Anete Pereira de. II. Garcia, Antônio Augusto Franco. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

(scs/ib)

Título em inglês: Development of EST-SSR molecular markers and functional mapping in sugarcane.
Palavras-chave em inglês: Sugarcane; Poliploidy; SUCEST; EST-SSR markers; Functional mapping.
Área de concentração: Genética Vegetal e Melhoramento.
Titulação: Doutora em Genética e Biologia Molecular.
Banca examinadora: Anete Pereira de Souza, Cláudio Lopes de Souza Júnior, Sabrina Moutinho Chabregas, Maria Imaculada Zucchi, Marcelo Menossi Teixeira.
Data da defesa: 30/10/2006.
Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 30 de Outubro de 2006

BANCA EXAMINADORA

Profa. Dra. Anete Pereira de Souza (Orientadora) _____ Assinatura

Prof. Dra. Sabrina Moutinho Chabregas _____ Assinatura

Prof. Dr. Cláudio Lopes de Souza Júnior _____ Assinatura

Profa. Dra. Maria Imaculada Zucchi _____ Assinatura

Prof. Dr. Marcelo Menossi Teixeira _____ Assinatura

Profa. Dra. Gláucia Mendes de Souza _____ Assinatura

Prof. Dr. Louis Bernard Klaczko _____ Assinatura

Prof. Dr. Michel Georges Albert Vincentz _____ Assinatura

Dedico

Aos meus pais **José Raimundo** e **Luciara**,

Aos meus irmãos **Gustavo** e **Emily**,

Ao meu namorado **Dario**,

pelo apoio incondicional.

Ofereço

À minha orientadora **Anete**,

Ao meu co-orientador **Augusto**,

pelos valiosos ensinamentos.

Agradeço

Aos meus pais José Raimundo e Luciara, por sempre acreditarem em mim e nos meus sonhos, por se orgulharem das minhas conquistas, pela preciosa educação, pela minha vida.

Aos meus irmãos Gustavo e Emily, por me ajudarem a crescer, por fazerem parte da minha vida, pela alegre e agradável convivência.

Ao meu namorado Dario, por ser um excelente companheiro e meu melhor amigo, por compreender os momentos de ausência durante o desenvolvimento deste trabalho, pelo amor, por me fazer muito feliz e me completar.

À Leandra, pela amizade e carinho, pelo interesse e apoio ao meu trabalho.

À minha orientadora Dra. Anete Pereira de Souza, pela grande orientação, pelos conhecimentos passados com muita paciência e dedicação, pelo apoio e incentivo, por me ensinar a fazer pesquisa.

Ao meu co-orientador Dr. Antônio Augusto Franco Garcia, por toda gentileza e disponibilidade no esclarecimento das minhas dúvidas, por me apresentar de forma clara a estatística como aliada.

Aos colegas do laboratório de Análises Genética e Molecular / CBMEG e Barracão, pelos conselhos e ajuda nos momentos de dúvidas, pelos momentos de descontração, pelo carinho durante estes anos de convivência.

Aos companheiros do Pite, Luciana Rossini, Laura e Thiago, pela valiosa e fundamental ajuda no desenvolvimento do projeto, pela união nas horas de dificuldade, pelo espírito de grupo, pela compreensão.

Aos colegas Marta e Gabriel (Esalq / USP), pelo imenso auxílio nas análises e discussão dos dados.

À professora Dra Eliana Regina Forni-Martins e Juan Domingos Urdampilleta (Dep. Botânica Unicamp), pelos ensinamentos e grande disponibilidade durante as análises citogenéticas.

Aos funcionários, Juverlande e Zildinha, aos técnicos, Georgina, Paulo e demais técnicos que passaram pelo laboratório, por serem sempre prestativos e gentis, pela participação essencial no bom andamento dos experimentos.

Aos professores participantes da pré-banca, Dra. Silvana Aparecida Creste Dias de Souza e Dr. Michel Georges Albert Vincentz, pela disponibilidade em participar da avaliação deste trabalho.

Aos professores participantes da banca, Dra. Sabrina Moutinho Chabregas, Dr. Cláudio Lopes de Souza Júnior, Dra. Maria Imaculada Zucchi e Dr. Marcelo Menossi Teixeira, pela análise da tese, comentários construtivos e tempo despendido.

À Fundação de Amparo à Pesquisa do Estado de São Paulo pela bolsa DD concedida (02/00197-4) e auxílio financeiro (FAPESP, 02/01167-1) e ao Centro de Tecnologia Canavieira, pelo financiamento de parte da pesquisa desenvolvida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa PDEE (0181-05-6) concedida, por possibilitar o desenvolvimento dos experimentos em laboratório no exterior.

À todos que contribuíram no andamento deste trabalho e tornaram possível esta conquista.

Sumário

Lista de tabelas

2 Revisão Bibliográfica

4 Artigo I

5 Artigo II

6 Artigo III

Lista de figuras

2 Revisão Bibliográfica

5 Artigo II

6 Artigo III

Resultados complementares

Prefácio

Os resultados obtidos durante o desenvolvimento deste trabalho de tese estão apresentados na forma de artigos científicos. O primeiro artigo, *Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats*, publicado na revista *Genome* (47: 795-804; 2004), descreve a busca de marcadores moleculares do tipo microssatélites no banco de dados do SUCEST, além do desenvolvimento e avaliação dos primeiros EST-SSRs. O segundo artigo, *Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs*, publicado na revista *Plant Breeding* (125: 378-384; 2006), apresenta a caracterização de um conjunto dos EST-SSRs desenvolvidos, a verificação do poder discriminatório destes marcadores e a comparação destes com SSRs genômicos. O terceiro artigo, *Functional integrated genetic linkage map based on EST-markers for a sugarcane (Saccharum spp.) commercial cross*, submetido à revista *Molecular Breeding*, apresenta o mapa funcional construído para a população de mapeamento 'SP80-180 x SP80-4966', a partir da integração dos EST-SSRs ao mapa prévio desta população.

Como resultado complementar, foram realizados estudos citogenéticos nos parentais da referida população de mapeamento e em alguns indivíduos da progênie F1, buscando um melhor entendimento dos resultados obtidos, isto é, objetivando relacionar os números e tipos de marcadores segregantes na população com os números de cromossomos dos indivíduos.

Durante o doutorado realizei estágio no CIRAD (*Centre de Coopération Internationale em Recherche Agronomique pour le développement – Montpellier/France*),

sob a orientação da Dra. Angélique D'Hont. Durante o estágio, trabalhei em um projeto de mapeamento genético de uma população de cana-de-açúcar proveniente de um cruzamento entre o cultivar 'R570' e o clone 'MQ76-53'. Integrei ao mapa genético desta população 13 dos locos EST-SSRs desenvolvidos no meu projeto de tese, além de 33 SSRs do programa do CIRAD. O resultado do mapeamento genético desta população encontra-se apresentado, em anexo, na forma de artigo científico, *Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene*, publicado na revista *Theoretical and Applied Genetics* (112: 1382-1391; 2006). Além deste projeto de mapeamento genético, também participei de outro projeto que tinha por objetivo o estudo da diversidade genética do fungo *Ustilago scitaminea*, que causa o carvão nas plantas de cana-de-açúcar. A partir dos esporos do fungo, realizei as análises com *primers* de SSRs marcados com fluorescência e posterior análise da diversidade. Este trabalho gerou o artigo, *Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea*, aceito para publicação na revista, *Fungal Genetics and Biology*. Durante este período também realizei a genotipagem de 70 dos EST-SSRs selecionados para mapeamento na população 'SP80-180 x SP80-4966', que encontram-se descritos no artigo 3 desta dissertação.

Resumo

A cana-de-açúcar (*Saccharum* spp.) está entre as espécies de maior importância econômica no mundo, constituindo uma das principais fontes de produção de açúcar e álcool. Apesar do Brasil ocupar posição de destaque, como o maior produtor mundial, os níveis de produtividade são considerados baixos. Os programas de melhoramento para obtenção de novas variedades de cana-de-açúcar, mais produtivas e resistentes a pragas e doenças, podem ser acelerados com o desenvolvimento de marcadores genéticos, PCR específicos, fortemente ligados a genes que controlam as características de interesse.

A utilização de marcadores em estudos de mapeamento genético e de QTL's (*Quantitative Trait Loci*) tem proporcionado um importante progresso no conhecimento da estrutura genômica e na genética da cana-de-açúcar. Sabe-se que os ESTs (*Expressed Sequence Tags*) apresentam grande potencial para serem utilizados com tais finalidades. O projeto de seqüenciamento de ESTs (SUCEST), do programa Genoma da FAPESP, identificou cerca de 43 mil *clusters* que representam os genes de cana-de-açúcar, potenciais para serem utilizados no desenvolvimento de marcadores genéticos.

Neste contexto, o presente trabalho teve como objetivo desenvolver e mapear marcadores EST-SSRs em uma progênie derivada do cruzamento entre duas variedades pré-comerciais de cana-de-açúcar, complementando os programas de mapeamento genético desta população. Uma busca no banco de dados do SUCEST resultou na identificação de marcadores microssatélites ou SSRs (*Single sequence repeats*) em 2005 *clusters*. Trezentos e setenta e dois locos EST-SSRs foram desenvolvidos e analisados e, destes, 149 foram selecionados para estudo de mapeamento genético.

Um total de 2303 marcadores polimórficos (SSRs, EST-SSRs, RFLPs, EST-RFLPs e AFLPs) foi identificado nos 100 indivíduos da progênie F1, dos quais 1669 (72%) eram marcadores em dose simples (1:1 e 3:1), segregantes na população. As análises de mapeamento foram baseadas na metodologia de identificação de marcadores em dosagem única no genoma, com o auxílio dos programas JoinMap (versão 3.0) e OneMap. Para a formação dos grupos de co-segregação (GCs) utilizou-se LOD 5 e fração de recombinação de 0.35. A função de mapeamento de Kosambi foi adotada para conversão das frequências de recombinação em distâncias de mapa em centiMorgans (cM). Dos 1669 marcadores segregantes, 664 (40%) foram distribuídos em 192GCs, gerando um mapa com 6261.1 cM de comprimento. Os 192 GCs foram agrupados em 14 prováveis grupos de homologia. Cento e treze dos 149 EST-SSRs avaliados foram mapeados e apresentaram homologia a genes conhecidos de outras espécies.

A adição dos marcadores provenientes de seqüências expressas, aumentou a cobertura e densidade do mapa prévio desta população, possibilitando também a construção do primeiro mapa funcional de cana-de-açúcar. Os EST-SSRs desenvolvidos têm potencial de utilização na detecção de QTL's associados a características de importância econômica para a cultura da cana-de-açúcar.

Abstract

Sugarcane (*Saccharum* spp.) is one of the most important cash crops, highly contributing towards the production of raw sugar and bioethanol produced worldwide. Even though Brazil is the major producer, sugar yield is considered low. Breeding programs for the attainment of new improved sugarcane varieties, which are more productive and more resistant to plagues and illnesses, could be sped up with the development of genetic markers that are PCR-specific and strongly linked to genes that control the desired agronomic traits.

The use of genetic markers in genetic mapping and QTL (*Quantitative Trait Loci*) studies has allowed important progress in the knowledge of the genomic structure and genetics of sugarcane. It is a known fact that the ESTs (*Expressed Sequence Tags*) have great potential to be used with such purposes. The Sugarcane Expressed Sequence Tag Project (SUCEST) has identified about 43000 clusters that represent the sugarcane genes with a potential to be used in the development of genetic markers.

In this context, the objective of the present work was to develop and map EST-SSR markers in a progeny obtained by the cross between two pre-commercial sugarcane varieties, complementing the genetic mapping programs of this population. A search in the SUCEST database resulted on the identification of the microssatellites or SSR (*Single sequence repeats*) markers in 2005 *clusters.* Thus, three hundred and seventy two EST-SSR loci had been developed and analyzed and, out of these, 149 were selected for mapping analyses.

A total of 2303 polymorphic markers (SSRs, EST-SSRs, RFLPs, EST-RFLPs and AFLPs) were identified among the 100 F1 individuals, of which 1669 (72.5%) were single-

dose (SD – segregated 1:1 and 3:1) markers. Map analyses were carried out using JoinMap (version 3.0) and OneMap algorithm, based on a single-dose marker approach. The linkage relationships of simplex markers were determined at a LOD score threshold of 5 and a recombination fraction threshold of 0.35 and map distances were derived from the recombination fraction using the Kosambi function. Out of these 1669 SD markers, 664 (40%) were scattered onto 192 co-segregation groups (CGs) with a total estimated map length of 6261.1 cM. Using both genomic and EST-derived SSR and RFLP, 120 of the 192 CGs were formed into fourteen putative homology groups (HGs). One hundred and thirteen of the 149 EST-SSRs evaluated were mapped and presented homology to previously studied genes of other species.

The addition of the EST-derived markers increased the coverage and density of the previous map of this population, which also enabled the construction of the first functional linkage sugarcane map. The EST-SSRs developed have the potential to be used in the detection of QTLs associated to important economic traits for sugarcane culture.

1 Introdução

A cana-de-açúcar (*Saccharum* spp) está entre as mais importantes espécies cultivadas. É a principal fonte de sacarose, além de possibilitar a extração e utilização de inúmeros subprodutos a partir do bagaço e caldo obtidos no seu processamento.

O Brasil encontra-se na posição de maior produtor mundial de cana-de-açúcar, participando com 25% da produção, seguido pela Índia, Cuba, México e China, e exportando cerca de 39% de sua produção (Matsuoka 1999). Apesar da posição de destaque em nossa economia, este produto enfrenta oscilações de preço no mercado internacional, influenciado diretamente pela oferta de açúcar de beterraba, que representa 35% da produção mundial de açúcar (Hoarau 2000).

Os países desenvolvidos, produtores de açúcar de beterraba, já recorrem às mais modernas tecnologias em busca do aumento de sua produção e na tentativa de domínio do mercado. Os países da Comunidade Econômica Européia subsidiam fortemente a produção do açúcar de beterraba para atenuar os custos com as importações do açúcar de cana produzido nos trópicos. Dessa forma, a competição internacional no setor sucro-alcooleiro torna-se crescente a cada ano, e o Brasil necessita de maiores investimentos em tecnologia que resultem no aumento da produtividade na indústria açucareira, visando manter sua capacidade produtiva e competitiva.

A utilização de variedades melhoradas constitui a principal estratégia para alcançar o incremento da produtividade. O melhoramento genético destaca-se por sua importância na obtenção de cultivares com requisitos adequados a interesses agronômicos e industriais. Entretanto, o alto nível de ploidia e a complexidade citogenética das espécies de *Saccharum*, envolvendo classes variadas de cromossomos e eventos de recombinação,

impõem dificuldades no melhoramento desta cultura (Heinz 1987; Arumuganathan *et al*. 1991). Igualmente, as variedades comerciais de cana-de-açúcar são aneuplóides com alto número cromossômico, desenvolvidas a partir de uma série de cruzamentos entre as espécies *S. officinarum*, a qual contribui com o alto teor de açúcar e, *S. spontaneum* responsável pelo vigor vegetativo e resistência aos estresses bióticos e abióticos, e retrocruzamentos para *S. officinarum.*

Além disso, um programa convencional de melhoramento genético em cana-de-açúcar leva um tempo relativamente longo para obtenção e distribuição de uma nova variedade aos produtores, em geral consome de 12 a 15 anos (Burnquist 2000), e pode ser efetivamente abreviado com o uso de marcadores moleculares.

Marcadores moleculares representam uma valiosa ferramenta, pois têm grande potencial na melhoria da eficiência dos programas de melhoramento, não apenas objetivando as características a serem selecionadas em uma geração, mas também precisando qual genótipo pode ser selecionado. Informações obtidas a partir destes marcadores têm contribuído no melhor entendimento da evolução e genética tanto de espécies diplóides quanto poliplóides (Soltis e Soltis 1993).

O número elevado de plantas analisadas nos programas de melhoramento requer um diagnóstico rápido e preciso, o qual é alcançado pelo emprego de marcadores revelados pela reação em cadeia da polimerase (PCR). Os marcadores microssatélites ou SSRs (*Simple Sequence Repeats*) apresentam grande valor no estudo de muitas espécies de plantas. Este sucesso é atribuído a sua natureza multialélica, herança codominante, facilidade de detecção pela PCR, abundância relativa, cobertura extensiva do genoma e necessidade de quantidades pequenas de DNA para dar início às reações (Powell *et al.* 1996).

A disponibilidade destes marcadores ao longo de todo o genoma permite a confecção de mapas genéticos, os quais possibilitam a aquisição de informações básicas sobre a estrutura e organização do genoma de uma espécie estudada, tais como: padrões de distorção de segregação Mendeliana de segmentos cromossômicos, associações de marcadores com caracteres qualitativos e localização dos mesmos e identificação de regiões genômicas associadas a caracteres quantitativos (Ferreira e Grattapaglia 1998).

O sequenciamento genômico tem contribuído significativamente na área de mapeamento de poligenes, principalmente na identificação de "marcadores candidatos" que apresentam uma maior probabilidade de estarem ligados a QTL's (*Quantitative Trait Loci*). O projeto SUCEST (*Sugarcane EST*) de seqüenciamento de ESTs (*Expressed Sequence Tags*) em cana-de-açúcar representa uma ótima fonte de "marcadores candidatos" (Camargo 2000a). Neste contexto, no sentido de explorar e aplicar as informações geradas pelo SUCEST e buscando identificar regiões genômicas associadas a genes que controlam características de importância agronômica e industrial, o presente trabalho teve por objetivo central desenvolver marcadores moleculares do tipo microssatélite a partir de ESTs.

Tais marcadores (EST-SSRs) foram integrados a um programa de mapeamento genético e de características qualitativas e quantitativas que vem sendo desenvolvido, empregando uma população F_1 obtida a partir do cruzamento de duas variedades 'SP 80-180' e 'SP 80-4966', híbridos pré-comerciais interespecíficos, altamente poliplóides e contrastantes quanto ao teor de sacarose. Este trabalho originou o primeiro mapa funcional de uma população de cana-de-açúcar.

2 Revisão Bibliográfica

2.1 Aspectos gerais da cultura da cana-de-açúcar

2.1.1 Domesticação e dispersão das primeiras formas cultivadas

A domesticação da cana-de-açúcar deu-se aproximadamente em 2500 a.C., na Nova Guiné, onde existia uma grande diversidade morfológica da espécie *Saccharum officinarum*. Inicialmente, a cana-de-açúcar era cultivada em jardins apenas para serem mascadas. Em seguida, *S. officinarum* difundiu-se para as ilhas do Sul do Pacífico, Índia e China, através de expedições australianas que ocorreram por volta de 1500 a 1000 a.C. (Brandes 1956; Daniels e Roach 1987). Duas outras espécies de cana-de-açúcar, *S. barberi* e *S. sinense*, apareceram nesta época, respectivamente na Índia e China. Foi exatamente na Índia e na China que a indústria de extração do açúcar se originou. Em 500 d.C., a cana-de-açúcar foi cultivada na Pérsia e de lá para a África do Norte e ilhas do mediterrâneo. Em seguida, no século XV, os Portugueses e Espanhóis propagaram a cultura nas ilhas do Atlântico. Na ocasião da sua segunda viagem, Cristóvão Colombo trouxe a cana-de-açúcar para as Américas e foi no Haiti que a cana-de-açúcar foi cultivada primeiramente. Durante os séculos XVI e XVII a extensão da cultura da cana-de-açúcar na América, principalmente no Brasil e no Caribe, estava estreitamente ligada às colonizações européias.

Até meados do século XVIII, o desenvolvimento de plantações de cana-de-açúcar realizou-se a partir de um único clone, ou de um número pequeno de clones, denominado Creoula. Tratava-se de um clone de *S. barberi* ou de um híbrido desta espécie com *S. officinarum*. Porém era pouco rústico e suscetível a doenças, tendo sido seu cultivo,

limitado a terras com alta fertilidade. Até meados do século XIX, este clone foi substituído pelo clone Bourbon. Devido a sua susceptibilidade, este clone, por sua vez, foi substituído pelo clone Cheribon e ainda Tanna (Stevenson 1965).

Oficialmente, foi Martim Afonso de Souza que, em 1532, trouxe a primeira muda de cana-de-açúcar ao Brasil e iniciou seu cultivo na Capitania de São Vicente. A partir das Capitanias de Pernambuco e da Bahia os engenhos de açúcar se multiplicaram, dando início a uma indústria que encontrou no Brasil seu campo fértil para uma rápida expansão e perpetuação. Após um início repleto de dificuldades, a produção de açúcar prosperou e passados menos de 50 anos o Brasil já detinha o monopólio mundial da produção (Machado 2003).

2.1.2 Início do melhoramento genético da cana-de-açúcar

Até o final do século XIX, a cultura da cana-de-açúcar deu-se através da propagação vegetativa de um número reduzido de clones provenientes das espécies *S. officinarum*, *S. sinense*, *S. barberi*. No final deste século, a fertilidade da cana-de-açúcar foi descoberta. A cana-de-açúcar era considerada como uma planta totalmente estéril até a descoberta de plântulas originadas de sementes naturais em 1858, em Barbados (Stevenson 1965). Porém esta aptidão só foi explorada para a produção de novas variedades a partir de 1888, em Java. Durante os vinte anos seguintes, diversos cruzamentos foram realizados pelos pioneiros de seleção em Java, na Índia, nas Ilhas Maurício e no Havaí (Bremer 1961a). Em um primeiro momento, os programas de melhoramento fundaram-se nos cruzamentos entre as canas 'nobres' (*S. officinarum*). Estes cruzamentos intra-específicos tiveram grande

sucesso. Os clones POJ100, 247B, EK28 obtidos em Java ou ainda BH10/12 ou B726 em Barbados foram sucessivamente cultivados nestas duas ilhas durante as duas primeiras décadas do século (Stevenson 1965). Estes trabalhos visaram controlar as principais doenças da época, como o mosaicismo.

O final do século XIX e início do século XX caracterizaram-se pelo aumento crescente das preocupações no setor açucareiro, no exterior e no Brasil, tanto com a baixa produtividade dos canaviais quanto com a melhoria da qualidade das canas (Stevenson 1965). Em consequência, passou a ocorrer intenso intercâmbio de tipos de cana entre países. A consciência da necessidade de pesquisa sistematizada acabou por levar a criação de estações experimentais, que tiveram como principais atividades a experimentação com variedades introduzidas e a criação de outras variedades locais (Matsuoka *et al.* 1999).

2.1.3 A importância econômica da cana-de-açúcar

A cana-de-açúcar, juntamente com a beterraba, é a base da indústria açucareira mundial. Atualmente, a produção mundial de sacarose ultrapassa 120 milhões de toneladas por ano, dos quais cerca de 70% são provenientes da cana-de-açúcar e 30% da beterraba. Este número mostra a importância atingida pelo açúcar da cana-de-açúcar na alimentação humana. Ela é cultivada em mais de 115 países, essencialmente em países tropicais devido à sua baixa tolerância ao frio (Feldmann *et al.* 1997).

O consumo mundial de açúcar continua a crescer por volta de 2.5% ao ano, atingindo aproximadamente 143 milhões de toneladas em 2004. Este aumento se iniciou pelo crescimento do mercado asiático, uma vez que o consumo nos países ocidentais

permaneceu estável ou apresentou uma leve queda. A proporção da produção mundial de açúcar proveniente da cana-de-açúcar se encontra em constante alta.

A maioria das regiões do mundo que cultivam a cana-de-açúcar tem programas de criação e seleção de variedades adaptadas às suas necessidades. Vale lembrar que o modo de propagação da cana-de-açúcar e sua cultura em numerosos países que não protegem as obtenções varietais tornam difícil a comercialização de cultivares. Existem duas coleções mundiais, uma na Índia e outra nos Estados Unidos, que reagrupam os recursos genéticos originados de numerosas prospecções efetuadas no centro de diversidade da cana-de-açúcar, na Ásia e na Oceania, desde o começo do século. Estas coleções são disponíveis a todos os países produtores de cana-de-açúcar agrupados na ISSCT (*International Society of Sugar Cane Technologists*). Os riscos fitopatológicos ligados às trocas destas plantas multiplicadas vegetativamente conduziram ao surgimento de instalações de quarentena (Feldmann *et al*. 1997).

No Brasil, o agronegócio sucroalcooleiro movimenta cerca de R$ 40 bilhões por ano, com faturamentos diretos e indiretos, o que corresponde a aproximadamente 2,35% do PIB nacional, além de ser um dos setores que mais empregam no país, com a geração de 3,6 milhões de empregos diretos e indiretos, e de congregar mais de 72.000 agricultores. Somente este ano o Brasil deve obter mais de US$ 3,5 bilhões em divisas com as exportações de 14,3 milhões de toneladas de açúcar e 2,5 bilhões de litros de álcool. Este setor faz do Brasil o maior produtor mundial de cana e açúcar e o principal país do mundo a implantar, em larga escala, um combustível renovável alternativo ao petróleo (www.jornalcana.com.br).

O setor sucroalcooleiro viveu em 2005 um momento de euforia com o sucesso do biocombustível. Vários países já estão estudando a utilização do álcool dentro de suas

matrizes energéticas. Nos Estados Unidos, por exemplo, vários Estados adicionarão 15% de etanol à gasolina até o ano de 2010, gerando uma demanda de cerca de 7,6 bilhões de litros. A China também irá adicionar 10% de álcool à gasolina, o que implicará na demanda de dois bilhões de litros de álcool ao ano. Outros países, como a Índia, vão adicionar 5% do produto à gasolina, estimando a necessidade de 400 milhões de litros de álcool. Da mesma forma, a Colômbia precisará de 750 milhões de litros, a Austrália de 350 milhões de litros e o México de três bilhões de litros ao ano. A estimativa é de que, nos próximos cinco anos, a demanda mundial por álcool ultrapassará os 10 bilhões de litros, o que levará à grande aceitação deste produto no mercado internacional, motivada principalmente por considerações de ordem ambiental, pela elevação dos preços do petróleo no mercado internacional e pela incerteza na oferta de combustíveis fósseis no médio e longo prazo.

O Brasil, além de maior produtor e consumidor de etanol, é também o maior exportador no cenário global. Até meados de 2002 as exportações brasileiras de álcool eram insignificantes, mas com o crescimento da demanda por esse biocombustível no mercado internacional, o volume exportado cresceu de 565 milhões de litros em 2003, para 2,1 bilhões de litros no período de Janeiro a Novembro de 2005 (Secex 2005).

Aliado ao crescimento das exportações brasileiras de açúcar, o cenário acima explica boa parte da significativa expansão do setor sucroalcooleiro nacional nos últimos anos e as perspectivas promissoras do mercado interno e externo para esse biocombustível num futuro bastante próximo. Sem dúvida, a necessidade de fornecer etanol para o mercado interno em expansão e para o mercado internacional, que anseia por fontes renováveis de energia, traz excelentes oportunidades para incrementos ainda maiores no crescimento do setor. Nos anos recentes, nota-se o aumento da produção de cana-de-açúcar e de seus

produtos derivados, açúcar e etanol, tanto nas tradicionais regiões produtoras como em estados que representam novas fronteiras agrícolas para a cultura canavieira no Brasil.

A Companhia Nacional de Abastecimento (CONAB) efetuou, recentemente, o primeiro levantamento da safra brasileira 2006/2007 de cana-de-açúcar e sua destinação (açúcar, álcool e outros) (Conab 2006). A produção nesta safra é estimada em 469,8 milhões de toneladas, superior em 8.9% a da safra anterior, que foi de 431,4 milhões de toneladas. A região Centro-Sul é responsável por 86.8% da produção nacional, sendo que a região Sudeste detém 69.4% da produção nacional, correspondendo a 80% da produção do Centro-Sul. Enquanto que a região Norte-Nordeste é responsável pelos 13.2% da produção nacional. O respectivo crescimento ocorreu em função da expansão de 5.4% na área, que passou de 5,8 para 6,2 milhões de hectares, e de 3.4% na produtividade média, que passou de 73,868 para 76,353 kg/ha. Este incremento é dito, segundo levantamento da Conab, fruto do clima e dos investimentos ocorridos nas indústrias atraídas pelos preços de mercado. Do total da produção, 50.5% são destinadas à produção de açúcar, 39.6% à produção de álcool, enquanto que os 9.9% restantes são destinados para fabricação de cachaça, alimentação animal, sementes, fabricação de rapadura, açúcar mascavo, dentre outros fins (Conab 2006).

O Brasil conta com uma posição privilegiada para atender às necessidades de maiores importações tanto de açúcar quanto de álcool anidro para fins combustíveis. O país tem duas regiões produtoras, Centro-Sul e no Norte-Nordeste, com safras alternadas, podendo manter sua presença no mercado mundial ao longo de todo o ano. Conta com uma avançada tecnologia de produção de álcool, além de ter o menor custo de produção do mundo e ainda possuir potencial de expansão de área plantada e de produtividade.

2.2 Classificação taxonômica

A cana-de-açúcar é uma planta herbácea, alógama, cultivada em regiões tropicais e subtropicais. Pertence à família das gramíneas (Poaceae), tribo Andropogoneae e gênero *Saccharum* (Tabela 1).

Tabela 1 Classificação taxonômica da cana-de-açúcar (Daniels e Roach 1987).

FAMÍLIA	SUBFAMÍLIA	TRIBO	GRUPOS DE SUBTRIBO	SUBTRIBO	GÊNERO
Gramineae	Pooideae				
	Panicoideae	Paniceae			
		Andropogoneae	Dimeriinae Hack	"Saccharastrae"	Imperata Cyr
			Saccharinae Benth		Eriochrysis P. Beauv.
			Germainiinae WD Clayton		Ecooilopus Steud
			Arthraxoninae Benth.		Spodiopogon Trin.
			Andropogoninae Presl.		Miscanthidium Stapf.
			Anthistiriinae Presl.		Erianthus Michx sect. Ripidium Henrard
			Ischaeminae Presl.		Miscanthus Anderss. sect. Diandra Keng
			Rottboeliinae Presl.		Sclerostachya (Hack.) A. Camus
			Tripsacinae Presl.		Narenga Bor.
			Coicinae		Saccharum L.
			Chionachninae WD Clayton	"Eulaliastrae"	
	Chloridoideae				
	Arundinoideae				
	Bambusoideae				

"Complexo *Saccharum*"

2.2.1 Gênero *Saccharum*

O gênero *Saccharum* é caracterizado pelo alto nível de poliploidia e aneuploidia. Seis espécies de cana-de-açúcar constituem este gênero (Naidu e Sreenivasan 1987; Roach e Daniels 1987), entre as quais distinguem-se duas categorias: (1) as espécies selvagens: *S.*

robustum Brandes e Jesweit ex Grassl e *S. spontaneum* L. e (2) as espécies domesticadas: *S. officinarum* L., *S. barberi* Jesw., *S. sinense* Roxb. e *S. edule* Hassk.

Espécies Selvagens

S. spontaneum é uma espécie rústica, altamente poliplóide, que pode colonizar regiões com características extremamente contrastantes, como temperatura, umidade e tipo de solo. Em conseqüência, esta espécie apresenta vasta expansão geográfica, que vai do Japão ao leste da África, passando pelo sudeste da Ásia, pelo continente indiano, pelo Oriente Médio e a bacia mediterrânea (Brandes *et al* 1939). Os clones desta espécie contêm baixo teor de sacarose, entretanto possuem resistência a pragas e doenças, capacidade de rebrota de soqueira, vigor e grande adaptabilidade (Naidu e Sreenivasan 1987). O número de cromossomos varia de 2n = 40 a 128, sendo os citótipos mais freqüentes os múltiplos de 8, sugerindo que o número básico desta espécie é x = 8 (Panje e Babu 1960; Sreenivasan *et al* 1987; Burner 1987). De fato, D'Hont e colaboradores (1998) confirmaram este número por hibridização *in situ* de rDNAs, além de indicar que o nível de ploidia varia entre 8 e 12, segundo os citótipos estudados.

S. robustum é a segunda espécie selvagem do gênero *Saccharum*. Nova Guiné e as ilhas de Melanésia constituem o centro de origem desta espécie, que permaneceu endêmica nesta região (Daniels e Roach 1987). Distingue-se de *S. spontaneum* pela ausência de rizomas, inflorescência grande, haste mais espessa e maior altura (Stevenson 1965). Foram encontrados dois citótipos euplóides, com 2n = 60 ou 2n = 80, e citótipos aneuplóides, variando de 2n = 63 a 205. Estes clones aneuplóides podem corresponder a híbridos naturais entre clones de *S. robustum* ou entre *S. robustum* e outras espécies (Price 1957a, 1965a).

Espécies domesticadas

S. officinarum é conhecida como a espécie produtora de açúcar ou 'cana nobre', por apresentar certas qualidades agronômicas e industriais correspondentes aos mais importantes critérios de seleção: colmos grossos, alto teor de sacarose e baixo conteúdo de fibra e amido (Bremer 1961a; Roach 1986). Acredita-se que esta espécie se originou das formas 2n = 80 da espécie selvagem *S. robustum*, na Nova Guiné, seu centro de origem (Brandes 1958). A partir desta região, os clones de *S. officinarum* se dispersaram para as ilhas vizinhas e, em seguida, para várias regiões que se tornaram os atuais países produtores de açúcar. Embora considerada, inicialmente, muito homogênea, *S. officinarum* apresenta características agronômicas que são suficientemente variáveis nos clones (Daniels e Roach 1987), como mostrado por um estudo da diversidade genética realizada através de marcadores moleculares (Jannoo *et al*. 1999a). *S. officinarum* é uma espécie euplóide, com 2n = 80 cromossomos, na sua grande maioria (Bremer 1930; Li e Price 1967; Price e Daniels 1968). Acredita-se que alguns poucos clones que não apresentam este número cromossômico se originaram de hibridação com outras espécies (Bremer 1924). *S. officinarum* é uma espécie octaplóide, apresentando um número cromossômico de base de x = 10 (D'Hont *et al.* 1998).

S. barberi **e** ***S. sinense*** são espécies que foram cultivadas, respectivamente, no norte da Índia e no sul da China (Stevenson 1965). Acredita-se derivarem de hibridações naturais entre as espécies *S. officinarum* e *S. spontaneum* (Price 1968), nas suas respectivas regiões de origem, sendo híbridos de primeira geração, uma vez que poucas recombinações interespecíficas entre os cromossomos foram observadas por hibridização *in situ* (D'Hont *et al* 1996). Distinguem-se dos clones de *S. officinarum* devido suas características florais, alto teor de fibras e a sua grande rusticidade devido a melhor tolerância aos estresses

ambientais. Um grande número de clones destas espécies é resistente às principais doenças da cana-de-açúcar (Daniels e Roach 1987). Seus números cromossômicos variam entre 2n = 81 a 124 em *S. sinense* e 2n = 111 a 120 em *S. barberi*. Observações citológicas mostram meioses extremamente irregulares (Sreenivasan *et al* 1987).

S. edule é um grupo menor de cana estéril, delimitado na Nova Guiné e nas ilhas vizinhas. Os clones desta espécie são provavelmente originados da espécie *S. robustum*, bem como de origem interespecífica com *S. robustum* como doadora do gameta feminino (Grivet *et al*. 2004). Os perfis moleculares mitocondriais e cloroplásticos destes clones de *S. edule* são identificados ao perfil mais freqüente observados na espécie *S. robustum* (D'Hont *et al* 1993; Sobral *et al*. 1994). Os clones constituem uma série de poliplóides com 2n = 60, 70 ou 80 cromossomos (Roach 1972).

2.2.2 Complexo '*Saccharum*'

Murkherjee (1954, 1957) revisou o gênero *Saccharum*, sua origem e distribuição. Demonstrou-se que os gêneros *Saccharum*, *Erianthus* (sect *Ripidium*), *Sclerostachya* e *Narenga* constituíam um grupo de intercruzamentos muito próximo. Murkherjee, então, cunhou o termo 'complexo *Saccharum*' para descrever este grande grupo, porém esta não é uma designação taxonômica formal. O complexo *Saccharum* foi revisado por Daniels e colaboradores (1975c) e o gênero *Miscanthus* (sect *Diandra*) foi adicionado ao grupo (Tabela 1).

Sugeriu-se que a cana-de-açúcar emergiu de complexos padrões de hibridações envolvendo espécies destes diferentes gêneros, principalmente, *Saccharum*, *Erianthus* e *Miscanthus* (Daniels e Roach 1987). Na realidade, parece que o conceito de 'complexo *Saccharum*' conduz a sobreestimar a contribuição de gêneros aparentados ao surgimento da cana-de-açúcar cultivada. No entanto, esta suposição não é confirmada por pesquisas moleculares que mostram a cana-de-açúcar originária especificamente do filo *Saccharum* (Besse *et al*. 1997; Grivet *et al.* 2004). Por exemplo, a ausência de hibridização de sondas específicas de *Miscanthus* e *Erianthus* sobre o DNA de diferentes cultivares de cana-de-açúcar é um elemento que clarifica tal tese (Alix *et al*. 1998; Alix *et al*. 1999). Os três gêneros *Saccharum*, *Erianthus* e *Miscanthus* podem ser seguramente diferenciados e *Erianthus* e *Miscanthus* não parecem ter contribuído com os cultivares de cana-de-açúcar (D'Hont *et al*. 1995; Besse *et al*. 1997; Alix *et al*. 1998; Alix *et al*. 1999; Nair *et al*. 1999; Piperidis *et al*. 2000).

2.3 Cultivares modernos de cana-de-açúcar

2.3.1 Nobilização e os primeiros híbridos interespecíficos

Estimulados pelo surgimento de doenças, os primeiros híbridos interespecíficos induzidos pelo homem foram produzidos em Java, envolvendo essencialmente as espécies *S. officinarum* e *S. spontaneum* e, simultaneamente na Índia, envolvendo *S. officinarum*, *S. spontaneum* e *S. barberi*. Os primeiros trabalhos neste domínio começaram em uma estação de seleção em Java, realizados por melhoristas holandeses. Estes designaram pelo termo

'nobilização' a ação consistente no cruzamento de um clone 'nobre', rico em açúcar, com um clone de uma espécie aparentada, vigoroso ou resistente a doenças, e diversos retrocruzamentos ao acaso do híbrido obtido com o mesmo parental 'nobre' ou outro, de maneira a recuperar um fenótipo cultivável mais rústico e vigoroso (Figura 1). Estes primeiros híbridos interespecíficos permitiram um progresso genético considerável apresentando grande utilidade nos programas de melhoramento da cana-de-açúcar, solucionando alguns problemas de doenças além de fornecerem benefícios adicionais no aumento da produção e da adaptabilidade de crescimento sob diversas condições de estresse (Roach 1972).

Os primeiros híbridos interespecíficos oriundos destes trabalhos de nobilização foram utilizados maciçamente em cruzamentos em todos os programas de seleção através do mundo. Trata-se notadamente dos clones POJ2878, POJ2725, POJ213 de Java e os clones Co281, Co290 de Coimbatore. Estes primeiros híbridos interespecíficos se encontram na genealogia de quase todas as variedades cultivadas atualmente.

De fato, os cultivares modernos de cana-de-açúcar derivam essencialmente dos cruzamentos entre estes primeiros híbridos que envolvem um pequeno número de clones parentais (Price 1965b). Trata-se de híbridos complexos interespecíficos e aneuplóides, com número de cromossomos variando entre 2n = 100 a 130 (Sreenivasan *et al* 1987). Este número elevado de cromossomos é conseqüência da transmissão cromossômica irregular nas primeiras etapas da nobilização. *S.officinarum*, utilizado como parental feminino, transmite seu número somático de cromossomos (2n = 80 cromossomos) no momento do primeiro cruzamento F1, enquanto que o parental masculino, *S. spontaneum*, transmite apenas seu número 'n' de cromossomos. Este fenômeno persiste até a segunda geração da nobilização, correspondente ao primeiro retrocruzamento ao *S. officinarum* (RC1). Somente

a partir da terceira nobilização (RC2) que o parental *S. officinarum* passa a transmitir 'n' cromossomos (Bremer 1961b, 1961c). O mecanismo exato da transmissão '2n + n' cromossomos ainda não foi perfeitamente estabelecido (Bhat e Gill 1985). Várias hipóteses foram levantadas para explicar este fenômeno: (1) duplicação cromossômica através de uma endoduplicação, (2) duplicação cromossômica após a primeira divisão meiótica, (3) fusão dos dois núcleos durante o estágio de tétrade. Esta transmissão '2n + n' leva a uma rápida redução do número relativo de cromossomos de *S. spontaneum* e um rápido retorno ao tipo cultivado *S. officinarum*.

Os híbridos naturais ou induzidos entre *S. officinarum* e *S. spontaneum* são geralmente férteis. As associações cromossômicas nestes híbridos são regulares, com formação de bivalentes e algumas raras univalentes na metáfase I (Bremer 1961a, 1961b, 1961c; D'Hont *et al* 1996). No entanto, o modo preciso de pareamento dos cromossomos ainda não foi definido, mas as associações entre os cromossomos interespécies existem, pois cromossomos recombinantes foram observados em algumas variedades comerciais (Grivet *et al* 1996; D'Hont *et al* 1996). *S. officinarum* e *S. spontaneum* têm o número básico de cromossomos diferente (D'Hont *et al* 1996). Assim, nos híbridos interespecíficos cromossomos homólogos coexistem, apresentando diferenças na organização estrutural (Grivet *et al* 1996; D'Hont *et al* 1996).

Até o momento sabe-se que o esquema de pareamento dos cromossomos durante a meiose é complexo, porém a disomia é muito improvável (Grivet *et al* 1996; Hoarau *et al* 2001) indicando que todos os cromossomos homólogos são susceptíveis à recombinação. Esta organização genômica infere que cada cópia simples do gene é representada por volta de 10 alelos, que correspondem potencialmente a um haplótipo diferente. Entre os 10

alelos, cerca de 8 a 9 devem ser herdados do *S. officinarum* e 1 ou 2 do *S. spontaneum* (Grivet *et al* 2003).

Segundo os estudados já realizados, observa-se que os cultivares comerciais de cana-de-açúcar apresentam complexa poliploidia e genomas aneuplóides, acarretando dificuldade de interpretação para a genética clássica, a genética molecular e os estudos de melhoramento genético, uma vez que as informações de estrutura e organização do genoma têm sido especulativas. Recentemente, estudos moleculares na cana-de-açúcar e em espécies relacionadas, como milho, arroz e sorgo, têm fornecido novos dados que podem ser úteis para o melhor entendimento da genética da cana-de-açúcar e seus progenitores.

2.3.2 Estrutura genômica

A cana-de-açúcar possui um genoma de grande tamanho e complexidade, principalmente devido ao alto nível de poliploidia. O tamanho do genoma monoplóide da cana-de-açúcar é sensivelmente equivalente ao do sorgo e duas vezes maior que o do arroz. O genoma monoplóide da *S. officinarum* compreende 930Mpb e da *S. spontaneum* 750Mpb (D'Hont e Glaszmann 2001). No genoma dos cultivares modernos de cana-de-açúcar cada cromossomo de base é potencialmente presente em dez a doze exemplares homólogos. Por conseguinte, o tamanho total do genoma de cultivares de cana-de-açúcar é maior que o do milho, 5500Mpb (2n = 20), sorgo, 1600Mpb (2n = 20), ou arroz, 860Mpb (2n = 24), refletindo a alta poliploidia dos cultivares de cana-de-açúcar (D'Hont e Glaszmann 2001).

Hibridização genômica *in situ* (GISH) mostrou que 15 a 25% do genoma de cultivares modernos são derivados de *S. spontaneum* (D'Hont *et al* 1996; Piperidis e D'Hont 2001) e juntamente com mapeamento genético (Grivet *et al* 1996; Hoarau *et al* 2001) demonstraram a ocorrência de recombinações entre cromossomos dos dois parentais. Por exemplo, o cultivar R570 é constituído de 80% de cromossomos provenientes de *S. officinarum*, de 10% de cromossomos herdados de *S. spontaneum* e 10% de cromossomos derivados de recombinações entre cromossomos das duas espécies ancestrais (D'Hont *et al*. 1996).

O número básico de cromossomos no gênero *Saccharum* já foi estudado por diversos autores que chegaram a conclusões divergentes. Vários números – 5, 6, 8, 10 e 12 - foram propostos, porém os números mais prováveis são 8 e 10 (Nishiyama 1956; Bremer 1961a, Sreenivasan *et al*. 1987). Hibridização fluorescente *in situ* (FISH) e genes rDNA indicaram que os 80 cromossomos de *S. officinarum* se encontram organizados em 8 cópias homólogas, de um conjunto básico de 10 cromossomos diferentes (2n = 10x = 80), enquanto que os 40 a 128 cromossomos de *S. spontaneum* estão organizados em 5 a 12 cópias homólogas, com conjunto básico de 8 cromossomos diferentes (2n = 8x = 40-128) (D'Hont *et al* 1998; Ha *et al* 1999). A diferença no número básico de cromossomos implica que diferenças estruturais separam os dois genomas. Como conseqüência, duas organizações cromossômicas distintas coexistem nos cultivares modernos de cana-de-açúcar. De fato, tal organização é particularmente intricada: (1) o número cromossômico básico de tais híbridos é provavelmente maior que 10 e a proporção de regiões duplicadas em seus genomas monoplóides é potencialmente maior que nas espécies ancestrais; (2) a freqüência de cromossomos não pareados em muitos grupos de homologia, devido a aneuploidia, implica a perda de alguns cromossomos em cada geração, o que pode resultar

em uma gradual erosão do número de cromossomos após uma série de cruzamentos recorrentes (Butterfield *et al* 2001).

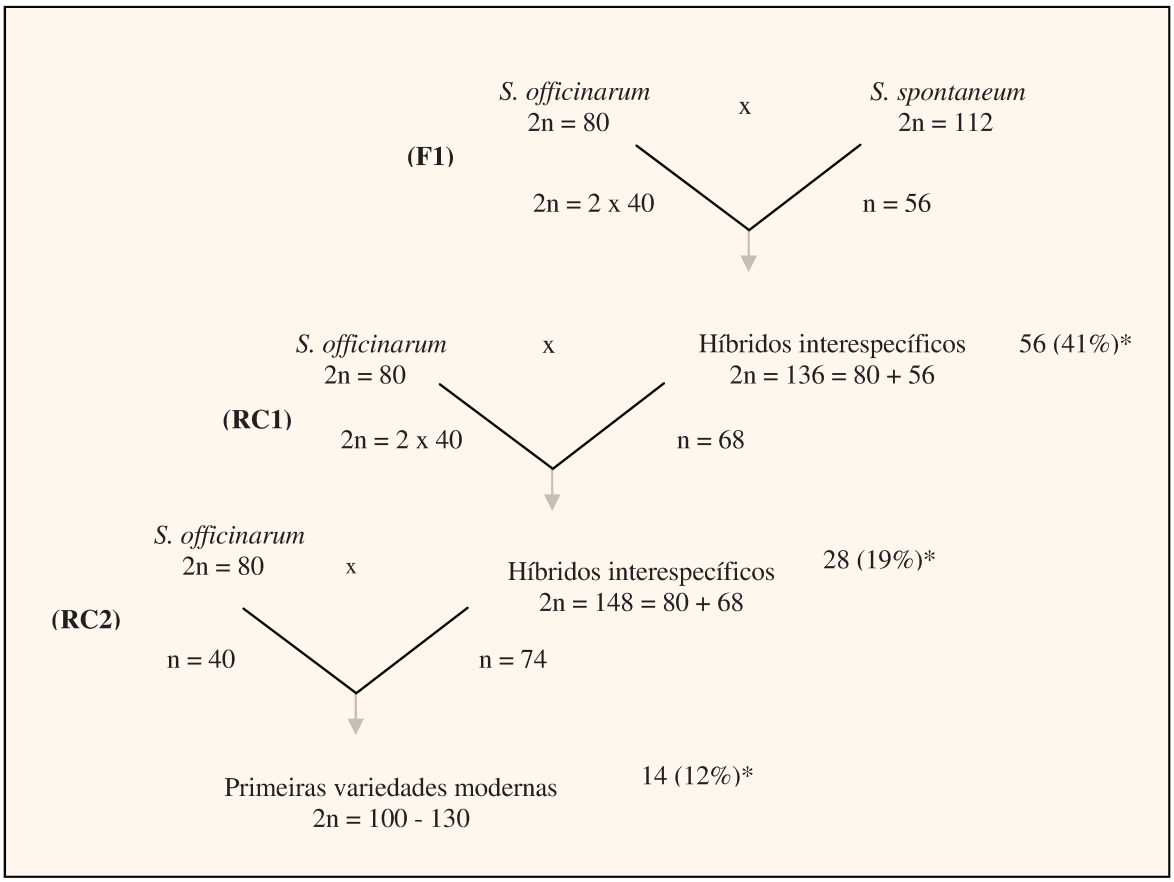


Figura 1 Transmissão dos cromossomos durante o processo de nobilização que gerou as variedades modernas de cana-de-açúcar. *S. officinarum*, parental feminino, e *S. spontaneum*, parental masculino. *Número e porcentagem estimada de cromossomos de *S. spontaneum* nos híbridos interespecíficos. (F1): primeira nobilização; (RC1): segunda nobilização; (RC2): terceira nobilização.

2.4 Diversidade genética no gênero *Saccharum*

Devido ao pequeno número de clones das espécies de *Saccharum* empregados nos primeiros cruzamentos, a base genética das variedades híbridas modernas apresenta-se estreita e esta pode ser a razão do lento progresso dos programas de melhoramento em cana-de-açúcar (D'Hont *et al.* 1995). O entendimento e manutenção da variação natural presente dentro dos cultivares domésticos e selvagens relacionados destas espécies são importantes no estabelecimento de um programa eficiente. Devido à predominância de *S. officinarum* no genoma dos cultivares, e da importância dos caracteres agronômicos inerentes a esta espécie, é particularmente importante avaliar a diversidade existente dentro das espécies e a proporção desta diversidade presente nos cultivares modernos.

Vários estudos moleculares foram conduzidos a fim de acessar a diversidade do germoplasma no gênero *Saccharum*. Análises isoenzimáticas (Glaszmann *et al* 1989), de RFLPs utilizando sondas citoplasmáticas heterólogas (D'Hont *et al* 1993), assim como rDNA (Glaszmann *et al* 1990), seqüências nucleares de baixo número de cópias (Burnquist *et al* 1992) e SSRs (Cordeiro *et al*. 2003) foram efetuadas. Os resultados obtidos estavam de acordo com os esquemas taxonômicos e filogenéticos baseados em dados morfológicos, citológicos e bioquímicos (Daniels e Roach 1987) e revelaram uma variabilidade limitada dentro de *S. officinarum*.

Lu e colaboradores (1994a, 1994b), através de análises de RFLP, concluíram que, dentro dos grupos, a similaridade genética foi mínima nas espécies selvagens, próxima da metade encontrada em *S. officinarum*. Sendo levemente mais alta em *S. robustum* que em *S. spontaneum*. Observou-se que *S. officinarum* é tão fortemente diferenciada de *S. spontaneum* que dois clones destas espécies apresentarão apenas 1/50 dos seus marcadores

em comum (Lu *et al*. 1994a). Logo, as três espécies *S. spontaneum*, *S. officinarum* e *S. robustum* são claramente distintas, embora as duas últimas apareçam mais relativamente próximas. Em suma, a variabilidade genética é mais alta em *S. spontaneum*, intermediária em *S. robustum* e mais baixa em *S. officinarum*. As espécies secundárias *S. barberi* e *S. sinense* combinam os perfis moleculares de *S. officinarum* e *S. spontaneum*, o que confirma suas origens como híbridos naturais entre as duas espécies. As variedades comerciais também foram localizadas em relação às espécies ancestrais, sendo colocadas entre os genótipos de *S. officinarum* e *S. spontaneum*, porém mais próximas à primeira espécie. Este resultado ilustra o efeito da nobilização. De fato, Jannoo e colaboradores (1999a) confirmaram que o essencial da diversidade da espécie *S. officinarum* é encontrada nos cultivares modernos, uma vez que apenas 15% dos alelos de *S. officinarum* não foram recuperados nos 109 cultivares estudados.

As análises de RFLP ainda mostraram que os cultivares modernos de cana-de-açúcar são altamente heterozigotos, apresentando vários alelos distintos em cada loco. Visto que o pequeno número de divisões meióticas desde os primeiros cruzamentos que originaram estes cultivares provê pouca oportunidade de recombinação entre os cromossomos ancestrais (Grivet e Arruda 2001), confirmou-se que uma importante parte desta diversidade deve-se ao complemento cromossômico herdado de *S. spontaneum* (Lu *et al*. 1994b; Jannoo *et al*. 1999a).

Cultivares de cana-de-açúcar representativos em vários programas internacionais de melhoramento foram estudados com marcadores moleculares RFLP, AFLP e SSR, em diversas investigações (Lu *et al* 1994a; Jannoo *et al* 1999a; Lima *et al* 2002). De modo geral, dendogramas baseados em índices de similaridade mostraram uma fraca estruturação entre os cultivares estudados. Estes resultados refletem a alta ploidia dos cultivares, sua alta

heterozigosidade, que os possibilita guardar individualmente uma grande proporção de alelos contribuídos pelos parentais dos cruzamentos interespecíficos iniciais. Além disto, a falta de clara estruturação entre variedades de diferentes programas de melhoramento (Lu *et al* 1994b) ilustra a tradição de larga troca de materiais parentais entre estações experimentais de cana-de-açúcar. Esta fraca diferenciação é, principalmente, atribuída aos marcadores específicos de *S. spontaneum* e ilustra o pequeno número de parentais *S. spontaneum* empregados na ancestralidade da maioria das variedades modernas.

2.5 Poliplóides

2.5.1 Conceitos fundamentais e poliploidia na cana-de-açúcar

Inúmeras plantas de interesse são poliplóides, como é o caso da cana-de-açúcar. Uma planta é dita poliplóide se dentro de suas células somáticas existir mais de dois cromossomos dentro de uma mesma classe de homologia (Allard 1960). Se o número total de cromossomos é múltiplo do número base de cromossomos, a planta é dita euplóide, caso contrário ela é dita aneuplóide.

Classicamente, os organismos poliplóides podem ser divididos em alopoliplóides e autopoliplóides com base em sua origem, nas relações fenotípicas e na evolução. Os autopoliplóides são derivados da multiplicação de um mesmo genoma de base, ou seja, da multiplicação do número básico de cromossomos do indivíduo, enquanto que os alopoliplóides envolvem uma origem ou domesticação a partir de hibridações interespecíficas ou intergenéricas e, portanto, são constituídos por conjuntos de cromossomos de espécies distintas (Da Silva e Sobral 1996). Entretanto, muitas vezes o

caminho evolutivo de determinadas espécies não é bem definido por ausência dos diplóides ancestrais, devido à extinção dos mesmos ou ao fato destes não serem conhecidos e, também, pode ser que haja ausência de relações filogenéticas que possam determiná-los.

No estudo de marcadores mendelianos em poliplóides, além do modo de formação e da origem filogenética, deve-se considerar o comportamento dos cromossomos no momento da meiose. Três casos extremos podem ser distinguidos: (1) formação de bivalentes e segregação dissômica, que compreende principalmente os alopoliplóides; (2) formação de bivalentes e segregação polissômica, na qual os cromossomos se pareiam ao acaso dentro de uma classe de homologia, que compreende principalmente os autopoliplóides e eventualmente alguns alopoliplóides; (3) e formação de multivalentes, onde os cromossomos homólogos se pareiam todos juntos dentro de uma mesma classe de homologia, categoria que concerne principalmente os autopoliplóides, mas pode ser identificado em alguns alopoliplóides (Jannoo 1998).

Jackson e Casey (1980) concluíram que o comportamento dos cromossomos durante o pareamento cromossômico representava uma chave para o entendimento da evolução dos poliplóides. Assim, baseando-se em fatores que afetam o pareamento cromossômico, considera-se que todo alopoliplóide tem interação dissômica. Para autopoliplóides é esperada a interação polissômica, a qual quase não é detectada em autogamia, porém é bem evidente em alogamia, propiciando a heterozigosidade e a heterose (Mac Key 1987; Uhl 1992). Entretanto, novos estudos demonstraram que o comportamento genético em poliplóides parece ser mais complexo, particularmente dentro das gramíneas, sugerindo o desenvolvimento de mecanismos de controle de pareamento de bivalentes (Jackson 1982).

Como dito acima, a segregação dissômica é característica dos organismos alopoliplóides. De fato, nos organismos alopoliplóides, os cromossomos homólogos têm tendência de se parear preferencialmente entre eles. Para cada genoma, uma vez que há a presença de dois cromossomos por classe de homologia, o comportamento cromossômico aproxima-se do observado nos organismos diplóides, nos quais há uma formação quase exclusiva de bivalentes na meiose e uma segregação do tipo dissômica. Assim, os alopoliplóides caracterizam-se por uma heterozigosidade fixa, resultante da combinação de genomas parentais divergentes, formação de bivalentes durante a meiose e herança dissômica. Enquanto que, os autopoliplóides podem apresentar até formação de multivalentes na meiose e são caracterizados por uma herança polissômica (Stebbins 1950).

A constituição genômica de muitos poliplóides, com alto nível de ploidia, ainda não é completamente entendida. Análises com metodologias especialmente desenvolvidas para poliplóides promoveram auxílio no entendimento e distinção entre auto e alopoliplóides (Wu *et al.* 1992), no entanto muitos mecanismos ainda precisam ser esclarecidos.

Na cana-de-açúcar, a poliploidia atinge todas as espécies do gênero *Saccharum*, que não apresentam nenhuma espécie diplóide aparentada conhecida, sendo as espécies aneuplóides as mais frequentes. A observação da progênie de uma autofecundação de uma espécie aneuplóide de *S. spontaneum* (2n=63) gerou uma série de plantas com 2n=50 a 110, entre as quais foram observadas diferentes formas de aberrações meióticas e variabilidades morfológicas. Estas observações sugerem que as plantas euplóides e aneuplóides da espécie *S. spontaneum* resultam de cruzamentos naturais e de autofecundações (Janaki Amnal 1936; Bremer 1961a). Da Silva *et al.* (1993a, 1995) e Al Janabi *et al*. (1993) observaram que a espécie *S. spontaneum* (2n=64) comportava-se como um autopoliplóide com herança polissômica, apesar de apresentar pareamento de bivalentes na meiose, encontrado na

caracterização de alopoliplóides. Na espécie *S. officinarum*, Jagathesan *et al* (1970) distinguiram três tipos de clones em função de seus comportamentos meióticos: (1) o tipo euplóide com comportamento meiótico normal e formação regular de bivalentes; (2) o tipo aneuplóide com uma baixa freqüência de anomalias meióticas; (3) e o tipo de origem híbrida com um nível elevado de anomalias meióticas.

Os clones das diferentes espécies do gênero *Saccharum* têm uma meiose geralmente normal com formação de bivalentes (Jagathesan *et al*. 1970). Entretanto, irregularidades meióticas, como a formação de univalentes ou até algumas multivalentes, são observadas e são mais freqüentes na espécie *S. spontaneum* que na *S. officinarum* (Sreenivasan 1975; Sreenivasan e Jagathesan 1975).

As associações cromossômicas das variedades comerciais de cana-de-açúcar são regulares com formação principalmente de bivalentes e algumas raras univalentes na metáfase I (Bremer, 1961a, b, c). Todavia, o modo preciso de pareamento dos cromossomos ainda não foi definido, porém as associações cromossômicas entre espécies existem, pois cromossomos recombinantes foram observados nestas variedades comerciais (D'Hont *et al*, 1996; Grivet *et al*, 1996).

O comportamento do pareamento dos cromossomos em plantas de cana-de-açúcar também foi investigado através das análises de repulsão entre marcadores mapeados. *S. robustum* (MOL5829, 2n=80) apresentou uma alta proporção de pareamento preferencial (50%) nos poucos grupos de co-segregação já definidos (Al-Janabi *et al*, 1994a; Ming *et al*, 1998). Em *S. officinarum*, alguns pareamentos preferenciais também foram observados, enquanto que nenhum pareamento preferencial foi encontrado em *S. spontaneum* (Al-Janabi *et al*, 1994a; Ming *et al*, 1998). No cultivar R570, Grivet e colaboradores (1996) observaram uma polissomia com poucos casos de pareamento preferencial. Este resultado

foi confirmado e estendido por Hoarau *et al*. (2001), cujos dados também sugeriram a possibilidade de dissomia local completa. No R570, o pareamento preferencial detectado concerne os cromossomos do *S. officinarum*, *S. spontaneum*, assim como a origem interespecífica-recombinante.

Diante de todas as informações disponíveis na literatura a respeito dos tipos de formação e comportamento dos cromossomos nas plantas de cana-de-açúcar, pode-se observar que muito ainda precisa ser esclarecido, apesar de tantos anos de estudos. A natureza poliplóide destes organismos dificulta o completo entendimento. Com o avanço das técnicas de biotecnologia, espera-se obter mais conhecimento a cerca dos mecanismos meióticos destas plantas e assim determinar o comportamento dos cromossomos, suas associações e tipos de segregação, além de confirmar os dados já existentes.

2.5.2 Detecção de polimorfismo em poliplóides

Uma vez definida a poliploidia e entendidas as dificuldades de se trabalhar com estes organismos complexos, pode-se levantar mais uma questão que torna ainda mais complicado o estudo com os poliplóides, que é a detecção de polimorfismo, durante o estudo de marcadores moleculares co-dominantes, como os microssatélites (SSRs) e os RFLPs.

Tais marcadores têm herança co-dominante devido à capacidade de distinguir entre os homozigotos e heterozigotos para o loco em questão. A existência de grande disponibilidade desses marcadores ao longo do genoma, abriu perspectivas para estudos genéticos, entre eles, a possibilidade da construção de mapas genéticos.

Entretanto, quando se trata de organismos poliplóides, a complexidade envolvida no uso de marcadores co-dominantes aumenta. A grande quantidade de fragmentos gerados, que podem representar os diferentes alelos do mesmo loco nos vários cromossomos homólogos envolvidos, não permite identificação dos genótipos pelo fenótipo visualizado.

Vários estudos foram realizados para tentar superar tal dificuldade, como a metodologia proposta por Wu *et al.* (1992), em que é considerado o pareamento de bivalentes na meiose. O mapeamento genético é realizado através de fragmentos de dosagem única no genoma, denominados marcadores em dose simples (MDS) ou SDRF (*Single Dose Restriction Fragment*) detectados pela segregação na proporção 1:1 (Figura 2a). O padrão de polimorfismo é representado pela ausência ou presença de bandas individuais. O grande número de marcadores leva a uma leitura dos perfis dos indivíduos da progênie, comparados aos parentais (Figura 2c). Nesse caso, os marcadores, como os SSRs e RFLPs, perdem a vantagem da co-dominância e passam a funcionar como um marcador dominante. No entanto, no caso de segregação 1:1, isso não representa nenhuma desvantagem.

Um MDS é equivalente a um único alelo em autopoliplóides, ou a um alelo no estado heterozigoto em um genoma diplóide e em alopoliplóides. Sua segregação equivale a um alelo simples ou no estado de heterozigoto nos gametas: metade dos gametas contém o DNA de um fragmento e a outra metade não. Estes tipos de gametas podem ser visualizados na progênie de um determinado cruzamento, onde tal fragmento está presente em um parental e ausente no outro. Assim, o fenótipo dos marcadores MDS representa o tipo de gameta de um parental ao invés de apresentar a combinação dos gametas de ambos.

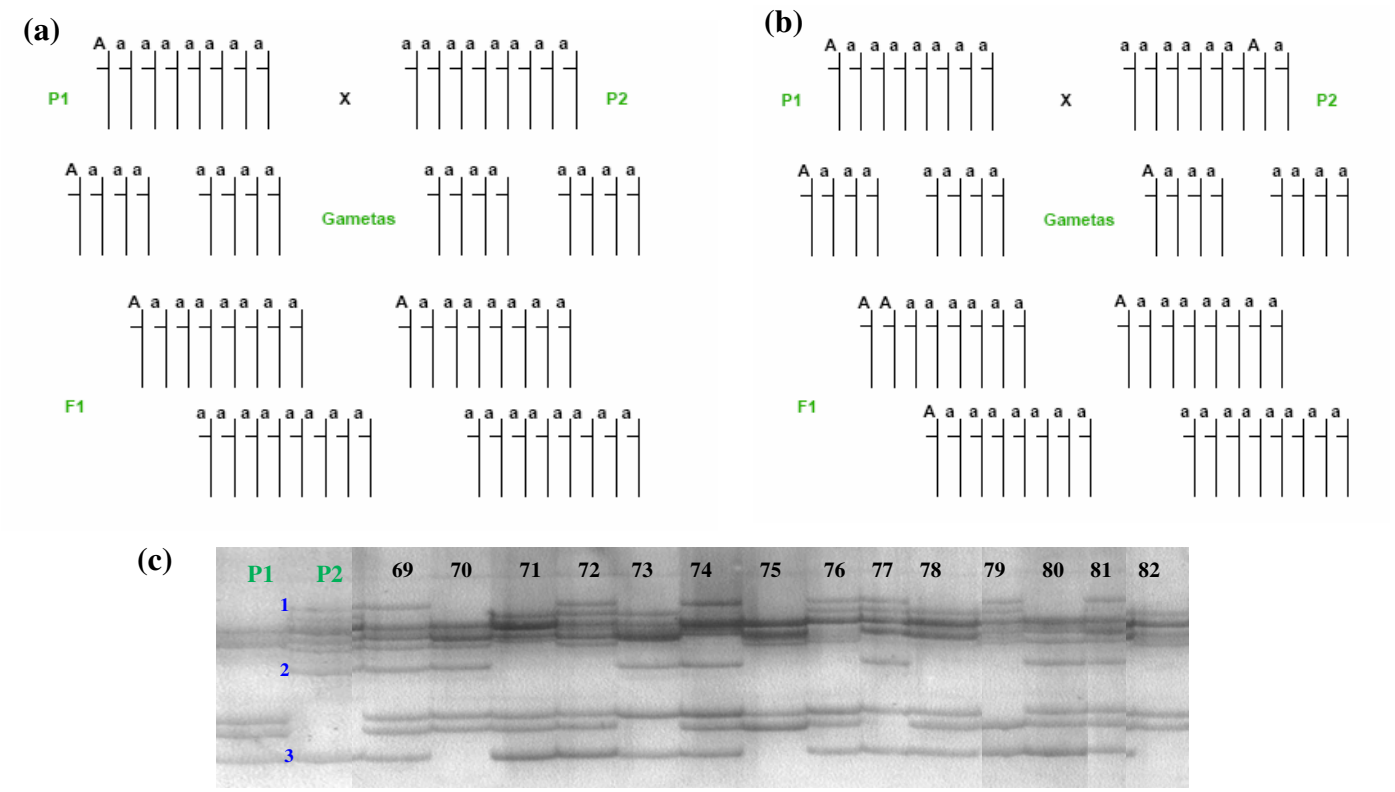


Figura 2 **(a)** Esquema ilustrando a distribuição dos alelos A e a, pertencentes ao mesmo loco, entre os gametas formados por cada um dos parentais (P1 e P2) e possíveis genótipos observados entre os descendentes da progênie F1. O alelo A é representado com cópia única entre os cromossomos homólogos do parental P1. Assim, haverá a formação de dois tipos de gametas: um tipo contendo este alelo e outro não. O alelo A não está presente no parental P2, resultando na formação de gametas sem este alelo em seu genótipo. Baseando-se no encontro aleatório dos possíveis gametas originados na meiose, os gametas do parental P2 combinam-se de duas formas diferentes: com um gameta do parental P1 que contêm o alelo A ou com um gameta sem a presença deste alelo. Desta forma, espera-se que metade dos indivíduos formados pela união desses gametas tenha o alelo A e a outra metade não, seguindo a relação esperada para um alelo em dosagem única no genoma (1:1). **(b)** Neste esquema o alelo A é representado como cópia única entre os cromossomos homólogos tanto no parental P1 quanto no parental P2. Verificando os possíveis genótipos formados para a progênie F1 constata-se que o alelo A estará presente na proporção de 3:1. Em ambos os casos, o alelo a está em dosagens variadas no genoma, possuindo um padrão de segregação mais complexo e de difícil análise. **(c)** Gel de poliacrilamida 6% mostrando um exemplo de segregação de marcadores SSR, em uma população de mapeamento de cana-de-açúcar. O grande número de marcadores leva a uma leitura dos perfis dos indivíduos da progênie, comparados aos parentais. Os marcadores 1 e 2 correspondem aos MDS que segregam na progênie na proporção de 1:1, enquanto que o marcador 3 está presente em ambos genitores com segregação 3:1 na progênie. P1 e P2: parentais da população de mapeamento; 69-82: amostra dos indivíduos da progênie F1. (Figura adaptada de Teixeira 2006).

Como se trata da genética de poliplóides, alternativas de segregação são possíveis de serem observadas. Por exemplo, observa-se a presença de um mesmo alelo nos dois genitores, estando ele ou não em dosagem semelhante no genoma de cada um dos genitores (Da Silva *et al.* 1993a).

Marcadores MDS também podem ser detectados com um comportamento segregante diferente do qual exposto até o momento. Tal situação é encontrada quando um mesmo alelo está presente nos dois genitores em dosagem única e a segregação deles na progênie F1 obedece à proporção 3:1 (Da Silva 1993b; Grivet *et al.* 1996) (Figura 2).

Wu *et al.* (1992) calcularam que para considerar um marcador como MDS com 95, 98 e 99% de nível de significância são necessários 54, 75 ou 92 indivíduos constituindo a família a ser estudada. Tanto para auto como para alopoliplóides, a detecção de ligação em fase de acoplamento de MDSs é uma função da fração de indivíduos recombinantes em uma população. Esses marcadores permitem a identificação de alelos em uma população relativamente pequena.

2.6 Marcadores moleculares

Microssatélites (SSRs) e ESTs como marcadores genéticos

Os marcadores moleculares são uma ferramenta valiosa no estudo de genomas complexos como o da cana-de-açúcar (Daugrois *et al* 1996). A incorporação destes na seleção de características econômicas durante os primeiros estágios de melhoramento, assim como, na escolha de melhores parentais em um cruzamento, pode reduzir significativamente o tempo de desenvolvimento de novas variedades. Estes objetivos

podem ser alcançados com sucesso com a disponibilidade de marcadores polimórficos robustos que co-segregam com características agrícolas economicamente importantes.

Um importante progresso no conhecimento da estrutura genômica e na genética dos cultivares modernos de cana-de-açúcar tem sido alcançado nas últimas décadas graças ao emprego de marcadores moleculares. As informações coletadas sobre a diversidade entre cultivares (Lu *et al* 1994a; Jannoo *et al* 1999a; Lima *et al.*, 2002), espécies parentais (Glaszmann *et al* 1990; Burnquist *et al* 1992; D'Hont *et al* 1993; Lu *et al* 1994b; Jannoo *et al* 1999a; Nair *et al* 1999) e outros gêneros do complexo *Saccharum* (Al-Janabi *et al* 1994b; Besse *et al* 1996, 1997; Alix *et al* 1998, 1999) ajudam os melhoristas a refinarem as estratégias gerais para a exploração dos germoplasmas existentes. Os marcadores podem ser também utilizados para identificar os cultivares, para controlar a progênie e monitorar introgressão (D'Hont *et al* 1995; Harvey *et al* 1998; Cordeiro *et al* 2000; Piperidis *et al* 2001). No entanto, muito deve ser feito antes que a seleção assistida de marcadores para características agronômicas quantitativas possa ser aplicada.

A seleção assistida por marcadores (SAM) fundamenta-se no conceito de que é possível inferir a presença de um gene a partir de um marcador fortemente ligado a este. Quando o marcador se encontra muito longe da região de interesse, a possibilidade de ambos serem transmitidos aos indivíduos da progênie é reduzida devido aos eventos de recombinação. Sendo assim, a existência de uma forte ligação entre a característica de interesse e o marcador é pré-requisito neste tipo de seleção (Kumar *et al.* 1999). A construção de um mapa de QTL's (*Quantitative Trait Loci*) requer a elaboração preliminar de um mapa de locos marcadores, o qual fornece a estrutura física para a alocação dos QTL's (Bearzoti 2000).

Os microssatélites ou SSRs (*Simple Sequence Repeat*) tornaram-se amplamente empregados nos estudos de marcadores de plantas. Estes marcadores, convencionalmente, são repetições em *tandem* de pequenas seqüências de nucleotídeos contendo uma a seis bases de comprimento. A variação no número de repetições resulta em locos polimórficos que são extremamente úteis, principalmente, em estudos de mapeamento. O polimorfismo dos marcadores SSRs é revelado pela reação em cadeia da polimerase (PCR) através da amplificação do DNA genômico total utilizando dois *primers* únicos compostos de seqüências curtas de nucleotídeos que flanqueiam e, portanto, definem o loco SSR. Estes marcadores são geralmente utilizados para mapeamento genético, *fingerprinting*, estudos populacionais e evolutivos, tanto para plantas como para animais (Bell e Ecker 1994; Yang *et al.* 1994; Senior *et al.* 1996; Russel *et al.* 1997; Petren *et al.* 1999).

O valor dos SSRs, como marcadores, é atribuído a sua natureza multialélica, herança codominante e transferibilidade entre espécies (Griffiths *et al*. 2002; Decroocq *et al*. 2003; Eujayl *et al*. 2004), facilidade de detecção pela PCR, abundância relativa e cobertura extensiva do genoma (Li *et al*. 2002) e necessidade de quantidades pequenas de DNA para dar início às reações (Powell *et al.* 1996). Os SSRs exibem também altas taxas de mutação (Vigoroux *et al*. 2002) e associação preferencial com regiões não-repetitivas do genoma (Morgante *et al*. 2002). Evidências substanciais sugerem que o tipo de motivo, comprimento e mutação dependem da localização dos SSRs (Morgante *et al.* 2002).

A habilidade dos SSRs em revelar alta diversidade alélica é particularmente proveitosa na discriminação entre os genótipos. O sucesso do uso deste marcador em outras espécies, como, cevada (Saghai *et al.* 1994; Russel *et al.* 1997), arroz (Wu e Tanksley 1993), trigo (Röder *et al.* 1995), maçã (Szewc-McFadden *et al.* 1996) e abacate (Lavi *et al.*

1994), estimularam a aplicação desta técnica para espécies mais geneticamente complexas, como a cana-de-açúcar (Cordeiro *et al*. 2000).

Análises prévias de todos os possíveis motivos de SSRs, em bancos contendo seqüências de plantas, revelaram freqüências variando a cada 29kb à 50kb, dependendo da espécie (Lagergrantz *et al.* 1993; Morgante e Olivieri 1993). A frequência de SSRs em plantas foi avaliada também por hibridização de *primers*, sugerindo uma variação de um SSR a cada 65kb a 80kb (Panaud *et al.* 1996; Eght e Maymarquardt 1997).

Os métodos convencionais de desenvolvimento de SSRs baseiam-se no isolamento e sequenciamento de clones contendo possíveis motivos SSRs, seguidos de desenho de *primers* para as regiões flanqueadoras dos motivos. Buscando a redução do custo e do tempo gasto, foram desenvolvidos processos de clonagem para criação de bibliotecas enriquecidas com SSRs (Edwards *et al*. 1996), no entanto, o desenvolvimento de marcadores SSR genômicos permaneceu ainda laborioso. A busca de SSRs em sequências expressas depositadas em bancos de dados públicos é uma estratégia alternativa, mais simples, rápida e econômica, no desenvolvimento dos SSRs.

A análise de ESTs (*Expressed Sequence Tags*) é uma estratégia simples para o estudo da porção expressa do genoma, mesmo em organismos com genomas grandes, complexos e altamente redundantes, como a cana-de-açúcar. A estratégia básica para a obtenção de EST constitui um método rápido e eficiente de amostragem do genoma para seqüências ativas de genes. O EST ou etiqueta de seqüência expressa representa uma seqüência parcial do cDNA de um gene que foi expresso em um tecido, num determinado momento (Sterky e Lundeberg 2000). Estudos da distribuição dos SSRs nos ESTs (EST-SSRs) foram realizados tanto em genomas eucariotos (Cardle *et al*. 2000; Kantety *et al*. 2002) quanto em procariotos (Gur-Arie *et al*. 2000).

Existem inúmeras vantagens na utilização de sequências expressas como marcadores genéticos, em relação a seqüências anônimas de DNA: (1) se o marcador EST estiver geneticamente associado a uma característica de interesse, é provável que o mesmo afete diretamente a característica (Cato *et al*. 2001; Chen *et al*. 2001; Thiel *et al*. 2003). Desta forma os marcadores derivados dos ESTs oferecem oportunidades para descoberta de genes; (2) os marcadores EST são provavelmente mais altamente conservados, logo, mais transferíveis entre espécies que sequências derivadas de marcadores anônimos (Cordeiro *et al*. 2001; Taylor *et al*. 2001; Decroocq *et al*. 2003). Como os marcadores derivados de ESTs encontram-se em regiões transcritas do genoma, os sítios de *primers* são mais conservados, tornando-os mais transferíveis entre espécies e aumentando seu valor em programas de melhoramento (Fraser *et al*. 2004); (3) os ESTs que apresentam homologia com genes candidatos podem ser utilizados de forma específica para o mapeamento genético e são úteis no alinhamento de genomas entre espécies relacionadas para análise comparativa (Holton *et al*. 2002). Além disso, os microssatélites derivados de EST (EST-SSR) ocorrem em alta frequência no genoma (Kantety *et al*. 2002), sendo a frequência estimada maior nas sequências codificadoras que nas regiões não codificadoras (Temnykh *et al*. 2001; Morgante *et al*. 2002),

No entanto, como marcadores genéticos, os EST-SSRs têm sido avaliados em vários estudos e tendem a ser consideravelmente menos polimórficos que os marcadores gerados a partir de sequências genômicas para arroz (Cho *et al*. 2000), cana-de-açúcar (Cordeiro *et al*. 2001) , trigo (Eujayl *et al*. 2002), cevada (Thiel *et al*. 2003).

Uma vez que os EST-SSRs amostram diretamente a variação nas regiões transcritas do genoma, podem prover estimativa da diversidade funcional. A instabilidade no número de repetições de SSRs implicaram na herança e severidade de algumas doenças humanas

(Brook *et al*. 1992; Vincent *et al*. 2000). A variação similar no número de repetições em plantas pode indicar variação em características vegetativas e de frutos, aumentando seus valores na seleção assistida por marcadores (Frase *et al*. 2004).

A ocorrência de SSRs em sequências EST tem sido reportada em diversas espécies de planta, incluindo arroz (Sasaki *et al.* 1994; Yamamoto e Sasaki 1997; Cho *et al.* 2000), uva (Scott *et al*. 2000), cevada (Holton *et al*. 2002; Thiel *et al*. 2003), milho (Wang *et al.* 1998; Sharapova *et al*. 2002), trigo (Eujayl *et al*. 2002; Kantety *et al*. 2002; Gupta *et al*. 2003; Gao *et al*. 2004; Nicot *et al*. 2004; Yu *et al*. 2004), algodão (Saha *et al*. 2003; Qureshi *et al*. 2004; Han *et al*. 2006) e laranja (Chen *et al*. 2006). A geração dos EST-SSRs tornou-se um atrativo complemento para as coleções existentes de SSRs.

Em cana-de-açúcar, a análise de 8.678 seqüências ESTs revelou aproximadamente 250 SSRs, a maioria composta por repetições perfeitas de trinucleotídeos sendo os motivos (CCG)n, (CGT)n e (CCT)n os mais comuns (Cordeiro *et al.*, 2001). Todos os EST-SSRs selecionados foram polimórficos nos gêneros correlatos *Sorgo* e *Erianthus*. O menor valor, para o conteúdo de informação de polimorfismo (PIC), foi obtido entre as variedades de cana-de-açúcar (0,23), aumentando entre as espécies *S. officinarum* e *S. spontaneum* (0,62) e, alcançando o maior valor (0,80) entre os gêneros *Sorgo* e *Erianthus*. Em virtude da base genética estreita das variedades de cana-de-açúcar, a utilização de EST-SSR pode auxiliar na caracterização da variabilidade genética disponível nas coleções de germoplasmas de gêneros correlatos utilizados em programas de introgressão. Desta forma, a limitação da introgressão do genoma *Erianthus* em cana-de-açúcar (*Saccharum*) pode ser superada pela utilização de EST-SSR na identificação da porção do genoma *Erianthus* nos híbridos intergenéricos (Cordeiro *et al*. 2001).

A aplicação dos ESTs mostrou-se um meio eficiente e bem sucedido de identificação de genes em cana-de-açúcar. Um estudo realizado por Carson e Botha (2000) revelou que de todos os clones de cDNA da bainha foliar, identificados na busca de homologia, 38% apresentaram similaridade significativa com seqüências de genes conhecidos. Este valor pode ser comparado ao observado na análise de bibliotecas de cDNA do endosperma e semente de milho (39.3%, Shen *et al.* 1994), sendo ainda melhor que os resultados obtidos utilizando bibliotecas de cDNA das folhas de milho (20%, Keith *et al.* 1993), tecidos de diferentes estágios de crescimento em arroz (25%, Yamamoto e Sasaki 1997) e porções de RNA de sementes estioladas, raiz, folhas e inflorescências de *Arabidopsis* (32%, Newman *et al.* 1994).

Várias razões têm sido apresentadas para os diversos resultados obtidos entre os diferentes projetos EST. Por exemplo, Van de Loo e colaboradores (1995) indicaram valores de identificação mais altos quando o tecido utilizado na construção da biblioteca de cDNA era especializado em processos envolvendo classes de proteínas bem caracterizadas. Em adição, mostrou-se que o sequenciamento da extremidade 5', em vez da extremidade 3', é mais informativo, e conseqüentemente a utilização de bibliotecas de clones direcionados fornecerá resultados mais significantes (Shen *et al.* 1994).

Projetos visando o sequenciamento de seqüências expressas têm sido desenvolvidos para diferentes espécies, com o intuito de fornecer informações básicas para o melhor entendimento da biologia das espécies e aprimoramento de programas de melhoramento, identificando uma grande coleção de novos genes candidatos. O número de sequências expressas depositadas no GenBank para trigo, milho, arroz e soja corresponde, respectivamente, a 416.000, 197.000, 113.000 e 308.000 sequências. Para cana-de-açúcar, projetos foram iniciados na África do Sul (Carson e Botha 2000), na Austrália (Casu *et al*.

2001) e no Brasil (http://sucest.lad.ic.unicamp.br/en). Na Austrália, o projeto está focado principalmente no mapeamento genético e na aplicação de marcadores genéticos no melhoramento genético da cana-de-açúcar, enquanto que a África do Sul está conduzindo um projeto pequeno (<500 ESTs). Uma busca mais recente de ESTs derivados do meristema apical, das folhas e colmos de cana-de-açúcar após a indução de floração (Ma *et al*. 2004) representam uma fonte adicional de ESTs, resultando numa combinação mundial total de 255.135 ESTs derivados de cana-de-açúcar (Casu *et al*. 2005).

As informações geradas nestes projetos têm sido utilizadas no mapeamento comparativo da família das Gramíneas, empregando-se marcadores comuns que hibridizam com cana-de-açúcar, arroz, milho, trigo hexaplóide, cevada e sorgo. No entanto, a informação molecular desenvolvida até o momento para a cana-de-açúcar é mínima comparada com a informação necessária para identificação e caracterização de locos que codificam características de importância agronômica.

O SUCEST (*Brazilian Sugarcane EST Project*), do programa Genoma da FAPESP, corresponde ao mais completo banco de dados de ESTs de cana-de-açúcar. Cerca de 200 pesquisadores de diversas universidades e centros de pesquisas do estado de São Paulo e outros lugares do Brasil contribuíram na construção deste banco de dados. Objetivando a obtenção de uma exaustiva visão do transcriptoma da cana-de-açúcar, 26 bibliotecas de cDNA foram construídas a partir de vários órgãos e tecidos de pelos menos 12 cultivares, em diferentes estágios de desenvolvimento. Destas bibliotecas, um total de 237.954 sequências ESTs foram obtidas, agrupadas em 43.141 prováveis transcritos únicos de cana-de-açúcar. Baseados na estimativa de redundância interna, concluiu-se que a coleção destes transcritos, possivelmente, indica cerca de 33.000 genes de cana-de-açúcar (Grivet *et al*.

2001; Grivet *et al*. 2003; Vettore *et al*. 2003). A anotação funcional destes produtos baseou-se na similaridade com sequências conhecidas do GenBank (Vettore *et al*. 2003).

Este banco representa uma ótima fonte de marcadores candidatos (Camargo 2000), que podem representar genes de importância econômica, em especial, aqueles relacionados ao metabolismo de sacarose, à resistência a pragas e doenças e à tolerância a condições adversas de clima e de solo (FAPESP 2001), podendo ser diretamente utilizados em programas de mapeamento genético. Uma vez mapeados estes marcadores podem ser avaliados para associação com as características de importância agronômica, bem como empregados na seleção assistida por marcadores, já que podem ser os genes responsáveis pela característica.

2.7 Mapeamento Genético

2.7.1 Princípios do mapeamento genético

A grande disponibilidade de marcadores moleculares, altamente polimórficos, aliada a procedimentos estatísticos complexos, tem permitido a construção de mapas genéticos para a maioria das espécies vegetais de interesse agronômico, até mesmo para aquelas de longo ciclo de vida, como as florestais e as frutíferas (Carneiro e Vieira 2002).

O mapeamento genético consiste em determinar a posição relativa de marcadores moleculares no genoma e visa igualmente determinar a distância genética entre dois marcadores consecutivos. O objetivo principal de um trabalho de mapeamento é ordenar o genoma, de maneira regular, visando à localização de genes de interesse e à identificação

de marcadores fortemente ligados a estes genes. Estando o genoma coberto, torna-se possível à localização precisa de um gene que codifica determinada característica.

Os principais passos para construção de um mapa genético são: (1) Produção de uma população segregante, sendo BC (*backcross*), F2, haplóides duplos, linhagens recombinantes (DH ou SSD) e cruzamentos entre plantas heterozigotas, os principais tipos de população utilizados. (2) Análise dos marcadores moleculares a fim de escolher os que são polimórficos entre os parentais da população de mapeamento. (3) Determinação dos alelos presentes em cada loco na progênie. (4) Construção dos grupos de ligação. (5) Ordenação dos marcadores dentro de cada grupo de ligação. (6) Computação das distâncias genéticas entre os marcadores. Os marcadores ideais para um trabalho de mapeamento genético apresentam como características, herança co-dominante, alto nível de polimorfismo, grande quantidade e sítio único no genoma com perfil e posição conhecidos.

Os marcadores empregados na construção dos mapas podem ser escolhidos com base em testes estatísticos da segregação dos alelos na população de mapeamento, comumente usando a estatística de Qui-quadrado (χ^2).

O nível de significância empregado na estatística Qui-quadrado é a probabilidade de se cometer um erro Tipo I, ou seja, a probabilidade de se rejeitar uma hipótese nula verdadeira. O nível de significância considerado para a realização dos testes é escolhido pelo pesquisador, sendo os valores mais comuns 5% e 1%. Isso implica na ocorrência de múltiplos testes, o que leva a um nível de significância conjunto ou genômico (α^*) do teste, o qual se apresenta crescente à medida que aumenta o número de testes realizados, como pode ser observado na equação abaixo (Liu 1998):

$\alpha^* = 1 - (1 - \alpha)^t$, em que:

$\alpha^* =$ nível de significância conjunto (ou genômico);

$\alpha =$ nível de significância para cada teste (individual);

$t =$ número de testes realizados (ou número de marcadores, no caso).

Assim, quando fixamos 5% para cada teste $(\alpha = 0{,}05)$ e realizamos, por exemplo, 100 testes $(t = 100)$, o nível de significância conjunto passa a ser $\alpha^* = 0{,}994079$, ou seja, certamente seriam detectados inúmeros falsos positivos.

Uma alternativa para contornar tal problema é a utilização da correção de Bonferroni (Rice 1989). Esta correção consiste em determinar o valor do nível de significância individual (α) que proporcionará o nível de significância conjunto (α^*) e, assim, contornar os problemas resultantes da realização de múltiplos testes. A fórmula empregada no cálculo da correção é apresentada a seguir, sendo os termos desta os mesmos já apresentados:

$$\alpha = -\exp\left[\frac{\ln\left(1-\alpha^*\right)}{t}\right] + 1$$

Tais análises estatísticas permitem a eliminação dos marcadores distorcidos do estudo de mapeamento, ou seja, descarte dos marcadores que não seguem as segregações esperadas na população de mapeamento. Assim, quando dispomos de vários marcadores em segregação em uma população, a primeira etapa consiste em testar a eventual ligação existente entre eles, seja através do teste de aderência de χ^2, seja através do *LOD score* (Mather 1957; Morton 1955), segundo uma análise de dois pontos. Esta etapa permite a

construção dos grupos de ligação. A ordem dos marcadores dentro de cada grupo é determinada em seguida através de uma análise de multiponto que se baseia geralmente no método de máxima verossimilhança. Estas análises e a construção final do mapa podem ser realizadas com o auxílio de programas especializados, como Mapmaker (Lander *et al* 1987) e JoinMap (Stam 1993).

Considerando a teoria da ligação genética, os mapas genéticos podem ser elaborados considerando-se a probabilidade de ocorrência de diferentes classes genotípicas, ou seja, utilizando a frequência de recombinação existente entre dois locos.

A frequência de recombinação traduz o número de *crossing-over* produzidos entre os dois marcadores durante a meiose. A relação entre a frequência de recombinação (distância genética) e o número de nucleotídeos separando dois marcadores (distância física) não é universal, podendo variar de um organismo para o outro. Pode igualmente variar de um lugar para o outro do genoma de um mesmo organismo, como encontrado, por exemplo, para o tomate (Ganal *et al* 1989), o arroz (Gustafson e Dillé 1992), o trigo (Chao *et al* 1989) e para o milho (Dooner *et al* 1986).

A frequência de recombinação é traduzida em distância genética, cuja unidade de medida é centiMorgan (cM). Esta conversão permite a análise dos eventos de recombinação entre os marcadores, além de dispor de uma medida aditiva da distância. Existem várias funções de mapeamento, que transformam a frequência de recombinação em distância genética e determinam de diferentes maneiras a ocorrência real de *crossing-over* em função da frequência de recombinação observada, levando ou não em consideração a interferência (interação entre *crossing-over*). As mais utilizadas são as funções de Haldane (1919), que assume não haver interferência, e de Kosambi (1944), que assume haver interferência entre os *crossing-over* adjacentes.

Em um mapa genético os cromossomos são representados por grupos de ligação, sob os quais são dispostos os marcadores moleculares, cujas distâncias são expressas em frequência de recombinação. Em poliplóides, os grupos de marcadores correspondentes aos cromossomos são denominados grupos de co-segregação, enquanto que o termo grupo de homologia designa o conjunto de grupos de co-segregação de uma classe de homologia.

O número de marcadores necessários para construção de um mapa genético depende do tamanho do genoma, do número de cromossomos e da frequência de recombinação genética. Um mapa pode ser considerado completo quando o número de grupos de ligação obtido pela análise dos marcadores for igual ao número de cromossomos do organismo e, também, quando todos os marcadores mapeados estiverem ligados, indicando que todas as regiões do genoma estão representadas (Guimarães e Moreira 1999a).

2.7.2 Mapeamento genético em cana-de-açúcar

Em alto poliplóides, como a cana-de-açúcar, o mapeamento genético é muito mais complexo que em espécies diplóides. A construção de mapas genéticos, em poliplóides, é viabilizada pela análise de segregação de marcadores em dose simples (MDS), proposta por Wu *et al.* (1992). Neste método, a segregação de cada marcador é analisada na progênie com base em sua presença e ausência. O marcador representado por dose simples estará presente em um genitor e ausente no outro, segregando na progênie na proporção de 1:1. Quando a poliploidia é alta e o pareamento é polisômico ou irregular, os alelos que estão presentes como simples cópia são muito mais informativos para a construção de mapas

genéticos que os demais. Logo, uma população ideal de mapeamento irá apresentar um grande número de MDS.

A relação entre marcadores e alelos, para poliplóides, foi claramente exemplificada por Burnquist (1991) em cana-de-açúcar. Em seu exemplo hipotético do padrão de bandas de RFLP obtido entre o cruzamento de dois genitores poliplóides, a presença de um alelo em cópia única em um genitor e ausente no outro, fornece na progênie uma segregação na proporção de 1:1. A presença de um alelo em cópia única, em ambos genitores, proporciona uma segregação na progênie de 3:1.

Dado o objetivo de compreender a genética das espécies estudadas de cana-de-açúcar, afim de melhor gerar os programas de melhoramento, diversos estudos de mapeamento foram desenvolvidos. Há alguns anos, várias equipes vêm tentando estabelecer o mapa genético da cana-de-açúcar, apesar da sua complexidade ligada à alta poliploidia. Com o auxílio de marcadores moleculares, mapas genéticos parciais foram produzidos para *S. spontaneum* (Al-Janabi *et al* 1993; Da Silva *et al* 1993a; Ming *et al* 1998, 2002a), *S. officinarum* (Al-Janabi *et al* 1993; Mudge *et al* 1996; Guimarães *et al* 1999b; Ming *et al* 1998) e para cultivares modernos (D'Hont *et al* 1994; Grivet *et al* 1996; Hoarau *et al* 2001; Aitken *et al*. 2005; Reffay *et al.* 2005; Raboin *et al.*2006; Garcia *et al*. 2006).

O primeiro mapa desenvolvido no gênero *Saccharum* explorou o genoma de um clone duplo haplóide de *S. spontaneum* (SES208, 2n = 64) (Da Silva *et al* 1995). Este mapa apresentou 424 marcadores MDS combinando marcadores RFLP (Da Silva *et al* 1993a) e marcadores RAPD (Al-Janabi *et al* 1993). O mapa genético apresentado por Al-Janabi e colaboradores (1993) foi o primeiro a ser construído diretamente a partir da análise do genoma de uma espécie poliplóide complexa, uma vez que até aquele momento os mapas

genéticos de espécies poliplóides eram construídos através da análise de cruzamentos envolvendo espécies diplóides relacionadas.

Mapeamento com clones cultivados foi iniciado com a progênie proveniente de autofecundações do cultivar SP70-1006 (D'Hont *et al* 1994) e mais tarde desenvolvido com o cultivar R570 (Grivet *et al* 1996) usando sondas de milho e de cana-de-açúcar, permitindo o alocamento de 408 marcadores em 96 grupos de co-segregação, cobrindo 2008cM do genoma. Este primeiro mapa para o cultivar R570 foi posteriormente completado com 600 marcadores RFLP derivados de diferentes sondas de Poaceae (Gramíneas) (Dufour *et al* 1997). Além disto, foi desenvolvido, para este mesmo cultivar, um mapa composto por 939 marcadores AFLP dos quais 887 estavam distribuídos em 120 grupos de co-segregação (Hoarau *et al* 2001). Este último mapa cobre 5849cM, o que representa um terço do comprimento total do genoma da cana-de-açúcar.

Rossi e colaboradores (2003) identificaram 88 genes análogos de resistência (RGAs) no banco de dados do SUCEST, com base em sua homologia típica para genes de resistência, incluindo 3 dos maiores grupos de genes de resistência com domínios NBS/LRR, LRR e S/T Kinase. Cinqüenta RGAs foram utilizados como sondas gerando um total de 148 marcadores do tipo RFLP em dose única, os quais foram integrados ao mapa genético do cultivar R570 (Hoarau *et al*. 2001). Até o momento, este é o mapa mais completo de um cultivar de cana-de-açúcar, apresentando 128 grupos de co-segregação, reunidos em 7 grupos de homologia e com comprimento de 7800cM. Porém vale acrescentar que este mapa não se encontra saturado, se compararmos com o tamanho total do genoma deste cultivar, estimado em 17000cM.

Cruzamentos interespecíficos envolvendo *S. officinarum* e *S. spontaneum* também foram explorados, gerando dois mapas de *S. officinarum* (Green German, 2n=97-117;

Muntok Java, 2n=140) e dois mapas de *S. spontaneum* (IND81-146, 2n=52-56; PIN84-1, 2n=96) com 418, 355, 385 e 297 marcadores RFLPs em simples dose, respectivamente. Duzentos e setenta, 206, 248 e 182 destes marcadores foram alocados em 72 grupos de co-segregação para ambos *S. officinarum* e 69 grupos de co-segregação para ambos *S. spontaneum* e cobriram, respectivamente, 2304, 1443, 2063 e 1103cM (Ming *et al* 1998).

Marcadores dos tipos RAPD, RFLP e AFLP também foram utilizados na construção de um mapa genético entre *S. officinarum* 'LA Purple' e *S. robustum* 'Mol 5829'. Os 642 marcadores moleculares, em dose simples, foram mapeados em ambas espécies gerando 74 grupos de co-segregação em 'LA Purple' e 65 em 'Mol 5829' (Guimarães 1999).

No último ano, três novos mapas genéticos foram construídos para cana-de-açúcar. Com base em uma progênie F1, derivada do cruzamento entre a variedade Australiana Q165 e o clone IJ76-514 (*S. officinarum*), 915 MDS foram mapeadas em 116 grupos de co-segregação, resultando em um mapa de 9058.3cM, contendo 8 grupos de homologia (Aitken *et al*. 2005). Este mapa é 1213cM mais longo que o mapa do cultivar R570, no entanto, apresenta menos marcadores, sendo o mapa do cultivar R570 ainda o mais completo. Uma comparação feita entre os dois mapas sugere que mapas genéticos de diferentes cultivares revelarão diferentes arranjos cromossômicos.

Reffay e colaboradores (2005) buscando um melhor entendimento da contribuição genética do clone Mandalay (*S. spontaneum*) nas variedades Australianas e nos parentais elites das populações de mapeamento, desenvolveram um mapa genético, a partir de 400 MDS. Estes foram agrupados em 101 grupos de co-segregação, originando um mapa com 3582cM de comprimento, apenas 20% saturado. Embora longe de estar saturado, este mapa possibilitou a identificação de regiões genômicas específicas de Mandalay em clones

parentais, de marcadores específicos de Mandalay e de QTL's com efeitos positivos e negativos em características importantes ao melhoramento.

O mais recente mapa genético de cana-de-açúcar foi desenvolvido para uma população derivada do cruzamento entre dois cultivares pré-comerciais, sendo o primeiro mapa baseado neste tipo de população (Garcia *et al*. 2006). Tal mapa possui 357 MDS, dos tipos AFLP, SSR e RFLP, alocados em 131 grupos de co-segregação, com comprimento total de 2602.4cM. Este mapa foi construído a partir de uma nova metodologia que possibilita a integração dos marcadores de ambos parentais em um único mapa, diferindo de todos os trabalhos publicados até o momento.

Desenvolver um mapa genético saturado de cana-de-açúcar requer muito mais trabalho que para diplóide. Para um determinado nível de diversidade molecular, o esforço para distinguir simultaneamente dez ou mais haplóides é muito maior que o necessário para distinguir apenas dois. Até o momento nenhum dos mapas genéticos publicados para cana-de-açúcar foi saturado. Estima-se que não mais de um terço do genoma está marcado, mesmo no mais refinado mapa (Rossi *et al* 2003). Para os mapas dos cultivares atuais, a cobertura dos marcadores é desigual, com os cromossomos de *S. spontaneum* sendo mais densamente cobertos que os cromossomos de *S. officinarum*. Novos modelos teóricos têm sido propostos a fim de obter-se maior precisão e compreensão das informações originárias dos experimentos em poliplóides complexos (Qu *et al* 2001). Estes esforços devem ser encorajados em vista da grande dificuldade de gerarmos dados completos (Grivet e Arruda 2001).

2.7.2.1 Mapeamento comparativo em cana-de-açúcar

A história evolutiva das gramíneas está bem estabelecida (Kellogg 2001). A conservação da organização cromossômica nesta família é de grande importância, visto que qualquer conhecimento relativo à determinada característica em uma espécie pode ser transferido para outra espécie de gramínea através do mapeamento comparativo (Gale e Devos 1998). O uso de conjuntos de marcadores genéticos comuns no mapeamento do genoma das gramíneas revelou que as espécies desta família possuem grande número de genes conservados e exibem extensas regiões de colinearidade nos mapas, implicando na conservação da ordem das seqüências de DNA nos cromossomos (Dufour *et al*. 1997; Glaszmann *et al*. 1997).

Devido à alta poliploidia da cana-de-açúcar e ausência de parentes diplóides próximos, a vantagem de investigar sintenia com outras Poaceae e, em particular, outros membros da tribo Andropogoneae tem sido explorada. O conhecimento do genoma da cana-de-açúcar potencialmente também beneficia programas do estudo de genomas de outros cereais, especialmente arroz, sorgo e milho. A sintenia da cana-de-açúcar com outras gramíneas tem sido utilizada com sucesso, por exemplo, para melhorar a resolução do mapeamento fino na região do gene de resistência à ferrugem (Asnaghi *et al*. 2000).

Sorgo parece ser um excelente modelo diplóide para a cana-de-açúcar, de maior sintenia, devido à divergência limitada e colinearidade entre os cromossomos da cana-de-açúcar e sorgo (Grivet *et al*. 1994; Glaszmann *et al*. 1997; Ming *et al*. 1998; Dufour *et al*. 1996; Dufour *et al*. 1997; Guimarães *et al*. 1997), além de ser a espécie mais adaptada para auxiliar no mapeamento genético da cana-de-açúcar (Dufour *et al*. 1997; Asnaghi *et al*. 2000).

No entanto, as primeiras comparações foram feitas com milho que tinha na época um mapa genético mais avançado. Sintenia foi dita conservada embora pouco perturbada pela descoberta da estrutura duplicada do genoma do milho e da presença de muitos rearranjos (D'Hont *et al*. 1994; Grivet *et al*. 1994; Dufour *et al*. 1996). A relação entre os genomas de milho e de cana-de-açúcar é a mais complexa: mais locos em cana-de-açúcar são ortólogos a dois locos de milho (Dufour *et al*. 1996; Grivet *et al*. 1994). Isto é provavelmente relacionado a alopoliploidia segmental, seguida de diploidização, e o aumento do tamanho do genoma causado por transposição ativa que afetam a linhagem de milho após sua divergência de sorgo e da cana-de-açúcar (Gaut *et al*. 2000).

Arroz também apresentou uma relativa sintenia global com a cana-de-açúcar, no entanto, com muitos rearranjos, explicados pela larga distância entre as duas espécies (Glaszmann *et al*. 1997). Embora distantemente relacionado, o arroz permanece um modelo interessante para cana-de-açúcar, dada a existência de um programa de sequenciamento do seu genoma e devido ao grande número de mutantes coletados. Assim, a colinearidade entre arroz e cana-de-açúcar é mais explorada que entre cana-de-açúcar e sorgo.

2.7.2.2 Mapeamento de QTL's em cana-de-açúcar

Uma importante aplicação dos mapas genéticos é a localização no genoma dos locos que contribuem para a variação de características fenotípicas. As espécies poliplóides apresentam também um maior desafio ao estudo destes locos e mapeamento dos QTL's. A primeira dificuldade é o extenso esforço requerido para a construção de um mapa genético que cubra todos os cromossomos homólogos. A segunda dificuldade é a coexistência de

muitos alelos segregantes em um único loco, o que gera a necessidade de ensaios em campo de grandes populações, com grande número de repetições, para a detecção de QTA's (*Quantitative Trait Alleles*).

Em cana-de-açúcar, potencialmente mais de 12 alelos coexistem em cada loco. De fato, em um determinado loco, o efeito de um alelo deve ser percebido apenas se exceder o efeito médio dos demais alelos segregantes no loco, e não somente se seu efeito exceder o efeito de um simples alelo alternativo, como ocorre nos organismos diplóides (D'Hont e Glaszmann 2001; Hoarau *et al.* 2002). Esta característica de complexidade genômica dita a utilização de parâmetros particulares para o mapeamento genético nesta cultura e os efeitos dos QTL's são freqüentemente pequenos devido a um complexo padrão de interações epistáticas e dominantes, nos locos e entre eles (Grivet e Arruda, 2001). Porém, apesar dos pequenos efeitos, em comparação aos diplóides, um QTL detectado em cana-de-açúcar, tende a ser mais robusto e diretamente útil, uma vez que foi diretamente confrontado com vários alelos presentes no loco (D'Hont e Glaszmann 2001).

Apesar das dificuldades, em cana-de-açúcar, locos que são associados a características importantes, como componentes de produção de açúcar e resistência a doenças, foram mapeados. Daugrois e colaboradores (1996) estudaram o padrão de herança da resistência à ferrugem, causada pelo fungo *Puccinia melanocephala*, em uma progênie proveniente de autofecundação do cultivar moderno R570 e detectaram um QTL de efeito maior ligado a 10cM de uma sonda de RFLP, além de outros QTL's de efeitos pequenos. Este foi o primeiro e, até o momento, o único gene maior localizado no genoma da cana-de-açúcar. Este gene chamado de *Bru1* é atualmente o foco de um projeto de clonagem baseada em mapa (D'Hont *et al.* 2001; Asnaghi *et al.* 2004).

Além deste estudo inicial, outros foram conduzidos nesta população de mapeamento oriunda da autofecundação do cultivar R570, com o intuito de avaliar componentes de produção da planta, como diâmetro do colmo, altura, número de colmos, conteúdo de açúcar (Hoarau *et al*. 2001; 2002). Numerosos QTA's com efeitos individuais de 3-7% foram detectados. Estes pequenos efeitos eram esperados devido a poliploidia presente. Dada a predominante polissomia, o efeito de um alelo é derivado da comparação entre sua presença e sua ausência independente dos demais alelos, que, provavelmente, têm efeito na mesma característica (D'Hont e Glaszmann 2001). Esses efeitos são especialmente importantes em cultivares modernos, uma vez que outros estudos baseados em populações interespecíficas entre *S. officinarum* e *S. robustum* (Sills *et al.* 1995) e *S. officinarum* e *S. spontaneum* (Paterson, dados não publicados) detectaram QTA's de efeitos mais fortes. Alelos favoráveis provavelmente concentraram-se nos cultivares modernos por melhoramento recorrente, assim diminuindo as contradições internas que determinam a segregação da característica e a magnitude dos efeitos dos QTA's (D'Hont e Glaszmann 2001).

Nesta mesma população, Jordan e colaboradores (2004) detectaram pequenos QTL's para o número de colmos e observou também que a mesma sonda de RFLP, quando utilizada em sorgo, também estava associada ao crescimento do colmo.

Além destes estudos com a população da R570, outros foram desenvolvidos. Guimarães (1999) utilizou metodologias de regressão e mapeamento por intervalo para detectar e mapear regiões genômicas associadas a QTL's referentes a número, diâmetro e peso do colmo, porcentagem de colmos com inflorescência, infectados com carvão (*Ustilago scitaminea*), porcentagem de fibra e teor de sacarose, em uma progênie proveniente do cruzamento entre *S. officinarum x S. robustum*. Foram detectadas

associações significativas para todos os caracteres avaliados. O grupo de ligação VII ancorou QTL's para três dos sete caracteres fenotípicos avaliados, dentre eles o florescimento. Os resultados demonstraram que mapas genéticos baseados em marcadores em dose simples podem ser empregados para identificar regiões genômicas que controlam caracteres quantitativos em poliplóides, assim como em diplóides. Em um estudo anterior, este mesmo grupo encontrou um marcador RFLP associado ao de dia curto (Guimarães *et al*. 1997).

O mapeamento de QTL's em autopoliplóides é complicado pela possibilidade de segregação de vários alelos em um mesmo loco, como já dito anteriormente, e também pela falta de pareamento preferencial. No entanto, o subconjunto de alelos polimórficos que apresentam taxa de segregação simples podem ser utilizados na localização de QTL's. Partindo destes princípios, Ming e colaboradores (2001) utilizando duas populações interespecíficas (*S. officinarum x S. spontaneum*) realizaram um estudo da base genética da variação no conteúdo de açúcar entre genótipos de cana-de-açúcar, a partir de marcadores em dose única. O objetivo era detectar o número e a localização dos QTL's para tal característica, com o intuito de investigar as bases moleculares dos autopoliplóides. Alguns QTL's foram detectados para os genótipos que apresentavam um maior teor de sacarose, sugerindo que alelos favoráveis foram fixados por seleção. No entanto, muitos dos QTL's não correspondiam a genes candidatos conhecidos, sugerindo que outros ensaios seriam necessários para isolar os determinantes genéticos do alto teor de sacarose.

Esta mesma população foi empregada no mapeamento de QTL's que afetam a altura da planta e florescimento, em um experimento conduzido em larga escala (Ming *et al.* 2002b). Os QTL's foram detectados em 8 regiões genômicas, sendo que 4 dos QTL's que controlam a altura da planta correspondem a 4 dos 6 QTL's previamente mapeados em

sorgo. Um QTL que controlava o florescimento em cana-de-açúcar correspondia a 1 dos 3 QTL's também mapeados em sorgo. A correspondência na localização de tais QTL's nestas duas culturas reforça a noção de que o simples genoma de sorgo é um molde válido para buscar o entendimento molecular do genoma da cana-de-açúcar.

Em suma, todos os estudos apresentados de mapeamento de QTL's em cana-de-açúcar, além de outros mais recentes (Reffay *et al* 2005; Raboin *et al* 2006) identificaram muitos alelos segregantes independentes, que puderam ser atribuídos a locos distintos. Baseando-se na comparação com mapas de outras gramíneas, particularmente milho, alguns destes alelos foram dados como candidatos. Vários QTL's foram detectados e puderam ser localizados em algumas regiões cromossômicas, sugerindo que o número atual de genes envolvidos no controle genético de tais características deve ser bem menor que o esperado. Em um cruzamento interespecífico a ampla segregação de fenótipo pode proporcionar um contexto favorável à detecção de QTL's.

Diante de todos as informações apresentadas anteriormente, observa-se que ainda há muito que estudar para se entender as heranças complexas em cana-de-açúcar. Sabe-se que o pagamento da cana-de-açúcar é feito sob um critério de qualidade que considera o teor de sacarose e a pureza do caldo. Este sistema incentiva os fornecedores a cultivarem variedades de cana-de-açúcar de melhor qualidade, livre de doenças e com alto teor de sacarose. Dada a importância econômica da cultura, e os problemas a ela associados, torna-se evidente os avanços que podem ser alcançados no melhoramento genético da cana-de-açúcar com o desenvolvimento de marcadores genéticos e a utilização destes em estudos de mapeamento genético e de QTL's.

3 Objetivo

O presente trabalho teve como objetivo central desenvolver marcadores moleculares microssatélites a partir de buscas no banco de dados de ESTs de cana-de-açúcar (SUCEST), bem como utilizá-los no mapeamento funcional de uma população segregante derivada do cruzamento entre duas variedades pré-comerciais de cana-de-açúcar contrastantes para características de importância econômica.

4 Artigo I

"Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats"

Luciana Rossini Pinto, Karine Miranda Oliveira, Eugênio César Ulian, Antônio Augusto Franco Garcia e Anete Pereira de Souza

Publicado na revista

Genome (47: 795-804; 2004)

Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats

Luciana Rossini Pinto, Karine Miranda Oliveira, Eugênio César Ulian, Antonio Augusto Franco Garcia, and Anete Pereira de Souza

Abstract: Sugarcane microsatellites or simple sequence repeats (SSR) were developed in an economical and practical way by mining EST databases. A survey in the SUCEST (sugarcane EST) database revealed a total of 2005 clusters out of 43 141 containing SSRs. Of these, 8.2% were dinucleotide, 30.5% were trinucleotide, and 61.3% were tetranucleotide repeats. Except for dinucleotides, the CG-rich motif types were the most common. Differences in abundance of trinucleotide motif types were observed between EST-SSRs and those isolated from sugarcane genomic libraries. Among the different cDNA libraries used for EST sequencing, SSRs were more frequent in the ones derived from leaf roll (LR). Twenty-three out of 30 tested SSRs produced scorable polymorphisms in 18 sugarcane commercial clones. These EST-SSRs showed a moderate level of polymorphism with some SSRs producing unique fingerprints. The number of alleles observed among the 18 clones evaluated varied from 2 to 15, with an average of 6.04 alleles/locus. The polymorphism information content (PIC) values ranged from 0.28 to 0.90 with a mean of 0.66. The EST-SSRs screened over both parents (SP 80-180; SP 80-4966) and 6 F_1 individuals produced 52 segregating markers that could potentially be used for sugarcane mapping. The EST-SSRs were found in clusters that had significant homology to proteins involved in important metabolic pathways such as sugar biosynthesis, proving that EST-SSRs are a valuable tool for the construction of a functional sugarcane map.

Key words: sugarcane, polyploid, expressed sequence tags (ESTs), microsatellites (SSRs), genetic mapping.

Résumé : Des microsatellites (SSR) ont été développés aisément et de façon économique chez la canne à sucre en exploitant des banques d'EST (étiquettes de séquences exprimées). Une interrogation de la banque SUCEST (« sugarcane EST database ») a révélé que 2 005 des 43 141 groupes d'EST contenaient des microsatellites. De ceux-ci, 8,2 % étaient dinucléotidiques, 30,5 % étaient trinucléotidiques et 61,3 % étaient tétranucléotidiques. À l'exception des microsatellites dinucléotidiques, les motifs riches en GC étaient les plus fréquents. Des différences quant à l'abondance des motifs trinucléotidiques ont été observées entre les microsatellites provenant d'EST et ceux isolés chez l'ADN génomique. Parmi les différentes banques d'ADNc employées pour le séquençage d'EST, les microsatellites étaient les plus fréquents parmi les transcrits de la gaine foliaire (« leaf roll »). Vingt-trois des 30 microsatellites analysés ont produit un polymorphisme détectable chez 18 clones commerciaux de canne à sucre. Ces microsatellites montraient un degré modéré de polymorphisme et certains microsatellites produisaient une empreinte unique. Le nombre d'allèles observés parmi les 18 clones étudiés variait entre 2 et 15 pour une moyenne de 6,04 allèles par locus. Les valeurs de l'indice PIC (« ploymorphism information content ») variait entre 0,28 et 0,90 pour une moyenne de 0,66. Les microsatellites provenant d'EST ont été examinés sur deux parents (SP 80-180 et SP 80-4966) ainsi que six individus F_1 et cette analyse a révélé 52 marqueurs en ségrégation et potentiellement utiles pour la cartographie génétique chez la canne à sucre. Les EST portant des microsatellites provenaient de gènes présentant une homologie significative avec des protéines impliquées dans des sentiers métaboliques importants, tels que la biosynthèse du sucre. Ces observations démontrent que les microsatellites provenant d'EST constituent un outil utile pour la construction d'une carte fonctionnelle de la canne à sucre.

Mots clés : canne à sucre, polyploïde, étiquettes de séquences exprimées (EST), microsatellites (SSR), cartographie génétique.

[Traduit par la Rédaction]

Received 15 September 2003. Accepted 11 May 2004. Published on the NRC Research Press Web site at http://genome.nrc.ca on 8 October 2004.

Corresponding Editor: P. Donini.

L.R. Pinto, K.M. Oliveira, and A.P. de Souza.[1] Centro de Biologia Molecular e Engenharia Genética (CBMEG) – Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz,CP 6010, CEP 13083-970, Campinas-SP, Brasil.
E.C. Ulian. Centro de Tecnologia COPERSUCAR, Piracicaba-SP, Brasil.
A.A.F. Garcia. Depto de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400,970, Piracicaba-SP, Brasil.

[1]Corresponding author (e-mail: anete@unicamp.br).

Genome 47: 795–804 (2004) doi: 10.1139/G04-055

Introduction

Sugarcane (*Saccharum* spp.) is a major economical crop that accounts for two thirds of the world's sugar production (Carson and Botha 2000). The improvement of sugar production, as well as the control of the main diseases affecting it, is achieved with the use of improved varieties.

Commercial varieties are derived from interspecific crosses between *Saccharum officinarum* ($2n$ = 80 chromosomes) and *Saccharum spontaneum* ($2n$ = 40–128 chromosomes), resulting in an asymmetric chromosome transmission that leads to the formation of varieties with different chromosome numbers, generally between 100 and 130 (Price 1965). Thus, the high polyploid and aneuploid nature of the interspecific hybrids allied with the difficulties of controlled hybridization (Hogarth 1987, Silva et al. 1993) and the time spent on developing new varieties (from 12 to 15 years) are the main drawbacks in sugarcane breeding programs.

Molecular genetic markers are a valuable tool in the studies of complex genomes such as sugarcane (Daugrois et al. 1996). Their incorporation in the selection of economic traits during the early stages of a breeding program, as well as in the choice of the best parents in a cross, may significantly reduce the time for the development of new varieties. These goals can be successfully achieved with the availability of robust polymorphic markers that cosegregate with economically important agronomic traits in sugarcane.

Microsatellites or simple sequence repeats (SSR) have been considered one of the most powerful Mendelian markers (Jarne and Lagoda 1996). This is due to their codominant nature, locus specificity, and high reproducibility. Basically, SSRs are stretches of DNA consisting of small motifs of 1 to 6 base pairs repeated in tandem and flanked by sequences that are sufficiently conserved, thereby allowing the design of specific primers for PCR amplifications (Powell et al. 1996; Wang et al. 1994). However, sequence information is required for SSR development.

Single-pass partial sequencing of the 5′ or 3′ end of a cDNA clone that represents an mRNA, known as expressed sequence tags (ESTs) (Wolfsberg and Landsman 2001), is a fast and efficient approach for sampling the transcribed portion of the genome (Sterky and Lundeberg 2000; Liang et al. 2000). At present, EST databases (dbEST) are an important source of candidate genes, as they can generate markers directly associated with a trait of interest (Cato et al. 2001). Moreover, the development of SSRs from dbESTs is much less expensive and laborious than that from genomic DNA libraries (Zane et al. 2002).

SUCEST (the sugarcane expressed sequence tag project) is the most complete dbEST for sugarcane in the world, with 237 954 EST sequences assembled in 43 141 clusters (Telles et al. 2001; Vettore et al. 2001). This number of ESTs is, approximately, 267 times greater than that currently available for *Saccharum* in the dbEST (http://ncbi.nlm.nih.gov/dbEST/) from NCBI (National Center for Biotechnology Information, Bethesda, Md.).

In sugarcane, a total of 250 SSRs were obtained from the analysis of 8 678 EST sequences. Despite the low levels of polymorphism detected in the sugarcane cultivars evaluated, they were highly transferable to related species or genera (Cordeiro et al. 2001). Similar results were observed in a preliminary electronic search in the SUCEST database, where 402 SSRs were identified and significant levels of polymorphism were achieved with a set of 20 primers screened over commercial cultivars and different *Saccharum* species (Silva 2001).

SSRs derived from ESTs can be associated with genes of known function and used as functional SSR markers, tagging genes of interest in a more efficient manner (Hackauf and Wehling 2002). This work describes a survey for microsatellites in the SUCEST database as part of an effort to generate molecular genetic markers that will be used in the construction of a functional map derived from two elite sugarcane clones selected from the COPERSUCAR (Piracicaba, São Paulo, Brazil) breeding program.

Materials and methods

SUCEST database

The Sugarcane EST database (http://sucest.lad.ic.unicamp.br/en/) comprises 237 954 expressed sequence tags assembled in 43 141 clusters. These ESTs, or reads, were obtained by single-pass sequencing of the 5′ and (or) 3′ ends of a cDNA clone. Twenty-six cDNA libraries were constructed from several organs and tissues sampled at different developmental stages. The reads were trimmed for low-quality sequences and grouped by a clustering procedure (CAP 3 program), giving rise to a consensus sequence (Telles et al. 2001; Vettore et al. 2001).

SSR identification

EST-SSRs were mined from the Sugarcane EST database. The BLASTn software was used to search for all possible repeat patterns with $n \geq 7$ for dinucleotide, $n \geq 5$ for trinucleotide, and $n \geq 3$ for tetranucleotide motifs against the Cluster Cap-3 consensus sequences. A total of 47 motif types were searched for in the SUCEST database. Each motif entry in the BLASTn represents all of the possible permutations and complementary repeats produced by the respective motif. To identify, count, and localize the SSR motif inside the cluster, the cluster FASTA sequence was submitted to a free, available, simple sequence repeat identification tool (http://www.gramene.org/microsat). Similarity scores over 80 were used as a threshold to identify the protein homology of EST clusters containing SSRs.

Primer design

Thirty primer pairs flanking the SSR motifs (Table 1) were designed from cluster consensus sequences using Primer Select software from the LaserGene program (versions 5.01 and 5.02, DNAStar, Inc., Madison, Wis.). The stringency criteria adopted was based on a GC content between 50% and 75%, melting temperature between 50 and 65 °C,and product length between 150 and 400 bp (Cordeiro et al. 2001). Primer pairs were synthesized by Invitrogen (Campinas, SP).

DNA extraction and PCR amplification

Total genomic DNA was extracted from 300 mg of powdery lyophilised young leaf tissues using a CTAB method (Hoisington et al. 1994) modified for sugarcane. PCRs were performed in a 20-μL final volume containing 40 ng of tem-

Table 1. Primer sequences for 30 EST-SSRs from the SUCEST database with annealing temperature (T_m), expected product size, and expected EST homology.

Primer[a]	Motif	Forward primer (5′→3′)	Reverse primer (5′→3′)	T_m (°C)[b]	Expected product size (bp)[c]	Expected EST homology
SCA01	$(AG)_{24}$	CGGCCCCGGGAAGAACT	GGCACAACCAAACCAACAGG	55	382	No hits found
SCA02	$(CG)_8$	TCCCCGCCTCGCAATCAGAC	CGCTTGCGGGAGGCTAAATC	59	244	(GenBank acc. No. AC025417) T12C24.22
SCA03	$(AC)_{20}$	CTAACCAGCAGCGTCACCAG	AGGCCGAGGTTTTGTCTTTG	53	181	(GenBank acc. No. AC006234) (1–4)-β-mannan endohydrolase
SCA04	$(TG)_9$	GGTACAAGCGGAACGGTGAC	ACACGGCAGTAGCACACGAG	54	170	(GenBank acc. No. AL391711) putative protein
SCA05	$(TA)_{145}$	ACCAAAGCAAGCAGGGAGAG	GACCAGTGACCCCCTCCTTA	53	415	No hits found
SCA06	$(TG)_{12}$	CATGCAGTCGTTGTCGTAGC	GCAAGCAAACCCAAGGACAT	54	238	(GenBank acc. No. U29176) lipid transfer protein precursor
SCA07	$(AC)_7$	TTGCAATGGAGGCGAAACAC	CAGGAGTATGAGCAACAGAGCAG	54	201	(GenBank acc. No. U36432) GNOM gene product
SCA08	$(CG)_7$	GGCAGGAGCGCGACAAGACG	GCCGGATCGCCAGGTAGAAGAAAC	63	217	(GenBank acc. No. AL031004) putative protein
SCA09	$(TC)_8$	GTCAGTCCGCGAACAAGGTAG	TCCAGCAACAGCAGACAACAC	54	176	No hits found
SCA10	$(GA)_7$	TCTAAGCCAAGCCGATTCCGTTCT	CAGCAGCCCAACCCACAGTCG	62	195	No hits found
SCB01	$(CGG)_7$	GCCGCCGTGATGGAGGTG	GGCCCCACCGCTCAGTTC	58	193	(GenBank acc. No. AF034945) glycine-rich RNA-binding protein
SCB02	$(AGA)_5$	GGCAGAAGGATGATGATGAGGAT	CTCCTCCTTGACCTCCACAG	51	164	(GenBank acc. No. AC082644) putative centromere / microtubule-binding protein
SCB03	$(CAG)_5$	CCCACTTCTCCCCTCCCGTTCC	GTCATCGGCGCCAGCACCAC	63	168	(GenBank acc. No. AJ278332) 12-oxophytodienoate reductase 3
SCB04	$(GGC)_5$	CAGTGTGCGTGCCAGTCC	CGATGTAGCGCTGGTAGAAC	50	228	(GenBank acc. No. AC021640) unknown protein
SCB05	$(CCA)_5$	CGCGCTGCATTAGAAAAG	AGCCATCTTTTAGGGGTAGC	50	143	(GenBank acc. No. AB054123) ribosomal protein S10
SCB06	$(GCC)_8$	CGATCTCCCTCCTCCTCCTCCAATG	TCCACTCCACACCTGACCTGACCAC	63	176	(GenBank acc. No. AC018363) putative protein phosphatase-2C (PP2C)
SCB07	$(CGA)_8$	ACGAGAACCACAGCCACCAG	GGAGGTAGTCGGTGAAGTGC	51	291	(GenBank acc. No. AB013392) pyrophosphate-dependent phosphofructo-1-kinase-like protein
SCB08	$(GGT)_6$	TCTCGTTATGATCGTGGTCGTG	CCACCACCATACCTGTCATCTCTG	57	176	(GenBank acc. No. AF315811) RNA-binding protein
SCB09	$(GAC)_6$	AGGCGGCAGGGGTTGTAGAGAT	CATCCTCAGCACCATCACCATT	56	251	No hits found
SCB10	$(CGC)_5$	GGCTCCACCGCACCAACTCC	CGCCTCGCTCGTCTTGGTCTC	61	196	No hits found
SCC01	$(GATA)_{14}$	GATGCTTGGGTCGTGATTTC	TCGCGTCCACCAATGAACC	56	298	(GenBank acc. No. Z97022) cysteine proteinase
SCC02	$(CGGC)_5$	GATCTTATGGGTGGGTCCGTTCG	GCCGCCGGTGGTCTCCTTCT	62	202	(GenBank acc. No. AF130975) plasma membrane intrinsic protein
SCC03	$(CCAC)_4$	ACTCCCTCCCCTCTTCCTC	ATGCTTTCCTTGGCTCCCGATTCC	64	227	(GenBank acc. No. AF241166) MAP kinase MAPK2
SCC04	$(GGAT)_6$	GGGGACCTGAAGATGACTGC	TCCTGCCTGCCTCATCATAC	52	218	(GenBank acc. No. AB019235) contains similarity to DNA-binding protein gene
SCC05	$(TGCT)_4$	TCGTCTTCCTCCTCTTGCTCTGGTG	CATCCTCCTCTGCTGCTCTCGTCTC	62	179	(GenBank acc. No. X75670) cytochrome *b5*

Table 1. Primer sequences for 30 EST-SSRs from the SUCEST database with annealing temperature (T_m), expected product size, and expected EST homology.

Primer[a]	Motif	Forward primer (5′→3′)	Reverse primer (5′→3′)	T_m (°C)[b]	Expected product size (bp)[c]	Expected EST homology
SCC06	$(CGTG)_5$	CCCTCCAGGCCGTTTTTG	ACAGAGCGGTAGCAGCAGTAGTAGT	55	201	(GenBank acc. No. X66014) cathepsin B
SCC07	$(CTAG)_3$	GGACGAGTGCCCCTGCTACA	AGCACGTCCGAGGATACAGTCATAC	58	228	(GenBank acc. No. AB015475) contains similarity to gibberellin-stimulated transcript 1 like protein–gene
SCC08	$(GGCC)_3$	TCGCCGGCCAGACCCAGAT	CCCCCACCCCATACCAACTCCAT	63	180	No hits found
SCC09	$(TCTT)_3$	GAATGCCCTCCAGTCAAGTG	ACCATTGACACGGAAACTGC	52	260	(GenBank acc. No. AL353814) protein kinase like protein
SCC10	$(AGGC)_3$	AGCTTCCTCGGCCATACCATTTC	GTGGAGCGGCATGGTGGAGAA	62	252	No hits found

[a]SCA, SCB, and SCC are, respectively, di, tri, and tetranucleotide primer pairs.
[b]Annealing temperature estimated by the Primer Select program.
[c]Expected product size (bp) estimated from the Primer Select program.

plate DNA, 0.2 μM of each forward and reverse primer, 100 μM of each dNTP, 2.0 mM $MgCl_2$, 10 mM Tris–HCl, 50 mM KCl, and 0.5 U *Taq* DNA polymerase (Invitrogen, São Paulo, Brazil). Reactions were performed on a PTC-100 thermocycler (MJ Research, Waltham, Mass.) using the following conditions: 94 °C for 3 min; 30 cycles of 94 °C for 1 min; annealing temperature specific to each primer for 1 min; extension at 72 °C for 1 min; a final elongation step at 72 °C for 2 min. The annealing temperature of each primer pair was identified using a gradient of temperature ranging from 50 to 65 °C on a PTC-200 thermocycler (MJ Research). Amplification products were separated by electrophoresis on 6% w/*v* denatured polyacrylamide gels using a 25-bp ladder as a size standard and silver stained according to Creste et al. (2002).

Polymorphism analysis

A total of 18 sugarcane commercial clones (CB 3642, SP 79-1011, CB 40-77, IAC 51-205, RB 739-359, SP 70-1284, IAC 64-257, SP 71-1406, SP 80-3280, RB 85-5035, SP 79-6134, SP 79-2313, RB 855563), including two parental clones (SP 80-180 and SP 80-4966) involved in a molecular genetic mapping program from COPERSUCAR and one clone from each of three *Saccharum* species represented by IJ76-314 (*S. officinarum*); Gandacheni (*S. barberi*) and Maneria (*S. sinense*) were used to assess the level of polymorphism of the EST microsatellites (EST-SSRs). To evaluate the level of polymorphism generated among the three types of repeats (di, tri, and tetranucleotides), 10 primers were designed for each repeat class (Table 1). The polymorphism information content (PIC) value was calculated by the following formula: $PIC = 1 - \sum p_{ij}^2$, where p_{ij} is the frequency of the j_{th} allele (marker) for the i_{th} EST-SSR locus. The capability of the EST-SSRs to distinguish between cultivars was estimated as the percentage of different banding patterns obtained among sugarcane commercial clones. To evaluate the potential of the EST-SSRs as markers for genetic mapping, both parents (SP 80-180 and SP 80-4966) and a random sample of 6 F_1 individuals derived from a mapping population were scored based on the presence (1) or absence (0) of SSR markers.

Results

Abundance of SSRs in the SUCEST database

According to the search criteria adopted, a total of 2005 clusters containing SSRs were found in the SUCEST database. Of these, 165 (8.2%) were composed of dinucleotide repeats, 613 (30.5%) of trinucleotide repeats, and 1227 (61.3%) of tetranucleotide repeats (Table 2). Among the dinucleotide motifs, $(CG)_n/(CG)_n$ was the least frequent (12%), with a 95% statistically significant confidence interval (95% CI 12.21–30.89), based on direct counts. The remaining dinucleotide motifs were almost equally frequent (29%). The most common trinucleotide motifs were $(GGT)_n/(ACC)_n$ with 18.8%, $(CGG)_n/(CCG)_n$ with 16.3%, $(CTC)_n/(GAG)_n$ with 12.4%, and $(TCA)_n/(TGA)_n$ with 11.0%. For tetranucleotides, the motifs $(CACG)_n/(CGTG)_n$, with 8.3%, and $(AGGC)_n/(GCCT)_n$, $(ACAT)_n/(ATGT)_n$, and $(TCAT)_n/(ATGA)_n$, with approximately 5.0%, were the most abundant.

Table 2. Distributions of SSRs according to the motif type searched in the SUCEST database.

Motif types	Frequency (%)	Total	95% Confidence interval[a]
Dinucleotides			
$(AG)_n/(CT)_n$	29	48	35.39–63.64
$(AT)_n/(AT)_n$	29.5	49	36.25–64.78
$(CG)_n/(CG)_n$	12	20	12.21–30.89
$(AC)_n/(GT)_n$	29.5	48	35.39–63.64
Trinucleotides			
$(ACG)_n/(CGT)_n$	7.2	44	31.97–59.07
$(CGG)_n/(CCG)_n$	16.3	100	81.36–121.63
$(GGT)_n/(ACC)_n$	18.8	115	94.94–138.04
$(CAA)_n/(TTG)_n$	8.6	53	39.70–69.32
$(TTC)_n/(GAA)_n$	9.5	58	44.04–74.98
$(AGC)_n/(GCT)_n$	8.8	55	41.43–71.60
$(CTA)_n/(TAG)_n$	3.4	21	12.99–32.10
$(TTA)_n/(TAA)_n$	4	24	15.38–95.71
$(TCA)_n/(TGA)_n$	11	67	51.92–85.09
$(CTC)_n/(GAG)_n$	12.4	76	59.88–95.12
Tetranucleotides			
$(ACAA)_n/(TTGT)_n$	4	49	36.25–64.78
$(CACG)_n/(CGTC)_n$	8.3	103	84.07–124.92
$(CTGT)_n/(ACAG)_n$	1.9	23	14.58–34.51
$(AGGC)_n/(GCCT)_n$	5.4	66	51.04–83.97
$(CAAT)_n/(ATTG)_n$	2.8	34	23.55–47.51
$(GCAT)_n/(ATGC)_n$	3.8	46	33.68–61.35
$(CTAC)_n/(GTAG)_n$	0.6	8	3.45–15.76
$(AAAG)_n/(CTTT)_n$	3.6	44	31.97–59.06
$(CCTC)_n/(GAGG)_n$	2.5	31	21.06–44.00
$(GTAC)_n/(GTAC)_n$	1.4	17	9.90–27.22
$(GCTC)_n/(GAGC)_n$	3.8	47	34.53–62.50
$(CGAT)_n/(ATCG)_n$	3.3	41	29.42–55.62
$(CAAG)_n/(CTTG)_n$	3.3	40	28.57–54.47
$(GGAT)_n/(ATCC)_n$	3.4	42	30.27–56.77
$(CCAC)_n/(GTGG)_n$	3.7	46	33.68–61.35
$(GAAG)_n/(CTTC)_n$	3.7	45	32.82–60.21
$(CTAT)_n/(ATAG)_n$	4	49	36.25–64.78
$(TTAG)_n/(CTAA)_n$	0.6	7	2.81–14.42
$(CTAG)_n/(CTAG)_n$	2.8	35	24.38–48.67
$(ACAT)_n/(ATGT)_n$	5	61	46.66–78.36
$(GTTG)_n/(CAAC)_n$	3.2	39	27.73–53.31
$(CCCG)_n/(CGGG)_n$	3.8	47	34.53–62.50
$(AAAT)_n/(ATTT)_n$	4.5	55	41.43–71.59
$(TCAT)_n/(ATGA)_n$	4.7	50	37.11–65.92
$(GAAC)_n/(GTTC)_n$	1.5	19	11.43–29.67
$(CGAC)_n/(GTCG)_n$	1.7	21	12.99–32.10
$(TCAG)_n/(CTGA)_n$	1.9	24	15.38–35.71
$(TAAT)_n/(ATTA)_n$	3.1	38	26.89–42.16
$(CACT)_n/(AGTG)_n$	0.7	9	4.11–17.08
$(TTAC)_n/(GTAA)_n$	0.2	2	0.24–7.22
$(CCAG)_n/(CTGG)_n$	4	49	36.25–64.78
$(AGTG)_n/(CACT)_n$	0.8	10	4.79–18.39
$(GCCG)_n/(CGGC)_n$	2.5	30	20.24–42.83

[a]Confidence interval based on Poisson distribution for direct counts.

The majority of the SSRs showed a perfect repeat structure (75.1%) with tetranucleotide repeats having the higher percentage (Table 3). The compound repeats were more abundant for dinucleotides (10.3%) and less abundant for tri and tetranucleotide repeats, with 3.4% and 2.0%, respectively. Low levels of imperfect repeats, i.e., consecutive repeats interrupted by stretches of nonrepeat bases, were observed in all repeat classes. Approximately 19% of the clusters had more than one region containing SSRs.

Across all repeat classes, the mean value of the repeat number was higher for the dinucleotides (15.8). The length of this class of repeat varied from $n = 7$ to $n = 145$ (Table 4) with a high variation in the repeat length, as illustrated in the distributions of the repeat number (Fig. 1). Approximately 20% of the dinucleotide SSRs had a length between 3 and 7 repeats, 40% between 8 and 14 repeats, and 5% between 36 and 43 repeats. The majority of the trinucleotide SSRs (80%) had a length close to the minimum repeat number ($n \geq 5$), as they were located in the 3–7 interval. Tetranucleotides were the least variable, with almost 100% of the SSRs between 3 and 7 repeats. The mode value showed that most of the SSRs had a repeat number close to the cut-off value used in the database search.

In relation to the distribution of the SSRs in different cDNA libraries (Fig. 2), it was noted that, regardless of repeat class, they were more frequent in the cDNAs derived from the leaf roll (LR), root (RT), leaf–root transition zone (RZ), and sugarcane flower (FL) libraries and almost absent in cDNA libraries derived from normalized tissues (NR).

Functional identification of the EST clusters containing SSRs

Of all the clusters containing microsatellites, 48.7% showed protein homology with similarity scores above 80. These clusters had homology to a vast range of proteins including enzymes, RNA- and DNA-binding proteins, and defense-related proteins, as well as transcriptional and translational factors (Table 5). Some enzymes related to sugar metabolism, such as a putative sugar transporter protein (AF119222), soluble acid invertase (AF062735), sucrose synthase (AC012396), and sucrose-6-phosphate hydrolase (U16123), were also found. However, most of the clusters were associated with hypothetical proteins or proteins with unknown function, i.e., proteins as yet uncharacterized. Few discrepancies were observed for some clusters that showed significant similarity scores in relation to an unnamed protein product (AK001129) derived from *Homo sapiens*, a vib gene product (AE003725) from *Drosophila melanogaster*, and a multifunctional β-oxidation protein (X80052) from *Neurospora crassa*. Besides this, most of the clusters showed protein homology to other plant species, mainly to those taxonomically related to sugarcane such as maize (*Zea mays*), rice (*Oryza sativa*), *Sorghum* (*Sorghum bicolor*), and gramines (*Hordeum vulgare*, *Triticum aestivum*, and *Cynodon dactylon*).

Polymorphism detected by EST-SSRs in sugarcane

Of the 30 designed EST-SSRs, 23 (77%) were able to produce scorable polymorphisms (Table 6), whereas the others failed to amplify or did not produce clearly interpretable patterns. All the EST-SSRs that did amplify produced fragments of predicted size.

Table 3. SSR structure types for each repeat class found in the SUCEST database.

Repeat class	Perfect (%)	Compound (%)	Imperfect (%)	More than one SSR region (%)
Dinucleotide	70.3	10.3	3.0	16.4
Trinucleotide	71.3	3.4	3.2	12.1
Tetranucleotide	78.0	2.0	2.5	17.5
Total	75.1	3.0	2.8	19.1

Table 4. Comparisons between SSR repeat length derived from SUCEST and sugarcane genomic libraries.

	Mean		Range	
Repeat class	SUCEST	Genomic[a]	SUCEST	Genomic[a]
Dinucleotide	15.80	15.5	7–145	6–57
Trinucleotide	6.44	12.0	5–33	8–92
Tetranucleotide	3.44	n.a.	3–25	n.a.

Note: n.a., not available.
[a]Extracted from Cordeiro et al. 2000.

Owing to the polyploid nature of sugarcane, most of the EST-SSRs produced more than two amplified products (Fig. 3) per individual that were assumed to be alleles. The number of alleles among the 18 sugarcane commercial clones (including the mapping parents), varied from 2 to 15, with polymorphism information content (PIC) values between 0.28 and 0.90. Apparently, no differences were observed in the level of polymorphism generated among the three sets (di, tri, and tetranucleotides) of EST-SSRs. The average number of alleles and the mean PIC values of each set of repeat type showed similar values. The mean PIC value was slightly superior for the dinucleotide primers (0.74) than for the trinucleotides (0.60) and tetranucleotides (0.66). Considering all the EST-SSRs, the mean number of alleles and the mean PIC value obtained by these functional markers were, respectively, 6.04 and 0.66. Almost all dinucleotide EST-SSRs were effective in distinguishing, in terms of percentage of different patterns, more than 50% of the 18 sugarcane commercial clones (Table 6). One trinucleotide (SCB07) and one tetranucleotide (SCC01) EST-SSR primer pair alone were able to produce 18 different sugarcane patterns.

Segregation analysis of the EST-SSR markers

The 23 EST-SSRs tested on the two parents (SP 80-180 and SP 80-4966) and 6 F_1 sampled individuals produced 41 markers that were polymorphic between the parents and within the progeny. Part of these markers should correspond to single dose markers segregating in a 1:1 ratio in the mapping population, while the other part should correspond to double dose markers with a 3:1 segregation ratio. Only 11 monomorphic markers between the parents were polymorphic in the 6 F_1 sampled individuals. Part of them should correspond to markers present in a single dose in both parents and should segregate in a 3:1 ratio in the progeny.

Discussion

The number of clusters found with repetitive sequences in the SUCEST database demonstrates its large potential for SSRs marker development in a simple, fast, and economical manner. In general, this number represents 4.6% of the entire cluster population of the bank and is somewhat underestimated, since 19% of these clusters contained more than one SSR region.

Despite the different search criteria reported for SSR research in the EST databases, the percentage of EST-SSRs found in the SUCEST database was similar to those observed in other species (Scott et al. 2000; Cardle et al. 2000; Hackauf and Wehling 2002) and superior to the value of 2.88% reported for sugarcane (Cordeiro et al. 2001).

Low frequencies of SSR are expected in EST sequences, which represent the coding portion of the genome. In sugarcane, the frequency of SSRs derived from enriched libraries (Cordeiro et al. 2000) was 20 times higher than the frequency of SSRs found in EST databases (Cordeiro et al. 2001).

Mining SSRs against sequences derived straight from individual ESTs can provide some degree of redundancy. Redundancy occurs because a group of ESTs could be transcribed from the same mRNA. To avoid this, the authors search for SSRs in the cluster consensus sequences that represent all ESTs likely derived from the same transcript molecule (Telles and Silva 2001).

Trinucleotide repeats tend to prevail in protein-coding exons, as well as in SSR-ESTs, as compared with dinucleotide and tetranucleotide repeats (Tóth et al. 2000; Varshney et al. 2002). In the SUCEST database, the percentage of trinucleotide repeats were almost half the percentage of tetranucleotides. This could be due to the high number of tetranucleotide motif types BLASTed in the search process being three times higher than that of trinucleotides. Moreover, the cut-off value was lower for tetranucleotides ($n \geq 3$) than for trinucleotides ($n \geq 5$).

The first studies carried out to investigate the abundance of different SSR motifs in plant databases pointed to AT/AT and GA/TC as the most common dinucleotide motifs (Lagercrantz et al. 1993); AAT/ATT, AAC/GTT, and AGC/GCT as the most common trinucleotides; and AATT/AATT and AAAT/ATTT as the most common tetranucleotides (Wang et al. 1994). However, significant differences in the motif distributions between coding and non-coding regions were found in different eukaryotic genomes, being also characteristic of the taxonomic group examined (Tóth et al. 2000). The SUCEST motifs, CGG/CCG and CTC/GAG, were among the most abundant trinucleotides, whereas AAT/ATT was among the least frequent. Interestingly, this same trend was observed for trinucleotides in maize coding sequences derived from cDNAs, ESTs, and known genes. It seems that these differences are related to the significance of codons having a direct effect on protein synthesis, such as TAA variants that code for stop codons (Chin et al. 1996). Moreover, the abundance of CCG motifs in EST-SSRs from dif-

Fig. 1. Distributions of dinucleotide, trinucleotide, and tetranucleotide SSRs according to the repeat number.

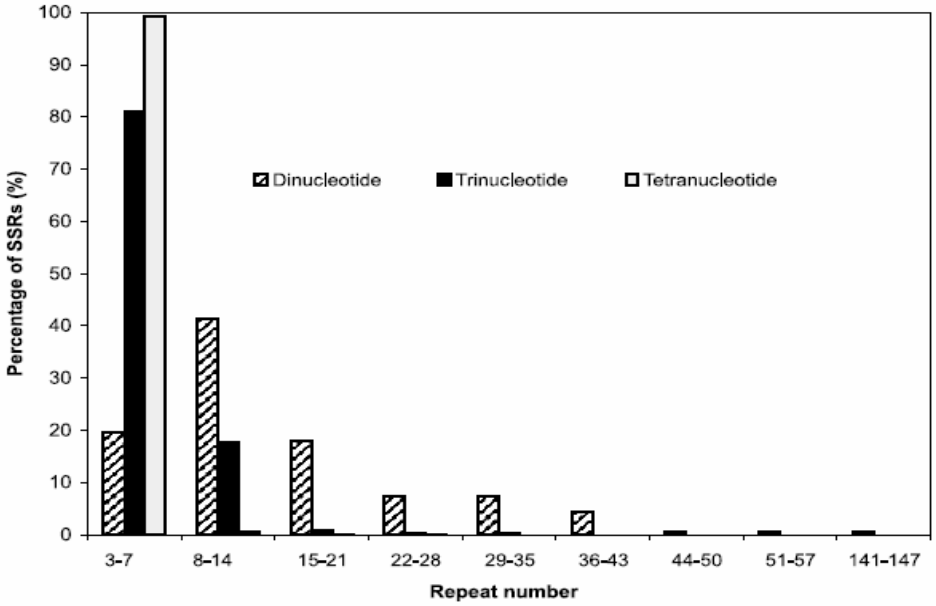
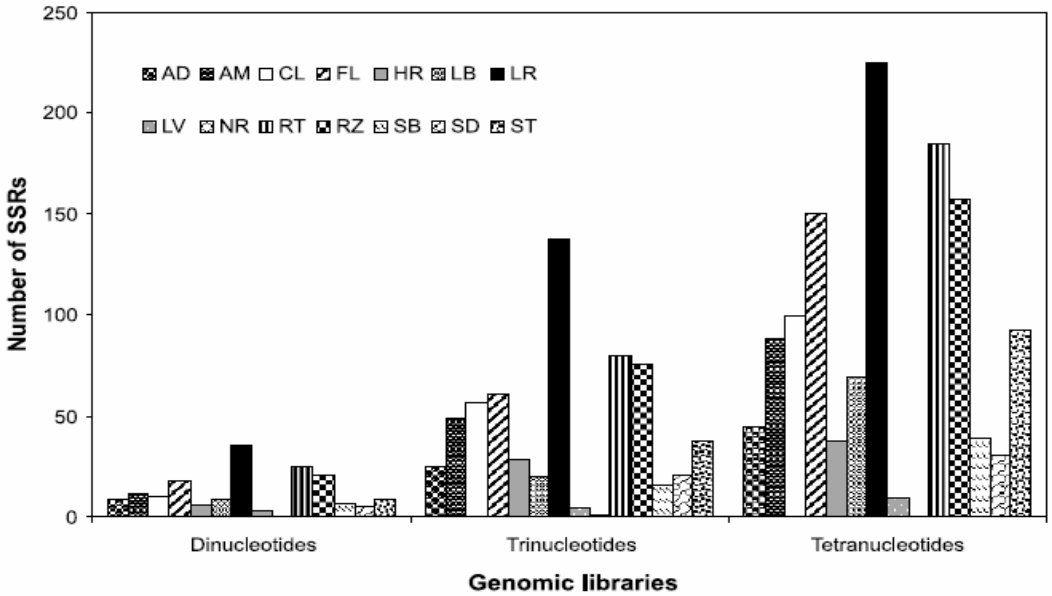


Fig. 2. Distribution of the SSRs in different cDNA libraries. AD, *Acetobacter diazotroficans*; AM, apical meristem; CL, callus; FL, flower; HR, *Herbaspirilum rubrisubalbicans*; LB, lateral bud; LR, leaf roll; LV, leaves; NR, all normalized tissue libraries; RT, root; RZ, leaf–root transition zone; SB, stem bark; SD, seeds; ST, stem.



ferent cereal species such as barley, maize, oats, rye, rice, and *Sorghum* was considered a feature of the monocot genomes that has an increased G+C content. Apart from this, except for barley, CG was the least common dinucleotide motif found in all cereal species (Varshney et al. 2002). These results are in perfect agreement with our findings for SSRs from the SUCEST database.

Differences in the abundance of the motif types were observed between the SUCEST SSRs and those derived from sugarcane genomic libraries, specifically for trinucleotides (Cordeiro et al. 2000). One of the least frequent trinucleotide motifs, $(CGG)_n/(CCG)_n$, isolated from the enriched process was among the most abundant in the SUCEST database. Although comparisons of this kind among dinucleotide motifs can be biased because of palindromic sequence exclusions, such as $(CG)_n$ and $(AT)_n$, from the enrichment process and because of the small number of different dinucleotide motif types used in the isolation of SSRs, the most frequent

Table 5. Classification of clusters containing SSRs according to their probable metabolic and biochemical functions.

Metabolic and (or) biochemical function[a]	Percentage (%)[b]
Stress- or defense-related proteins	2.0
RNA-binding proteins	5.0
DNA-binding proteins	8.6
Structural proteins	3.3
Enzymes	18.0
Regulatory proteins	11.7
ATP synthases, electron transport, and transport-like proteins	6.3
Membrane protein	2.0
Others	20.0
Unknown or hypothetical protein	23.0

[a]Classification based on Carson and Botha (2000).
[b]Percentage calculated only with clusters showing scores ≥ 80.

Table 6. Polymorphism characterization of the 23 EST-SSR.

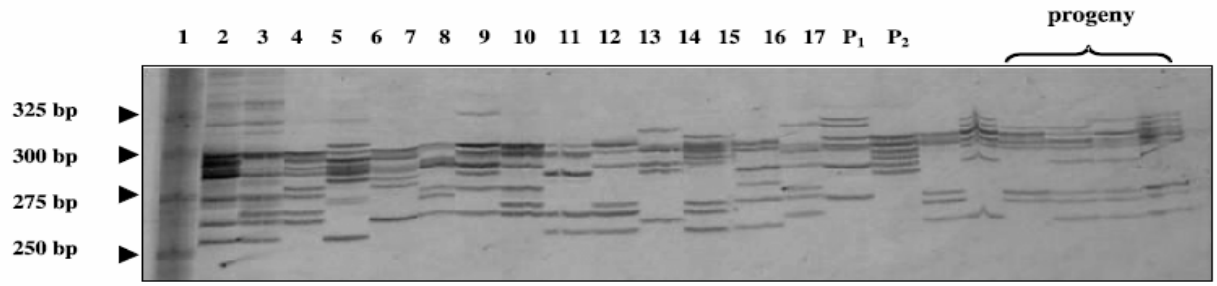
Primer	T_g (°C)[a]	No. of alleles	Mean no. of alleles	PIC	Mean PIC value	Different patterns (%)	Segregation analysis[b] 1:1[c]	Segregation analysis[b] 3:1[d]
SCA03	52	5		0.71		33	2	1
SCA04	52	3		0.59		11	0	1
SCA06	52	9		0.81		83	4	0
SCA07	63	8	6.4	0.82	0.74	78	3	1
SCA08	62	8		0.79		62	3	1
SCA09	62	5		0.71		86	n.a.	n.a.
SCA10	63	7		0.8		89	4	0
SCB01	56	4		0.7		50	1	0
SCB02	60	6		0.34		37	n.a.	n.a.
SCB03	56	6		0.72		44	3	0
SCB06	60	5	5.4	0.67	0.6	35	0	1
SCB07	62	13		0.9		100	7	1
SCB08	50	3		0.59		17	0	1
SCB09	50	4		0.58		17	0	0
SCB10	62	2		0.28		11	0	1
SCC01	50	15		0.9		100	5	1
SCC02	62	5		0.73		11	1	0
SCC03	50	2		0.3		11	1	0
SCC04	52	9	6.4	0.84	0.66	83	2	1
SCC05	62	8		0.82		78	4	0
SCC07	63	4		0.62		33	0	0
SCC09	50	2		0.36		11	0	0
SCC10	63	6		0.69		28	1	1

Note: n.a., not available.
[a]Annealing temperature of each primer pair identified by a Gradient thermocycler.
[b]Number of markers between the parents (SP 80-180 and SP 80-4966), potentially useful for sugarcane mapping. Values obtained in a test sample of 6 F_1 individuals derived from the cross between SP 80-180 and SP 80-4966.
[c]Marker present in only one parent and absent in at least one of the 6 F_1 individuals with potential to segregate in a 1:1 ratio in the population.
[d]Marker present in both parents and absent in at least one of the 6 F_1 individuals with a potential to segregate in a 3:1 ratio in the population.

dinucleotide motif $(GT)_n/(AC)_n$ derived from sugarcane genomic libraries (Cordeiro et al. 2000) was also one of the most abundant dinucleotide motifs in the SUCEST database.

Several studies have shown that repeat length is positively correlated to the informativeness of a microsatellite (Weber 1990; Temnykh et al. 2001; Jones et al. 2001). Our EST-SSRs had a low number of repeat units, especially for the tri- and tetranucleotides that were close to the cut-off value used in the search criteria. In EST-SSRs derived from different cereal species, only 25% fit the category of more than 10

Fig. 3. Silver-stained denaturing polyacrylamide gel of the EST-SSRs SCC01: lane 1, 25-bp size marker; lane 2, CB 3642; lane 3, SP 79-1011; lane 4, CB 40-77; lane 5, IAC 51-205; lane 6, RB 739-359; lane 7, SP 70-1284; lane 8, IAC 64-257; lane 9, SP 71-1406; lane 10, SP 80-3280; lane 11, RB 85-5035; lane 12, SP 79-6134; lane 13, SP 79-2313; lane 14, RB 855563; lane 15, Gandacheni (*S. barberi*); lane 16, IJ76-314 (*S. officinarum*); lane 17, Maneria (*S. sinense*); lane 18, SP 80-4966; lane 19, SP 80-180, final lanes, F_1 individuals.



repeat units, and in the tetrameric, pentameric, and hexameric SSRs, all (100%) fall into the category of fewer than 10 repeat units (Varshney et al. 2002). Furthermore, differences in the length of the repeats were evident between trinucleotide motifs derived from sugarcane genomic libraries (Cordeiro et al. 2001) and those from the SUCEST database, with the former having longer repeats.

The majority of clusters with SSRs were derived from the cDNA leaf roll libraries. This may be due to the fact that leaf roll constitutes a meristematic region of the plant that is highly active metabolically (Carson and Botha 2000) with transcription of several genes occurring at a higher rate compared with other plant parts.

In relation to the polymorphism levels reveled by SUCEST SSRs, the mean number of alleles and the average PIC value (6.04 and 0.66, respectively) were somewhat inferior to those observed for genomic SSR primers (8.00 and 0.72, respectively) screened over five *Saccharum* genotypes (Cordeiro et al. 2000). However, the values of polymorphism detected in our study were superior to that displayed by EST-SSRs within sugarcane species (average PIC value of 0.23) reported by Cordeiro et al. (2001). This was expected, since our EST-SSR primers were screened over sugarcane commercial clones (interspecific hybrids) previously selected according to their genetic divergence and reported in a cluster analysis based on AFLP markers (Lima et al. 2002).

The large number of markers (alleles) displayed in a single individual can provide unique fingerprints, useful for sugarcane variety identification. Some EST-SSRs were promising in terms of distinguishing between the commercial clones evaluated providing fingerprinting at the functional level.

The high polyploidy of sugarcane cultivars generates constraints for genetic mapping. In particular, only single-dose markers, and sometimes double-dose markers, are used for genetic mapping, although double-dose markers are less informative. Single-dose markers (present in one of the parents and in one copy) segregate in a 1:1 ratio in an F_1 progeny. The 23 EST-SSRs evaluated on both parents and in the 6 F_1 progeny of our mapping population revealed 52 polymorphic markers that will be worth analyzing on the mapping population to select and map the one corresponding to single-dose markers.

For species in which a sufficient number of ESTs exist, as is the case of the SUCEST database, EST-SSR markers will lead to the direct mapping of genes and provide a transcriptional map. The EST-SSRs derived from the SUCEST database will complement the 259 genomic SSRs developed by the International Sugarcane Microsatellite Consortium, thereby extending the assessment of variability to the whole-genome level.

Acknowledgements

The authors would like to thank Dr. Giovanni M. Cordeiro, Centre for Plant Conservation Genetics, Southern Cross University, Lismore, Australia, for supplying the PIC calculator program. The present research was financed by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 02/01167-1) and COPERSUCAR (Coperativa dos Produtores de Açúcar e Álcool do Estado de São Paulo). L.R.P. and K.M.O. receive, respectively, post-doctoral and doctoral fellowships from FAPESP (01/14656-8; 02/00197-4). A.P.S. receives a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Cardle, L., Ramasay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, **156**: 847–854.

Carson, D.L., and Botha, F.C. 2000. Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci. **40**: 1769–1779.

Cato, S.A., Gardner, R.C., Kent, J., and Richardson, T.E. 2001. A rapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. **102**: 196–306.

Chin, E.C.L., Senior, M.L., Shu, H., and Smith, J.S.C. 1996. Maize simple repetitive DNA sequences: abundance and allele variation. Genome, **39**: 866–873.

Cordeiro, G.M., Taylor, G.O., and Henry, R.J. 2000. Characterization of microsatellite markers from sugarcane (*Saccharum* spp.), a highly polyploid species. Plant Sci. **155**: 161–168.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Sci. **160**: 1115–1123.

Creste, S., Tulmann Neto, A., and Figueira, A. 2002. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gel by silver staining. Plant Mol. Biol. Rep. **19**: 299–306.

Daugrois, J.H., Grivet, L., Roques, D., Hoarau, J.Y., Lombard, H., Glaszmann, J.C., and D'Hont, A. 1996. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. Theor. Appl. Genet. **92**: 1059–1064.

Hackauf, B., and Wehling, P. 2002. Identification of microsatellite polymorphisms in an expressed portion of rye genome. Plant Breed. **121**: 17–25.

Hogarth, D. 1987. Genetics of sugarcane. Sugar improvement through breeding. Elsevier Press, Amsterdam, the Netherlands.

Hoisington, D., Khairallah, M., and González-De-León, D. 1994. Laboratory protocols. CIMMYT Applied Molecular Genetics Laboratory, Mexico City, Mexico.

Jarne, P., and Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. Trends Mol. Ecol. **11**: 424–429.

Jones, E.S., Dupal, M.P., Kölliker, R., Drayton, M.C., and Foster, J.W. 2001. Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor. Appl. Genet. **102**: 405–415.

Lagercrantz, U., Ellegren, H., and Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**: 1111–1115.

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L., and Quackenbush, J. 2000. An optimized protocol for analysis of EST sequences. Nucleic Acids Res. **28**: 3657–3665.

Lima, M.L.A., Garcia, A.A.F., Oliveira, K.M., Matsuoka, S., Arizono, A., Souza Jr., C.L., and Souza, A.P. 2002. Anaysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor. Appl. Genet. **104**: 30–38.

Powell, W., Machray, G.C., and Provan, J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. **1**: 215–221.

Price, S. 1965. Interspecific hybridization in sugarcane breeding. Proc. Int. Soc. Sugar Cane Technol. **12**: 1021–1026.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723–726.

Silva, J.A.G. 2001. Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). Genet. Mol. Biol. **24**: 155–159.

Silva, J.A.G., Sorrells, M.E., Burnquist, W.L., and Tanksley, S.D. 1993. RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome, **36**: 782–791.

Sterky, F., and Lundeberg, J. 2000. Sequence analysis of genes and genomes. J. Biotechnol. **76**: 1–31.

Telles, G.P., and Silva, F.R. 2001. Trimming and clustering sugarcane ESTs. Genet. Mol. Biol. **24**: 17–23.

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

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11**: 1441–1452.

Tóth, G., Zoltán, G., and Jurkaet, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. **10**: 967–981.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. 2002. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell. Mol. Biol. Lett. **7**: 537–546.

Vettore, A.L., Da Silva, F.R., Kemper, E.L., and Arruda, P. 2001. The libraries that made SUCEST. Genet. Mol. Biol. **24**: 1–7.

Wang, Z., Weber, J.L., Zhong, G., and Tankesley, S.D. 1994. Survey of plant short tandem DNA repeats. Theor. Appl. Genet. **88**: 1–6.

Weber, J.L. 1990. Informativeness of human (dC-dA)n.(dG-dT) polymorphisms. Genomics, **7**: 524–530.

Wolfsberg, T.G., and Landsman, D. 2001. Expressed sequence tags (ESTs). Bioinformatics: a practical guide to the analysis of genes and proteins. John Wiley and Sons, New York, N.Y.

Zane, l., Bargelloni, L., and Patarnello, T. 2002. Strategies for microsatellite isolation: a review. Mol. Ecol. **11**: 1–16.

5 Artigo II

"Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs"

L.R. Pinto, K.M.Oliveira, T.G. Marconi, A.A.F. Garcia, E.C. Ulian, A.P. de Souza

Publicado na revista

Plant Breeding (125: 378-384; 2006)

Plant Breeding **125**, 378—384 (2006)

Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs

L. R. PINTO[1], K. M. OLIVEIRA[2], T. MARCONI[2], A. A. F. GARCIA[3], E. C. ULIAN[4] and A. P. DE SOUZA[2]

[1]Centro Avançado da Pesquisa do Agronegócio de Cana, Instituto Agronômico de Campinas-IAC, Anel Viário Contorno Sul, km 321, CP 206, CEP 14001-970, Ribeirão Preto, São Paulo, Brasil, E-mail: lurossini@iac.sp.gov.br; [2]Centro de Biologia Molecular e Engenharia Genética (CBMEG) – Universidade Estadual de Campinas (UNICAMP), Cidade Universitária 'Zeferino Vaz',CP 6010, CEP 13083-970, Campinas-São Paulo, Brasil; [3]Depto de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-São Paulo, Brasil; [4]Centro de Tecnologia Canavieira, Piracicaba-São Paulo, Brasil

With 2 figures and 3 tables

Received October 21, 2005/Accepted January 23, 2006
Communicated by W. E. Weber

Abstract

Microsatellites or simple sequence repeats (SSRs) are one of the most suitable markers for genome analysis as they have great potential to aid breeders to develop new improved sugarcane varieties. The development of SSR derived from expressed sequence tags (EST) opens new opportunities for genetic investigations at a functional level. In the present work, the polymorphism obtained with a subset of 51 EST–SSRs derived from SUCEST was compared with those generated by 50 genomic SSRs (gSSR) in terms of number of alleles, polymorphism information content, discrimination power and their ability to establish genetic relationships among 18 sugarcane clones including three *Saccharum* species (*S. officinarum*, *S. barberi*, *S. sinense*). The majority of EST–SSRs loci had four to six alleles in contrast to the seven to nine observed for the gSSRs loci. Approximately, 35% of the gSSRs had PIC values around 0.90 in contrast to 15% of the EST–SSRs. However, the mean discrimination power of the two types of SSR did not differ significantly as much as the average genetic similarity (GS) based on Dice coefficient. The correlation between GS of the two types of SSRs was high ($r = 0.71/P = 0.99$) and significant. Although differences were observed between dendrograms obtained with each SSR type, both were in good agreement with pedigree information. The *S. officinarum* clone IJ76-314 was grouped apart from the other clones evaluated. The results here demonstrate that EST–SSRs can be successfully used for genetic relationship analysis, extending the knowledge of genetic diversity of sugarcane to a functional level.

Key words: *Saccharum officinarum* — microsatellites — EST

Modern sugarcane varieties are mainly derived from interspecific crosses between the noble cane *S. officinarum* (2n = 80) and the wild species *S. spontaneum* (*2n* = 40–128) that was chosen for its resistance to pathogens and other stresses. Successive backcrosses of the hybrids with *S. officinarum* were carried out to recover the high-sugar producing types. Thus, present-day sugarcane cultivars possess a complex interspecific aneu-polyploid genome with chromosome numbers ranging from 100 to 130 (Hoarau et al. 2005). This genomic complexity added to the multigenic and/or multi-allelic nature of most agronomic traits makes sugarcane breeding a difficult task (Casu et al. 2005).

Molecular markers have a great potential to help breeders to develop new improved sugarcane varieties. Microsatellites or simple sequence repeats (SSRs) are one of the most suitable markers for genome analysis. They are short segments of DNA of one to six tandem repeated base pairs that are commonly isolated from enriched libraries in a laborious and time-consuming process (Jarne and Lagoda 1996). Microsatellite markers have many applications in sugarcane genetics and breeding processes including germplasm analysis, cultivar identification, parent evaluation, genetic mapping and marker-assisted selection (Cordeiro et al. 2003, Pan et al. 2003a,b, Aitken et al. 2005). The International Sugarcane Microsatellite Consortium funded by the International Consortium for Sugarcane Biotechnology (ICSB) developed 259 genomic microsatellites (gSSRs) to be used for these purposes (Cordeiro et al. 1999, 2000).

Nowadays microsatellites have been developed for many plant species through SSR mining in expressed sequence tags (EST) obtained from partial sequencing of cDNAs (Cho et al. 2000, Scott et al. 2000, Thiel et al. 2003). A large quantity of these ESTs has been deposited in the EST databases (http://ncbi.nlm.nih.gov/dbEST/). As ESTs represent part of expressed genes, they can provide direct mapping of genes with known functions allowing candidate genes, involved in important metabolic pathways to be directly evaluated for associations with important agronomic traits. Moreover, they are ideal for marker-assisted selection as they themselves may be the gene which is responsible for the trait of interest (Cato et al. 2001, Ma et al. 2004).

The SUCEST, Sugarcane Expressed Sequence Tag Project (http://sucest.lad.ic.unicamp.br/en/) has produced a large collection of ESTs with a great potential to generate molecular markers. By mining dbSUCEST for SSR, 2005 clusters containing SSR (EST–SSRs) that are being used as molecular markers in a sugarcane mapping project (Pinto et al. 2004, Garcia et al. 2006) were found. The development of microsatellites derived from expressed sequence tags (EST) opens new opportunities for genetic investigations at a functional level. In an attempt to explore the large amount of sequence information deposited in the SUCEST database for sugarcane breeding, the Plant Molecular Biology Laboratory (UNICAMP, Campinas, São Paulo, Brazil), in a collaborative effort with the Centro de Tecnologia Canavieira (CTC, Piracicaba, São Paulo, Brazil), is developing a working primer set of EST–SSRs.

The present study was undertaken to evaluate the potential of a novel set of (EST–SSRs) developed from the SUCEST database as molecular markers for sugarcane breeding. The objectives were to (i) characterize the polymorphism levels of these novel EST–SSRs, (ii) verify the discrimination power of the EST–SSRs for fingerprinting, and (iii) compare the polymorphism between EST–SSRs and gSSR considering a common subset of sugarcane clones.

Materials and Methods

Plant materials: The level of polymorphism generated by EST–SSRs and gSSRs was evaluated against 13 commercial clones of sugarcane, *Saccharum officinarum*, three *Saccharum* species (*S. officinarum*, *S. barberi* and *S. sinense*) and the parents (SP80-180 and SP80-4966) of a mapping population from the Centro de Tecnologia Canavieira (CTC-Piracicaba, Sao Paulo, Brazil, Table 1).

DNA extraction and PCR amplification: A modified CTAB-method (Hoisington et al. 1994) was used to extract total genomic DNA from 300 mg of lyophilized young leaf tissue ground to a fine powder. PCR reactions were performed in a 20-μl final volume containing 40 ng of template DNA, 0.2 μM of each forward and reverse primers, 100 μM of each dNTP, 2.0 mM $MgCl_2$, 10 mM Tris-HCl, 50 mM KCl, and 0.5 unit *Taq* DNA polymerase (Invitrogen, São Paulo, Brasil). Reactions were amplified on a PTC-100 thermocycler (MJ Research, Watertown, MA, USA) as follows: 94°C for 3 min; followed by 30 cycles of 94°C for 1 min; annealing temperature specific for each primer for 1 min; extension of 72°C for 1 min and a final elongation step at 72°C for 2 min. Amplification products were separated by electrophoresis on 6% denatured PA gels using a 25-bp ladder as a size standard and silver-stained according to Creste et al. (2002).

SSRs: EST–SSRs were developed from cluster consensus sequences derived from the Sugarcane Expressed Sequence Tag database (SUCEST/http://sucest.lbi.dcc.unicamp.br/en/) as detailed described by Pinto et al. (2004). A set of 100 EST–SSR was designed using the Primer Select software (LaserGene versions 5.01/5.02, DNAStar, Inc., Madison, WI, USA) and synthesized by Invitrogen. The stringency criteria used were the same as described by Cordeiro et al. (2000). The results obtained with EST–SSRs were compared with 50 gSSRs (randomly chosen) developed from the International Consortium for Sugarcane Biotechnology (ICSB; Cordeiro et al. 1999, 2000) screened over the same sugarcane clones and three *Saccharum* species.

Polymorphism analysis: The sugarcane clones were scored for the presence (1) or absence (0) of common bands present in the different patterns. The amplified fragments produced by the SSRs (EST–SSRs and gSSRs) were considered alleles of a single locus. The polymorphic information content for each locus was calculated according to Cordeiro et al. (2003). The discrimination power was calculated for each EST–SSR based on Tessier et al. (1999) and was used as a measure of SSR loci efficiency for variety identification.

Genetic-similarity estimate: The final presence–absence matrix, constructed only with polymorphic bands, was used to estimate the genetic similarity between all the clones evaluated, adopting the Dice similarity coefficient: (Dice 1945). The genetic relationships among the clones were visualized by a dendrogram based on the Unweighted Pair Group Method with Arithmetic Averages, UPGMA (Meyer et al. 2004). The cophenetic coefficient (r_{coph}) was calculated to measure the correspondence between the genetic similarity matrix and the cluster analysis (dendrogram). The above analyses were performed with NTSYS-PC software, version 2.0 (Exeter Software, NY, USA; Rohlf 1993). A bootstrap method with 1000 replicates was performed to verify if the number of polymorphic markers used for genetic similarity estimation was enough to support precise estimates among the sugarcane clones evaluated by the programme (Lima et al. 2001). This was performed using a programme developed in the R software (http://www.R-project.org).

Genetic similarity comparison between EST–SSRs and gSSRs: The two genetic similarity matrices obtained with EST–SSR and gSSRs data were compared with measure the degree of relationship between them by computing the product–moment correlation (r) and the Mantel test statistic (Z) using NTSYS-PC software, version 2.0 (Exeter Software; Rohlf 1993). Such comparison was carried out to verify if the genetic similarities generated by the two types of SSRs data (EST–SSRs and gSSRs) provide similar genetic measures.

Results

Polymorphism analysis

The 100 EST–SSRs were initially tested in agarose gels to verify the presence and quality of product amplification. Out of these, 62 were screened in denaturing polyacrylamide gels for banding patterns of high quality, with 51 EST–SSRs being selected for genetic similarity analysis (Table 2). Most of these EST–SSRs have homology to genes of interest (Table 3). The

Table 1: Sugarcane clones investigated in the present study with their pedigrees relationship covering two generations and their breeding origin

Clone	Pedigree[1]	Origin
CB3624	[POJ2364 × EK28] POJ 2878 × ?	Campos, Brazil
SP791011	[CO419 × CO419] NA5679 × CO775 [POJ2878 × CO371]	São Paulo, Brazil
CB4077	[POJ2364 × EK28] POJ2878 × CO290 [CO221 × D74]	Campos, Brazil
IAC51205	[POJ2364 × EK28] POJ2878 × ?	Campinas, Brazil
RB739359	[CO419 × MZ336] IANE5534 × ?	Republic of Brazil
SP701284	[POJ2878 × ?] B4176 × ?	São Paulo, Brazil
IAC64257	[POJ2878 × CO290] CO419 × IAC49-131 [CP27108 × ?]	Campinas, Brazil
SP711406	[CO419 × CO419] NA5679 × ?	São Paulo, Brazil
SP803280	[CP5530 × CP5376] SP71-10088 × H575028 [H49134 × ?]	São Paulo, Brazil
RB855035	[CP521 × CP48103] L6014 × SP701284 [CB4176 × ?]	Republic of Brazil
SP796134	[H53263 × H507209] H634644 × ?	São Paulo, Brazil
SP792313	[CP5659 × ?] SP71-6106 × ?	São Paulo, Brazil
RB855536	[IAC48-65 × ?] SP70-1143 × RB72454 [CP5376 × ?]	Republic of Brazil
SP49666	[NA5679 × ?] SP711406 × ?	São Paulo, Brazil
SP80180	[B3337 × ?] B3337 × ?	São Paulo, Brazil
Ganda Cheni (*S. barberi*)	*S. barberi* × ?	Saretha, India
IJ76-314 (*S. officinarum*)	*S. officinarum* × ?	Iryan, Java
Maneria (*S. sinense*)	*S. sinense* × ?	Pansahi, China

[1]Extracted from Caiana/IAC software.
?, polycross.

Table 2: Polymorphism characterization for 51 EST–SSRs and 50 gSSRs. Number of alleles (A), expected product size, allele range (bp), polymorphism information content (PIC) and discrimination power (*D*)

EST–SSR[1]	A	Expected size[2]	Allele range (bp)[3]	PIC	*D*	gSSR	A	Expected size[2]	Allele range (bp)[3]	PIC	*D*
SCA03	5	181	187–209	0.71	0.79	SMC863CG	7	296	268–309	0.82	1.00
SCA04	3	170	184–227	0.59	0.21	SMC765BS	8	298	46–326	0.83	0.98
SCA06	9	238	215–233	0.81	0.98	SMC39BUQ	10	143	145–163	0.87	0.97
SCA07	8	201	207–227	0.82	0.96	CIR67	12	142	72–139	0.89	0.98
SCA10	7	195	204–248	0.80	0.99	SMC415MS	4	338	284–304	0.86	0.99
SCA12	11	269	223–251	0.87	0.98	SMC2055FL	10	318	248–272	0.83	0.96
SCA14	7	186	193–207	0.78	0.97	SMC21AS	15	133	121–146	0.91	1.00
SCA16	10	165	167–189	0.87	0.99	SMC119CG	12	119	109–168	0.89	1.00
SCA17	11	212	146–192	0.75	0.82	CIR36	16	166	128–158	0.91	1.00
SCA19	2	191	202–216	0.21	0.30	CIR55	9	333	288–341	0.83	0.99
SCA22	4	185	180–194	0.65	0.80	SMC36BUQ	6	118	117–163	0.76	0.93
SCA26	10	183	171–200	0.87	0.98	SMC319CG	10	183	139–166	0.87	0.93
SCB01	4	193	165–196	0.70	0.91	SMC1055HA	7	224	212–278	0.79	0.99
SCB02	6	164	182–205	0.34	0.81	CIR35	8	173	154–182	0.80	0.98
SCB06	5	176	163–193	0.67	0.70	CIR14	5	240	223–236	0.64	0.79
SCB07	13	291	223–300	0.90	1.00	CIR11	7	235	214–156	0.83	0.56
SCB08	3	176	176–183	0.59	0.45	CIR21	18	181	106–149	0.91	1.00
SCB12	6	147	107–182	0.71	0.60	CIR56	17	184	128–170	0.90	1.00
SCB13	5	133	132–145	0.67	0.86	SMC843BS	12	201	172–208	0.90	0.97
SCB14	12	297	263–289	0.90	1.00	SMC687CS	5	203	195–211	0.67	0.81
SCB16	6	234	262–306	0.71	0.87	SMC1011HA	8	276	174–233	0.83	0.81
SCB17	6	158	94–166	0.80	0.86	CIR58	7	318	287–299	0.90	1.00
SCB25	6	191	192–207	0.78	0.86	CIR12	14	279	151–279	0.81	1.00
SCB27	6	213	203–232	0.78	0.94	SMC1235FL	9	88	76–107	0.83	1.00
SCB35	6	180	199–217	0.74	0.82	SMC448MS	17	187	113–232	0.92	1.00
SCB37	6	187	182–214	0.77	0.87	SMC1047HA	15	166	137–171	0.89	1.00
SCB38	5	210	207–229	0.72	0.81	CIR24	7	265	216–242	0.78	0.93
SCB39	8	146	111–144	0.83	0.96	CIR25	5	300	274–293	0.68	0.90
SCB40	7	155	165–188	0.79	0.91	CIR26	10	134	116–138	0.83	0.92
SCB41	21	163	81–165	0.93	0.98	CIR28	12	419	255–315	0.89	1.00
SCB43	4	330	255–290	0.67	0.78	CIR50	9	263	185–298	0.83	0.97
SCB45	12	110	94–160	0.84	0.99	CIR4	7	289	263–298	0.81	0.98
SCB47	7	236	220–242	0.82	0.83	SMC232MS	3	157	162–210	0.47	0.21
SCB48	3	218	199–213	0.57	0.29	SMC477CG	7	168	88–139	0.79	0.64
SCB49	3	454	322–373	0.59	0.60	SMC31CUC	13	225	149–190	0.89	1.00
SCB51	6	416	339–372	0.76	0.97	SMC2017FL	14	229	215–252	0.90	1.00
SCB52	11	145	89–149	0.87	0.95	SMC1069HA	5	195	155–206	0.55	0.81
SCB53	12	113	95–142	0.90	0.81	SMC222CG	9	198	158–183	0.85	0.97
SCC01	15	298	266–306	0.90	1.00	CIR23	9	281	261–332	0.85	0.95
SCC02	5	202	198–217	0.73	0.28	CIR51	6	272	251–279	0.72	0.75
SCC03	2	227	230–242	0.30	0.35	SMC2039FL	7	275	146–163	0.79	0.91
SCC04	9	218	198–214	0.84	0.98	SMC2024FL	8	275	193–206	0.84	0.97
SCC05	8	179	149–196	0.82	0.97	SMC260MS	5	236	201–223	0.71	0.93
SCC07	4	228	201–228	0.62	0.83	SMC280CS	3	238	215–221	0.41	0.11
SCC10	6	252	220–262	0.69	0.67	CIR32	15	211	162–248	0.90	1.00
SCC13	7	356	303–352	0.79	0.97	SMC236CG	8	210	161–225	0.82	0.53
SCC15	5	203	152–215	0.71	0.84	CIR18	8	215	170–189	0.84	0.96
SCC16	3	229	227–251	0.45	0.31	CIR31	7	215	183–197	0.81	0.94
SCC17	14	137	115–150	0.89	1.00	CIR1	21	227	118–244	0.93	1.00
SCC18	4	129	121–132	0.53	0.65	CIR74	8	226	210–232	0.78	0.75
SCC19	8	238	242–281	0.78	0.94						
Total	369			—	—		474			—	—
Mean	7.2			0.73	0.80		9.5			0.82	0.89
$CI_{(95\%)}$	—			0.69–0.77	0.74–0.86		—			0.79–0.85	0.84–0.94

[1]EST–SSR primer sequences are under a confidentially agreement. Contact A. P. de Souza (anete@unicamp.br).
[2]Expected size of the amplified product obtained by primer select (Laser gene) programme.
[3]Allele range estimated by linear regression.

51 EST–SSRs detected 369 alleles of which 64 (17%) were monomorphic among the clones evaluated. Allele number ranged from 2 (SCA19; SCC03) to 21 (SCB41) with a mean of 7.2 alleles per EST–SSR. Most of the EST–SSRs showed high values of polymorphic information content (PIC) with an average of 0.73. The greatest PIC value was observed for SCB41 (0.93) whereas the lowest one (0.21) was for SCA19.

The 50 gSSRs used for polymorphism comparison generated 474 alleles of which 44 (9%) were monomorphic. The number of alleles varied from 3 (SMC232MS) to 21 (CIR1), with a mean of 9.5 alleles per gSSR. The PIC values ranged from 0.41 (SMC280CS) to 0.93 (CIR1) with a mean value of 0.82.

Overall, the polymorphism obtained with gSSR, in a similar sample of sugarcane clones, was higher than those obtained with EST–SSRs. This was based on the comparative distributions of polymorphism shown in Fig. 1. The majority of the EST–SSR loci had four to six alleles in contrast with the seven to nine observed for gSSR loci. Approximately, 35% of gSSRs

Table 3: Relation of the 51 EST–SSRs with their motif and expected EST homology

EST–SSR[1]	Motif	EST Homology[2]
SCA03	(AC)20	(AC006234) (1–4)-beta-mannan endohydrolase
SCA04	(TG)9	(AL391711) putative protein
SCA06	(TG)12	q42976|nlt4_orysa non-specific lipid-transfer protein 4 precursor
SCA07	(AC)7	S65572 pattern-formation protein GNOM
SCA10	(GA)7	No hits found
SCA12	(TA)29	No hits found
SCA14	(AG)23	No hits found
SCA16	(AG)25	(AB009048) gene_id:K15E6.9 unknown protein
SCA17	(AG)32	(AP002482) ESTs D41739(S4522)
SCA19	(CG)9	(AC006592) putative peroxidase
SCA22	(TA)30	No hits found
SCA26	(TG)11	T05613 hypothetical protein F9D16.
SCB01	(CGG)7	(AF034945) glycine-rich RNA binding protein
SCB02	(AGA)5	(AC082644) putative centromere/microtubule binding protein
SCB06	(GCC)8	(AC018363) putative protein phosphatase-2C (PP2C)
SCB07	(CGA)8	(AB013392) pyrophosphate-dependent phosphofructo-1-kinase-like protein
SCB08	(GGT)6	(AF315811) RNA-binding protein
SCB12	(GCC)5	(AB007649) gene_id:MLE2.11
SCB13	(ACG)5	No hits found
SCB14	(CGT)8	(D28861) RNA binding protein, RZ-1
SCB16	(CGG)5	(AP001168) Similar to mRNA for DREB1A (AB007787
SCB17	(GTC)6	(AP002524) contains ESTs AU032708(S13763), D47951(S13763)
SCB25	(AAT)7	(AC003673) putative protein kinase
SCB27	(AGT)6	(AL163815) putative protein
SCB35	(CCG)5	(X67324) MFS18
SCB37	(GCC)6	(L14063) O-methyltransferase
SCB38	(GTC)8	No hits found
SCB39	(CGC)8	(AB023482) ESTs AU058081(E30812)
SCB40	(GCC)5	(X97022) lamin
SCB41	(CGA)8	(AL049525) putative protein
SCB43	(CGA)9	(AP002743) putative pyrophosphate-dependent phosphofructo-1-kinase
SCB45	(GTA)7	(AC004665) putative phosphomannomutase
SCB47	(CGC)5	(AP002746) putative zinc finger transcription factor
SCB48	(CTA)5	(AC004665) hypothetical protein
SCB49	(GCC)5	(AF310215) glycine-rich RNA-binding protein
SCB51	(TCA)6	(AC008261) hypothetical protein
SCB52	(CCG)5	No hits found
SCB53	(GCG)5	No hits found
SCC01	(GATA)14	(Z97022) cysteine proteinase
SCC02	(CGGC)5	(AF130975) plasma membrane intrinsic protein
SCC03	(CCAC)4	(AF241166) MAP kinase MAPK2
SCC04	(GGAT)6	(AB019235) contains similarity to DNA-binding protein~gene_id:MMI9.9
SCC05	(TGCT)4	(X75670) cytochrome b5
SCC07	(CTAG)3	(AB015475) contains similarity to gibberellin-stimulated transcript 1 like protein~gene_id: MMN10.7
SCC10	(AGGC)3	No hits found
SCC13	(TGCG)5	(AC084320) putative actin-depolymerizing factor
SCC15	(TAAT)13	No hits found
SCC16	(TCTA)4	(M57249) phospholipid transfer protein
SCC17	(ACGC)5	(AF060198) PsbY precursor; putative photosytem II peptide
SCC18	(AGGA)3	(AB009053) permease 1
SCC19	(AGCC)5	No hits found

[1]EST–SSR primer sequences are under a confidentially agreement. Contact A. P. de Souza (anete@unicamp.br).
[2]EST homology: annotation of the best homologue identified by BLASTX (available in SUCEST database).

had PIC values around 0.90, in contrast to 15% of the EST–SSRs.

Discrimination power

The discrimination power (*D*) was high for the majority of EST–SSRs (with an average value of 0.80 reaching the maximum value (1.00) for EST–SSRs SCB07, SCB14, SCC01 and SCC17 (Table 2). For gSSRs, it ranged from 0.11 (SMC80CS) to 1.0 (SMC863CG, SMC21AS, SMC119CG, CIR36, CIR21, CIR56, CIR58, CIR12, SMC1235FL, SMC448MS, SMC1047HA, CIR28, SMC31CUQ, SMC2017FL, CIR32, CIR1) with a mean value of 0.89. Although parameters such as number of alleles and PIC were less informative for the EST–SSR loci, no differences were found between discrimination power averages for the two types of SSR (Fig. 1). In fact, for this parameter, the mean value obtained by the two types of SSRs did not differ considering the significant confidence interval of 95% (Table 2).

Genetic similarity

Genetic similarities (GS) based on Dice coefficients were obtained for all possible 153 pairwise comparisons calculated with 305 polymorphic markers amplified by 51 EST–SSRs.

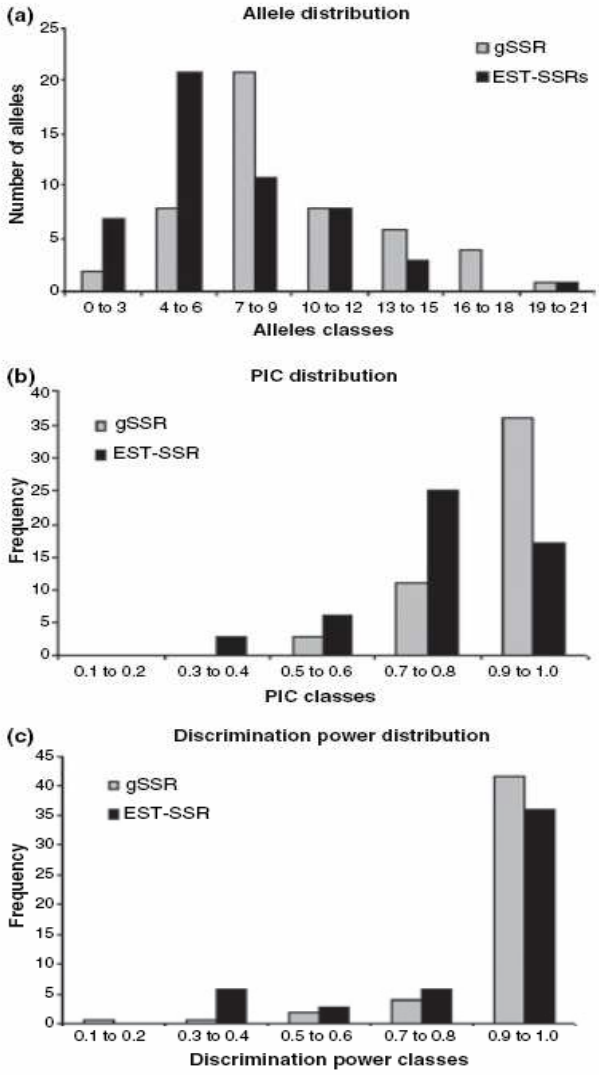


Fig. 1: Comparative distributions of (a) number of alleles, (b) Polymorphism information content (PIC) and (c) discriminatory power (d) between gSSRs and EST–SSRs

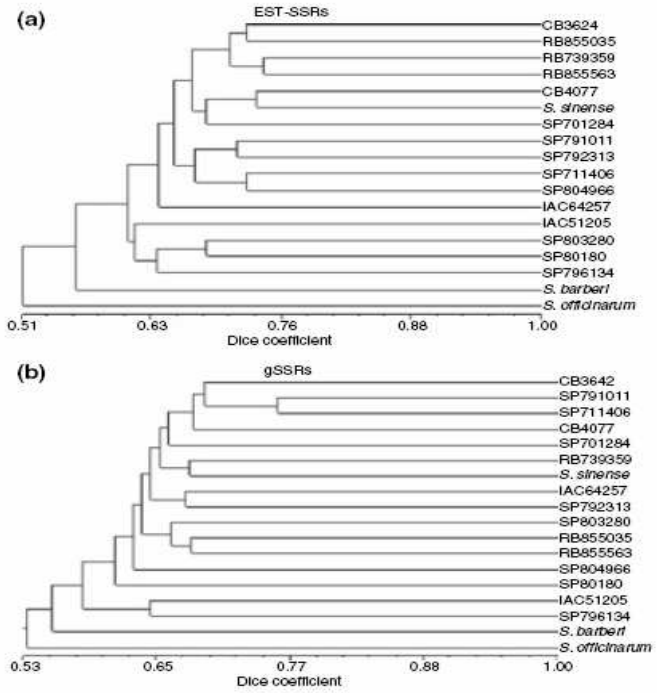


Fig. 2: Dendrogram base on the Dice similarity coefficient and UPGMA clustering method. (a) Calculated with EST–SSRs data. (b) Calculated with gSSRs data

These 153 pairwise combinations gave an average GS value of 0.62. Genetic similarity values varied from 0.46 (SP791011 and IJ76-314) to 0.74 (RB855563 and RB739359). Among the species evaluated, *S. barberi* was more related to *S. sinense* (0.57) than *S. officinarum* (0.52) while *S. barberi* vs. *S. officinarum* was more distant (0.48). The mean GS value among these species was 0.52. A genetic similarity value of 0.69 was found between SP80180 and SP804966. These two sugarcane elite clones were selected from the CTC (Centro de Tecnologia Canavieira, Piracicaba, São Paulo, Brazil) breeding programme and are being used in the construction of a genetic map (Garcia et al. 2006) and for the mapping of QTL.

For gSSR, GS estimates were based on 430 polymorphic markers derived from 50 loci. The highest GS-value (0.75) was observed between the SP791011 and SP711406 clones that are half sibs (Table 1). The most genetically distant clones were *S. officinarum* and IAC51205; *S. officinarum* and SP804966 showed a GS-value of 0.49. On average, the GS-value obtained with gSSR was 0.61. The GS between *S. officinarum* and *S. sinense* was almost the same (0.58) as that observed between *S. barberi* and *S. sinense* (0.57) while *S. barberi* and *S. officinarum* was the most genetically distant (0.51). The mean GS-value among these species was 0.55. For the SP80180 and SP804966 mapping parents, the GS-value was 0.61 and equivalent to the average.

Based on bootstrap analysis (data not shown), the mean coefficient variation (CV) value obtained with 430 gSSR and 305 EST–SSR polymorphic markers was 4.73% and 5.42%, respectively. This implies that the number of polymorphic markers used were sufficient to produce reliable GS estimates. Interestingly, the number of EST–SSR polymorphic markers (80 markers) needed to reach a 10% CV-value that is usually recommended, was lower than that required for gSSRs (90 markers).

Although differences in pairwise GS values were observed for each type of marker (gSSRs and EST–SSRs), the correlation between GS was high ($r = 0.71$/P = 0.99) and significant. This result shows that EST–SSRs can be successfully used for sugarcane genetic relationship analysis.

Cluster analysis

The genetic relationships among the sugarcane clones evaluated for each type of SSR are represented by dendrograms in Fig. 2. The cophenetic values were high and significant for both types of SSR ($r_{EST-SSR} = 0.85$; P = 0.99; $t = 4.4$ and $r_{gSSR} = 0.87$; P = 0.99; $t = 4.6$) indicating a good fit with genetic similarity values. The clones were not separated into clear distinct groups but into a large number of sub-clusters

demonstrating a weak form of structuration among the sugarcane varieties evaluated.

Discussion

In the present work, the polymorphisms obtained with a subset of 50 EST–SSRs were compared with those generated by 50 gSSR, considering parameters such as number of alleles, polymorphism information content, discrimination power and their ability to establish genetic relationships among sugarcane clones.

The results demonstrate that EST–SSR loci are able to generate sufficient polymorphism to ensure their use as molecular markers in sugarcane genetic investigations. The number of alleles and PIC values revealed by the 51 EST–SSRs were close to those reported in sugarcane polymorphism analyses using a few gSSR (Cordeiro et al. 2003, Pan et al. 2003a).

The power of SSRs to distinguish sugarcane varieties have been shown in several studies (Jannoo et al. 2001, Pan et al. 2003b). In fact, the polyploid nature of sugarcane makes SSR markers ideal for fingerprinting (Cordeiro 2001). The lack of difference between discrimination power averages of the two types of SSRs confirms the ability of EST–SSR to generate unique sugarcane fingerprints.

The high value of the average GS (0.61) exhibited with SSRs (EST–SSRs and gSSRs) reveals the narrow genetic base of sugarcane clones, attributed to crosses involving common parents (Table 1). As has been seen, the exchange of materials among breeding programmes has lead to the repeated use of varieties derived from good performing crosses (Nair et al. 2002, Arro 2005) as parents. This is the case of POJ2878, introduced into most sugarcane research stations of the world (Lu et al. 1994), which was used as the progenitor parents of varieties derived from several different breeding programmes (Table 1). Either with gSSRs or EST–SSRs, the highest GS values were observed among related sugarcane clones. This was the case of SP804966 derived from SP711406 and clones SP791011 vs. SP711406, CB3624 vs. CB4077, CB4077 vs. IAC51205 and CB3624 vs. IAC51205 that are half sibs.

The dendrograms generated with each type of SSR did not cluster sugarcane clones into well-defined groups. This result is somewhat expected in sugarcane cluster analysis as a consequence of the high levels of ploidy and heterozygosity which retains a large proportion of alleles derived from the parents involved in the initial interspecific crosses (Lu et al. 1994, Lima et al. 2001) allied to the small number of *S. spontaneum* parents used in the ancestry of most modern varieties. Dendrogram differences observed between EST–SSR and gSSR are probably due to the number of markers and the different portion of genome sampled by each SSR type. Nevertheless, both dendrograms are in agreement with pedigree information covering two generations (Table 1), reinforcing that GS measures derived from SSR markers can be used for planning crosses between divergent sugarcane clones.

Out of the six species that encompass the genus *Saccharum* (*S. officinarum*, *S. barberi*, *S. sinense*, *S. spontaneum* L., *S. robustum*, and *S. edule* Hassk.), three were investigated in this study (*S. officinarum*, *S. barberi*, *S. sinense*) for polymorphism comparisons between EST–SSRs and gSSRs. Because only one accession was used to represent each species, it is difficult to draw well-defined genetic relationships among them. However, for both types of SSR, *S. officinarum* (IJ76-314) formed a single group that was closer to *S. barberi* (Ganda Cheni) and was clearly separated from *S. sinense* (Maneria). This same result was observed using 1121 polymorphic markers derived from AFLP (Lima et al. 2001).

The low levels of polymorphism commonly displayed with molecular markers derived from gene sequences can limit their wide-spread use in genetic analysis. However, compared with other gene-targeted marker systems, the degree of genetic diversity sampled by EST–SSRs seemed promising for generating reliable genetic parameter estimates for assessment of sugarcane variability. As reported by Arro (2005), the cluster patterns generated by TRAP (Target Region Amplification Polymorphism) markers in 63 sugarcane clones did not reflect their pedigree relationships, as half-siblings, and even full-siblings failed to group in the same cluster. On the contrary, as SSRs have high mutation rates and therefore may show high variation between individuals, some degree of polymorphism is more expected, even in conserved regions.

Finally, the fact that most of the EST–SSRs have homology to genes (Table 3) counterbalances their relatively low levels of polymorphism in relation to those derived from anonymous markers such as gSSRs. Thus, EST–SSRs will contribute to the knowledge of genetic diversity of sugarcane, extending it to the functional level.

Acknowledgements

The authors thank Dr Marcos Guimarães de Andrade Landell for supplying pedigree information with the Caiana software. The present research was financed by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 02/01167-1) and CTC (Centro de Tecnologia Canavieira). LRP received postdoctoral fellowship from FAPESP (01/14656) and KMO received a doctorate fellowship from FAPESP (8; 02/00197-4). APS and AAFG received a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Aitken, K. S., P. A. Jackson, and C. L. A. Mcintyre, 2005: A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor. Appl. Genet. **110,** 789—801.

Arro, J., 2005: Genetic Diversity Among Sugarcane Clones Using Target Region Amplification Polymorphism (Trap) Markers and Pedigree Relationships. MS Thesis, Louisiana State University, Louisiana.

Casu, R. S., J. M. Manners, G. D. Bonnett, P. A. Jackson, C. L. McInTyre, R. Dunne, S. C. Chapman, A. L. Rae, and C. P. L. Grof, 2005: Genomics approaches for identification of genes determining important traits in sugarcane. Field Crops Res. **92,** 137—147.

Cato, S. A., R. C. Gardner, J. Kent, and T. E. Richardson, 2001: A rapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. **102,** 196—306.

Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCouch, W. D. Park, N. Ayres, and S. Cartinhour, 2000: Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100,** 713—722.

Cordeiro, G. M., 2001: Molecular marker systems for sugarcane germplasm analysis. In: R. J. Henry (ed.), Plant Genotyping – The DNA Fingerprinting of Plants, 129—146. CABI Publishing, Wallingford.

Cordeiro, G. M., T. L. Maguire, R. J. Henry, and K. J. Edwards, 1999: Optimisation of a microsatellite enrichment technique in *Saccharum* spp. Plant Mol. Biol. Rep. **17,** 225—229.

Cordeiro, G. M., G. O. Taylor, and R. J. Henry, 2000: Characterization of microsatellite markers from sugarcane (*Saccharum* spp.) a highly polyploid species. Plant Sci. **155,** 161—168.

Cordeiro, G. M., Y. B. Pan, and R. H. Henry, 2003: Sugarcane microsatellites for assessment of genetic diversity in sugarcane germplasm. Plant Sci. **165,** 181—189.

Creste, S., A. Tulmann Neto, and A. Figueira, 2002: Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gel by silver staining. Plant Mol. Biol. Rep. **19,** 299—306.

Dice, L. R., 1945: Measures of the amount of ecologic association between species. Ecology **26,** 297—302.

Garcia, A. A. F., E. A. Kido, A. N. Meza, H. M. B. Souza, L. R. Pinto, M. M. Pastina, C. S. Leite, J. A. G. da Silva, E. C. Ulian, A. V. O. Figueira, and A. P. Souza, 2006: Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross based on a maximum-likelihood approach. Theor. Appl. Genet. **112,** 298—314.

Hoarau, J. Y., G. M. Souza, A. D'Hont, M. Menossi, L. R. Pinto, A. P. Souza, L. Grivet, C. F. M. Menck, E. C. Vlian, M. Vincentz. Sugarcane, a tropical crop with a highly complex genome. In: J.F. Morot-Gaudry, P. Lea and J.F. Briat. (Org.). Plant genomics. Sciences Publishers, 2006, in press.

Hoisington, D., M. Khairallah, and D. González-De-León, 1994: Laboratory Protocols: CIMMYT. Applied Molecular Genetics Laboratory, Mexico, DF.

Jannoo, N., L. Forget, and A. Dookun, 2001: Contribution of microsatellites to sugarcane breeding program in Mauritius. Proc. Intl Soc. Sugar Cane Technol. **24,** 637—639.

Jarne, P., and P. J. L. Lagoda, 1996: Microsatellites, from molecules to populations and back. Trends Mol. Ecol. **11,** 424—429.

Lima, M. L. A., A. A. F. Garcia, K. M. Oliveira, S. Matsuoka, H. Arizono, C. L. Jr de Souza, and A. P. de Souza, 2001: Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor. Appl. Genet. **104,** 30—38.

Lu, Y. H., A. D'Hont, F. Paulet, L. Grivet, M. Arnaud, and J. C. Glaszmann, 1994: Molecular diversity and genome structure in modern sugarcane varieties. Euphytica **78,** 217—226.

Ma, H. M., S. Schulze, S. Lee, M. Yang, E. Mirkov, J. Irvine, P. Moore, and A. Paterson, 2004: An EST survey of the sugarcane transcriptone. Theor. Appl. Genet. **108,** 851—863.

Meyer, A. S., A. A. F., Garcia, and A. P. Souza, 2004: Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L). Genet. Mol. Biol. **27,** 83—91.

Nair, N. V., A. Selvi, T. V. Sreenivasan, and K. N. Pushpalatha, 2002: Molecular diversity in Indian sugarcane cultivars as revealed by randomly amplified DNA polymorphisms. Euphytica **127,** 219—225.

Pan, Y. B., G. M. Cordeiro, E. P. Jr Richard, and R. J. Henry, 2003a: Molecular genotyping of sugarcane clones with microsatellite DNA markers. Maydica **48,** 319—329.

Pan, Y., J. D. Miller, R. J. Schnell, E. P. Richard Jr, and Q. Wei, 2003: Application of Microsatellite and RAPD fingerprints in Florida Sugarcane Variety Program [abstract] . Plant and Animal Genome Abstract XI Abstract Book **W189,** p.43.

Pinto, L. R., K. M. Oliveira, E. C. Ulian, A. A. Garcia, and A. P. Souza, 2004: Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome **47,** 795—804.

Rohlf, F. J., 1993: NTSYS-pc Numerical Taxonomy and Multivariate Analysis System, Version 2.0. Exeter Publications, New York.

Scott, K. D., P. Eggler, G. Seaton, M. Rossetto, E. M. Ablett, L. S. Lee, and R. J. Henry, 2000: Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100,** 723—726.

Tessier, C., J. David, P. This, J. M. Boursiquot, and A. Charrier, 1999: Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. **98,** 171—177.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner, 2003: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106,** 411—422.

6 Artigo III

"Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross"

K.M. Oliveira, L.R. Pinto, T.G. Marconi, G.R.A. Margarido, M.M. Pastina, L.H.M. Teixeira, A. Figueira, E.C. Ulian, A.A.F. Garcia, A.P. Souza

Artigo submetido à revista

Molecular Breeding

Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross

K.M. Oliveira[1], L.R. Pinto[2], T.G. Marconi[1], G.R.A. Margarido[3], M.M. Pastina[3], L.H.M. Teixeira[1], A. Figueira[4], E.C. Ulian[5], A.A.F. Garcia[2], A.P. Souza[1]

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

[2]Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana - IAC/Apta, Anel Viário Contorno Sul, Km 321, CP 206, CEP 14.001-970, Ribeirão Preto-SP, Brasil.

[3]Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba-SP, Brasil.

[4]Centro de Energia Nuclear na Agricultura (CENA) – Universidade de São Paulo, Avenida Centenário, 303, CP96, SP 13400-970, Piracicaba-SP, Brasil.

[5]Centro de Tecnologia Canavieira -CTC, Caixa Postal 162, 13400-970, Piracicaba, São Paulo, Brasil.

*To whom correspondence should be addressed: Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, 13083-085, C.P. 6010, Campinas/SP, Brasil.

E-mail: anete@unicamp.br

Fax: (55-19) 3788-1089

Phone: (55-19) 3788-1132 or 3788-1156

Abstract

Breeding in sugarcane (*Saccharum* spp.) is time-consuming due to the highly heterozygous nature of this tropical crop, coupled with interespecific origin of present-day cultivars. Commercial sugarcane plants are the result of a limited series of crosses and backcrosses derived from the domesticated species *S. officinarum* (2n=80) and the wild species *S. spontaneum* (2n=40-128), presenting the most complex genome of any crop. Their genome has been unraveled because of the development of molecular markers and genetic linkage maps. The growing availability of EST (*Expressed Sequence Tags*) sequences provides a potentially valuable source of new DNA markers. We have examined the SUCEST (*Sugarcane expressed Sequence Tag Project*) database and developed EST-SSRs. Thus to enhance the resolution of an existing linkage map and to identify putative functional polymorphic gene loci in a sugarcane commercial cross, 149 EST-derived SSRs and 10 EST-derived RFLPs were screened among the 100 F1 of the SP80-180 x SP80-4966 mapping population. Evaluation of these ESTs resulted in 576 EST-SSR and 47 EST-RFLP polymorphic markers. With the AFLP, SSR and RFLP markers already analyzed in the previous map, we generated 2303 polymorphic markers, of which 1669 (72.5%) were single-dose (SD – segregated 1:1 and 3:1) markers. Map analyses were carried out using JoinMap and OneMap algorithm. Of these 1669 SD markers, 664 (40%) were scattered onto 192 co-segregation groups (CGs) with a total estimated length of 6261.1 cM. Eighty-five CGs from the previous map were conserved, 30 CGs were modified due to EST-derived marker integration and 77 new CGs were formed. Using SSR and RFLP, both genomic and EST-derived, we succeeded in forming 120 of the 192 CGs into fourteen putative homology groups (HGs). The genome coverage was significantly increased in the

current map. The 149 EST-SSRs and 10 EST-RFLP were subjected to BLASTX search against all GenBank databases, of which putative function was assigned to 113 EST-SSRs and 6 EST-RFLPs based on high nucleotide homology to previously studied genes. Some important enzymes from sugarcane metabolism were identified. The integration of EST-derived markers improves the map, makes it possible to consider additional fine mapping of the genome, and provides the means for developing 'perfect markers' associated with key QTLs and for comparative genomics. In summary, we have constructed the first genetic linkage map of sugarcane that is predominantly populated by functionally associated markers.

Keywords Sugarcane, Functional map, ESTs, Polyploid, Mapping

Introduction

Modern sugarcane cultivars have a highly polyploid, aneuploid genome with chromosome number ranging from 100 to 130 (Roach 1969). These modern cultivars are interspecific hybrids derived essentially from crosses between *Saccharum officinarum* (2n=80, x=10), a species that has stalks with high sugar content, and *Saccharum spontaneum* (2n=40-128, x=8), a wild, vigorous species that is resistant to several sugarcane diseases. In order to reduce the contribution of *S. spontaneum* and to enhance the sugar content of the hybrids, a series of backcrosses to *S. officinarum* was undergone in early breeding programs, in a process known as nobilization. Hence, modern cultivar genomes are predominantly derived from *S. officinarum*, with only 15-25% of their chromosomes contributed by *S. spontaneum*, as has been shown using genomic *in situ* hybridization (GISH) (D'Hont et al. 1996; Piperidis et al. 2001) and by molecular marker approaches (Grivet et al. 1996; Hoarau et al. 2001). Because of their high polyploidy and interespecific origin, these cultivars produce progenies with aneuploid chromosome numbers.

Due to the complexity of the sugarcane genome, which appears difficult to be deciphered, sugarcane genetics had received little attention from plant scientists. However, a vast array of genomic tools became available and new opportunities have been opened to refine our understanding of the genetic architecture of sugarcane and to explore its functional system (Hoarau et al. 2006). The development of ubiquitous genetic marker technologies in conjunction with increased computational capability has provided abundant resources for whole-genome linkage analysis.

Among the DNA marker systems available, microsatellite or simple sequence repeat (SSR) is a preferred technology for plant genome analysis (Morgante and Olivieri 1993).

These markers are characterized by their simplicity, abundance, variation, co-dominance and multi-alleles among genomes (Powell et al. 1996). Recent studies revealed that gene transcripts can also contain SSRs, and the abundance of ESTs (*Expressed Sequence Tags*) has become an attractive potential source of microsatellite markers (Kantley et al. 2002). Indeed, electronic searches in genome databases may represent a simple, fast and economical way to obtain SSRs. With the spread of EST projects, SSRs have been developed for many plant species through SSR mining in EST sequences, including *Triticum aestivum* L. (Gupta et al. 2003; Gao et al. 2004; Nicot et al. 2004), *Medicago truncatula* (Eujayl et al. 2002), *Vitis nivifera* (Decroocq et al. 2003), *Grossypium* (Saha et al. 2003; Han et al. 2004; Qureshi et al. 2004; Han et al. 2006) and *Citrus* (Chen et al. 2006). In the past few years, several projects for the sequencing of sugarcane ESTs have been initiated in South Africa (Carson et al. 2000), Brazil (http://sucest.lad.ic.unicamp.br/en/) and Australia (Casu et al. 2001), allowing the development of EST-SSRs for *Saccharum* (Cordeiro et al. 2001; Pinto et al. 2004). The EST collections provide the opportunity for intensive development of functionally associated markers. ESTs can provide direct mapping of genes with known function allowing candidate genes involved in important pathways to be directly evaluated for associations with important agronomic traits (Cato et al. 2001; Ma et al. 2004). Moreover, their presence in conserved transcribed regions makes them more transferable across species and to closely related genera, an aspect that would increase their value in a breeding programme.

The Sugarcane Expressed Sequence Tag Project (SUCEST) produced 237.954 EST, from 26 cDNA libraries constructed from several organs and tissues sampled at different developmental stages, assembled in 43.000 clusters (Vettore et al. 2001), giving rise to the

most complete dbEST for sugarcane (http://sucest.lad.ic.unicamp.br/en/). These 43.000 transcript products are believed to represent 90% of the sugarcane genes with over 30% corresponding to new genes (Vettore et al. 2003). Pinto *et al.* (2004) mined EST-SSRs in this bank and identified 2005 clusters containing SSR that in turns represents around 5% of the entire cluster population of the bank.

The first-generation molecular marker-based genetic maps for agronomically important plant species have been largely based on anonymous genetic markers. Due to the fact that polyploidy dictates particular constraints for mapping, the development of a low-density genetic map for sugarcane requires much more work than for a diploid species. When polyploidy is high and pairing is polysomic or irregular, alleles presented as a single copy are much more informative for genetic maps construction than any other (Wu et al. 1992). Based on this method, partial genetic maps have been produced for *S. spontaneum* (Ming et al. 1998; Al-Janabi et al. 1993; da Silva et al. 1993; Ming et al. 2002), *S. officinarum* (Mudge et al. 1996; Guimarães et al. 1999) and modern cultivars (Grivet et al. 1996; D'Hont et al. 1994; Hoarau et al. 2001; Aitken et al. 2005; Raboin et al. 2006; Garcia et al. 2006). However, coverage of the genomes surveyed is still incomplete and none of the published maps of sugarcane are saturated yet. Although the current set of genetic markers provides the means to anchor maps across different pedigrees and to establish linkage with QTL for agronomic traits, they are not in general saturated enough to narrow the search to small areas in the genome.

In this paper, we report the development of the first functionally associated marker-based genetic linkage map for sugarcane, using EST-RFLP and EST-SSR marker technologies. These EST-derived markers were integrated into an existing framework of a genetic map for a commercial cross (Garcia et al. 2006), increasing the marker density and

coverage of the existing map. This gene-based map will enable the precision of QTL mapping and would be useful for sugarcane breeding programs.

Materials and Methods

Plant Material

Genetic mapping was conducted using a population made of 100 random chosen individuals obtained by the cross between the commercial cultivars SP80-180 [B3337 x polycross] and SP80-4966 [SP71-1406 x polycross]. SP80-180, the female parent, has lower sucrose content and high stalk production, whereas SP80-4966, the male parent, has higher sucrose and lower stalk production. Both parents and population were developed at the Experimental Station of the Centro de Tecnologia Canavieira-CTC (Camamu-BA, Brazil). Total genomic DNA of mapping progenies was extracted from 300mg of powdery lyophilized young leave tissue according to the method described by Hoisington et al. (1994), with minor modifications.

Genotypic data and marker notation

Five types of markers were used to genotype the parents and the 100 progeny individuals. RFLP, AFLP and SSR markers had already been generated and coded as described in detail by Garcia et al. (2006). EST-SSR and EST-RFLP markers were generated as follows:

EST-SSR

From the 372 EST-SSRs that were developed from the SUCEST db (Pinto et al. 2004; Pinto et al. 2006), 149 were selected for genetic mapping in the SP80-180 x SP80-4966 population. Part of these EST-SSRs were screened in the mapping population according to Pinto et al. (2004). Amplification products were separated by electrophoresis on 6% denatured polyacrilamide gels (*w/v*) using a 25bp ladder as a size standard and silver stained according to Creste et al. (2002). The other part of these EST-SSRs was performed following the protocol described in Raboin et al. (2006). Then, the samples were resolved on 5% denatured polyacrilamide gels (*w/v*), dried and exposed to X-ray film (Fuji RX) for 4 days. In these cases, a 10bp ladder was used as a size standard.

The nomenclature of these markers was: EST followed by a letter and a two digit code number that represent the motif type and EST-SSR number, respectively (A: for di-; B: for tri-; C: for tetranucleotide primer pairs); followed by a number referring to the amplified allele and a letter to denote parental polymorphism origin. Parental polymorphism origin was designated according to the cross type of marker locus, following notation of Wu et al. (2002) and detailed in the previous map of this population (Garcia et al. 2006). "D1" corresponds to marker locus heterozygous in SP80-180 and homozygous in SP80-4966, "D2" represents marker locus heterozygous in SP8049-66 and homozygous in SP80-180, while "C" indicates marker locus heterozygous in both parents.

EST-RFLP

20 microgrammes of genomic DNA were individually digested with 10 restriction enzymes (*BamH I, Bgl II, Dra I, EcoR I, EcoR V, Hind III, Pst I, Sal I, Sst I e Xba I*) in order to prepare the membranes that were used with EST probes. Restriction fragments were loaded

on 0.8% (*w/v*) agarose gels that were submitted to electrophoresis in a TAE buffer (40mM Tris acetate, pH 8.0, 2mM EDTA) at 25V for 22h and transferred to nylon membranes (Hybond-N^+, Amersham, UK). 40 EST probes deriving from SUCEST (*Sugarcane EST Project*, FAPESP), obtained from BCC Center (Brazilian Core Collection Center, Jaboticabal, SP, Brazil), were radioactively labeled with α-$dCTP^{32}$ (Feinberg and Volgelstein 1983). Hybridizations were performed in a HYB solution (0.5 M Na_2PO_4 pH 7.2, 1% BSA, 7% SDS, 100 μg/mL sheared herring sperm DNA) at 65^0C for 18-24 h. The membranes were washed once for 20 min at 65^0C in each of the following solutions: solution I (2 X SSC; 5%, SDS), solution II (1 X SSC; 5% SDS 5%) and solution III (0.5X SSC; 5% SDS) and placed on an X-Omat (Kodak) film for at least 7 days at -80^0C.

Probes from expressed sequences (EST-RFLP) evaluated in the population were named Est plus a number according to the corresponded gene. Markers from the same EST probe were identified by probe code followed by a letter indicating a decrease in the size of the fragment and a letter to denote parental polymorphism origin (D1, D2, C).

Marker scoring

All segregating bands that were distinct and unambiguous were scored independently as a dominant marker, based on the presence (1) or absence (0) in the progeny of the mapping population. Since sugarcane is highly polyploid, only single-dose markers (Wu et al. 1992) were used for map analysis. Each marker was tested against the expected ratios using chi-square test (χ^2) for single dose markers, considering the ratios 1:1 and 3:1. Thus, marker segregation types were identified for deviation from the segregation ratios of 1:1 and 3:1,

expected for markers in single-dose in only one of the parents and markers in single-dose in both parents (referred here as double single-dose markers), respectively. All loci with strong deviation from expected proportions were discarded after controlling type I error for multiple tests (Garcia et al. 2006). The molecular data from EST-SSR and EST-RFLP single-dose markers were then added to data from AFLP, SSR and RFLP markers used to construct the previous genetic map developed by Garcia et al (2006), which was used as a framework.

Linkage analysis and map construction

We basically made the analysis following the procedures proposed by Wu et al. (2002) and implemented in the OneMap software (Garcia et al. 2006), using the previous map as a framework. As pointed out by Garcia et al. (2006), this procedure has several advantages over the ones used by the JoinMap software (Stam 1993). In short, two-point analysis was carried out using minimum LOD Score threshold of 5 and 35 cM (Kosambi) for the recombination fraction. The clusters of linked markers (co-segregation groups, CGs) were then ordered using the Rapid Chain Delineation algorithm (Doerge 1996) and the final position of markers was refined using three-point analysis based on the likelihood. Markers that could not be correctly positioned were set as accessories. Then, the ordered CGs were assembled into homology groups (HGs) based on common RFLP, SSR, EST-RFLP or EST-SSR markers linked in coupling or repulsion phase.

Results

EST-derived markers evaluation

EST-SSR

A total of 372 EST-SSRs were developed and screened against the parents and 6 F1 progeny individuals from the F1 mapping population (SP80-180 x SP8049-66) in order to determine polymorphism levels. This includes the EST-SSRs already described by Pinto et al. (2004, 2006). Of these, 149 polymorphic loci, with at least two band differences and segregating in the population sample as dominant genetic markers, were selected and scored across the population. These 149 EST-SSRs produced 576 polymorphic bands of which 332 (57.6%) were single dose markers (1:1) (Table 1). Out of these 332 markers, 161 (48.5%) segregated in the SP80-180 gametes (D1 –cross type "*ao* x *oo"*) and 171 (51.5%) in the SP80-4966 gametes (D2 – cross type "*oo* x *ao"*) (Table 2). An additional 183 (31.8%) EST-SSR double single-dose markers (C – cross type "*ao* x *ao"*) were added to the map (Table 1). The number of segregating markers identified with each EST-SSR ranged from 1 to 10 with an average of 3.5. Neither the markers with two or higher allele dosage nor the markers that did not fit expected segregation ratios (1:1 or 3:1) were exploited in this work.

EST-RFLP

Out of 36 EST probe/enzyme combinations, 10 were selected and screened in the mapping population. A total of 47 polymorphic markers were generated, of which, 37 (78.7%) were segregating in expected ratios (1:1 and 3:1) (Table 1). Of these 37 markers, 13 (35.1%) were polymorphic in SP80-180 (D1 – cross type "*ao* x *oo"*), 14 (37.8%) were polymorphic

in SP80-4966 (D2 – cross type "*oo* x *ao*") and 10 (27.1%) were polymorphic in both parental genotypes (C – cross type "*ao* x *ao*") (Table 2). The 10 remaining markers had distorted segregation and were discarded.

Table 1 Overall markers screened across progeny of SP80-180 x SP80-4966 cross: number of polymorphic markers and single- and double single-dose markers.

Markers		RFLP		SSR		
	AFLP	Genomic	EST	Genomic	EST	Total
Number of markers evaluated	1104	221	47	355	576	2303
Number of polymorphic markers between parents	304	112	35	190	385	1026
Number of monomorphic markers between parents	800	109	12	165	191	1277
Single-dose markers (1:1)	212	100	27	129	332	800
Double-single-dose markers (3:1)	507	88	10	81	183	869
Total number (1:1 and 3:1)	719	188	37	210	515	1669
Number of markers with distorted segregation	385	33	10	145	61	634

Modified from Garcia *et al*. (2006)

Integration of EST-derived markers into the previous map

The 515 EST-SSR and 37 EST-RFLP markers were integrated into a framework map of the SP80-180 x SP8049-66 population (Garcia et al. 2006), in combination with AFLP, RFLP and SSR data already mapped. Hence, a total of 1669 segregating bands were scored and classified as being from SP80-180 (D1 type) or SP80-4966 (D2 type), or as being a common marker (present in both parents) (C type). Polymorphic markers included 719 AFLP markers generated by 23 primer pair combinations, 210 SSR markers generated by 52 primer pairs, 188 RFLP markers generated by 55 probe/enzyme combinations and the EST-derived markers (Table 2).

Table 2 Distribution of the different marker types according to their cross type

Cross type	Nº of markers					
	RFLP		SSR			
	Genomic	EST	Genomic	EST-SSR	AFLP	Total
D1 (*ao* x *oo*)	47	13	55	161	80	356 (21.3)
D2 (*oo* x *ao*)	53	14	74	171	132	444 (26.6)
C (*ao* x *ao*)	88	10	81	183	507	869 (52.1)
Total	188	37	210	515	719	1669
Cross type	Nº of linked markers					
	RFLP		SSR			
	Genomic[a]	EST [a]	Genomic [a]	EST[a]	AFLP[a]	Total [b]
D1 (*ao* x *oo*)	19 (1.1)	5 (0.3)	19 (1.1)	97 (5.8)	39 (2.3)	179 (27.0)
D2 (*oo* x *ao*)	16 (1.0)	4 (0.2)	30 (1.8)	102 (6.1)	76 (4.6)	228 (34.4)
C (*ao* x *ao*)	24 (1.4)	1 (0.1)	24 (1.4)	76 (4.6)	132 (7.8)	257 (38.6)
Total	59	10	73	275	247	664

Modified from Garcia *et al*. (2006)

[a] Percentage of linked markers in relation to the total number of markers available for mapping (1669)

[b] Percentage in relation of the total number of linked markers (664)

In order to avoid false positives when testing for linkage, co-segregation groups were built up at a stringent LOD score threshold of 5. This high value was chosen with respect to the high number of markers and the high number of expected CGs, which is supposed to be over 100 CGs due to the high number of chromosomes in sugarcane species. As a result, 664 (40%) markers formed 192 CGs. Of these 664 markers, 179 (27%) were polymorphic on SP80-180 (D1) and linked into the map resulting in 62 CGs; 228 (34.4%) were polymorphic on SP80-4966 (D2) and linked, forming 78 CGs. The remaining 257 (38.6%) linked markers were common from both parents (C) and were used as bridges for combining the information of markers (D1 and D2) into a single map (Table 2). The CGs varied in length from 0 cM to 149.0 cM, with a cumulative length of 6261.1 cM and an average length of 35.2cM, with mean distance between markers of 9.4 cM. An irregular

distribution of markers along the chromosomes was observed, with an average of 3.6 markers per linkage group (Figure 1). The 275 EST-SSR and 10 EST-RFLP markers were incorporated into 30 CGs of the previously constructed genetic map and formed 77 new CGs, corresponding to 43% of the linked markers. From these new groups, 39 (51.3%) assembled only EST-derived markers. The remaining 85 CGs were already present in the previous map.

Assignment of HGs using SSR and RFLP markers

Homologous linkage groups (HGs) were identified using common allelic bridges, provided by SSR and RFLP markers, both genomic and EST-based, due to their co-dominance inheritance. Of the 192 CGs, 120 could be regrouped into 14 putative HGs, each containing between 2 and 24 CGs. Markers were not evenly distributed over the different HGs, with the largest HG containing 115 markers and the smallest containing 5 (Figure 1). Only three (HGIV, HGVII, HGXI) of these 14 HGs obtained did not encompass EST-derived markers. 54 CGs containing two mapped markers (Table 3) and 18 CGs containing more than two markers (Figure 1) became stranded as independent groups because of the lack of common markers to discriminate HGs. In general, the order of SSR and RFLP markers was consistent between CGs found within an HG. But, as large gaps were still often formed, more markers were needed for the precise positioning of and to order these markers.

Homology groups in the present map were organized based on the groups of previous map. Three homology groups (HGIV, HGVII and HGXI) were exactly the same on both our map and the previous one, and four homology groups (HGV, HGXII, HGXIII and HGXIV) were established as new ones, based on EST-SSR markers. Except for HGIII and

HGVI that were formed by two groups from the first map (HGIII and HGV, HGVI and HGXII, respectively), the others were basically the same with new co-segregation groups derived from EST-based markers or previous co-segregation groups with new linked markers.

24 of the SSRs and 3 of the RFLPs detected duplicated loci within a co-segregation group. All but two homology groups (HGIV and HGXII) detected these duplications (Table 4). Besides, markers generated by 7 EST-SSRs were mapped to different homology groups: ESTA09 (HGIII and HGX), ESTB03 (HGVI and X), ESTB20 (HGVIII and X), ESTB110 (III and VI), ESTB149 (HGI and HGII), ESTC57 (HGIII and HGVIII) and ESTC123 (HGV and HGX) (Table 4).

Putative functions of the EST products containing SSR and RFLPs

To explore the potential utility of the EST-based markers, 113 EST-SSR loci and 6 EST-RFLP mapped probes were compared to genes of other species using BLASTX, directly from the SUCEST database. The EST was identified as the protein showing the highest score among the candidate proteins. Blast result revealed that, amongst the EST-SSRs sequences, 90.3% were derived from ESTs with functionally annotated hits in other organisms, predominantly *Arabidopsis thaliana*, followed by species taxonomically related to sugarcane such as, rice, maize, sorghum and other species. The remaining 9.7% showed no homology to sequences in the database interrogated (Table 4). From the 102 EST sequences homologous to functional hits, only 13 were associated with hypothetical or unknown protein. In the majority of the EST-SSRs, the BLAST homology comparisons indicated that the sequences had characterized functions.

Some important enzymes related to sugarcane metabolism were identified. They included putative sugarcane transporter (*cluster* SCEZRZ1014H07.g), fructose-bisphosphate adolase (*cluster* SCSGLV1008C05.g), peroxidase (*clusters* SCCCLB1002D05.g, SCEQRT1024F02.g, SCSGFL4C02D07.g), fructose -1,6-bisphosphatase (*clusters* SCAGLR1021D10.g, SCSGLV1008C05.g), diphosphate-fructose-6-phosphate 1-phosphotransferase (*clusters* SCCCCL4007E05.g, SCCCST1006B01.g) and glyceraldehyde-3-phosphate dehydrogenase (*cluster* SCMCFL5009H03.g). Besides the EST-SSR, the EST-RFLP probes also mapped directly genes such as: sucrose synthase (*cluster* SCCCLR1001A05.g), triosephosphate isomerase 1 (*cluster* SCCCRZ1001B08.g), cellulose synthase-4 (*cluster* SCCCLR1066B10.g) and S-adenosylmethionine synthetase (*cluster* SCCCLR1001G06.g).

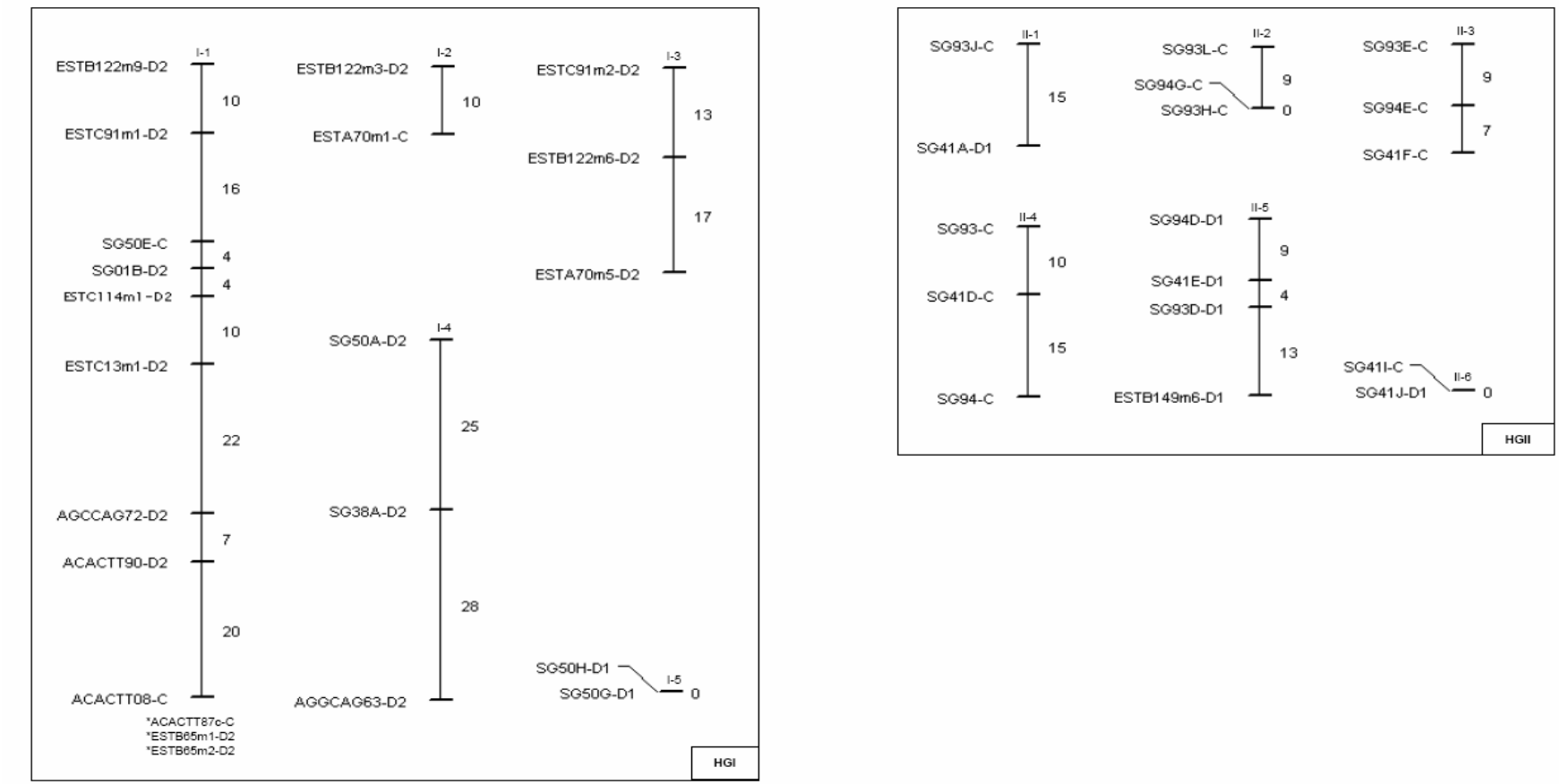
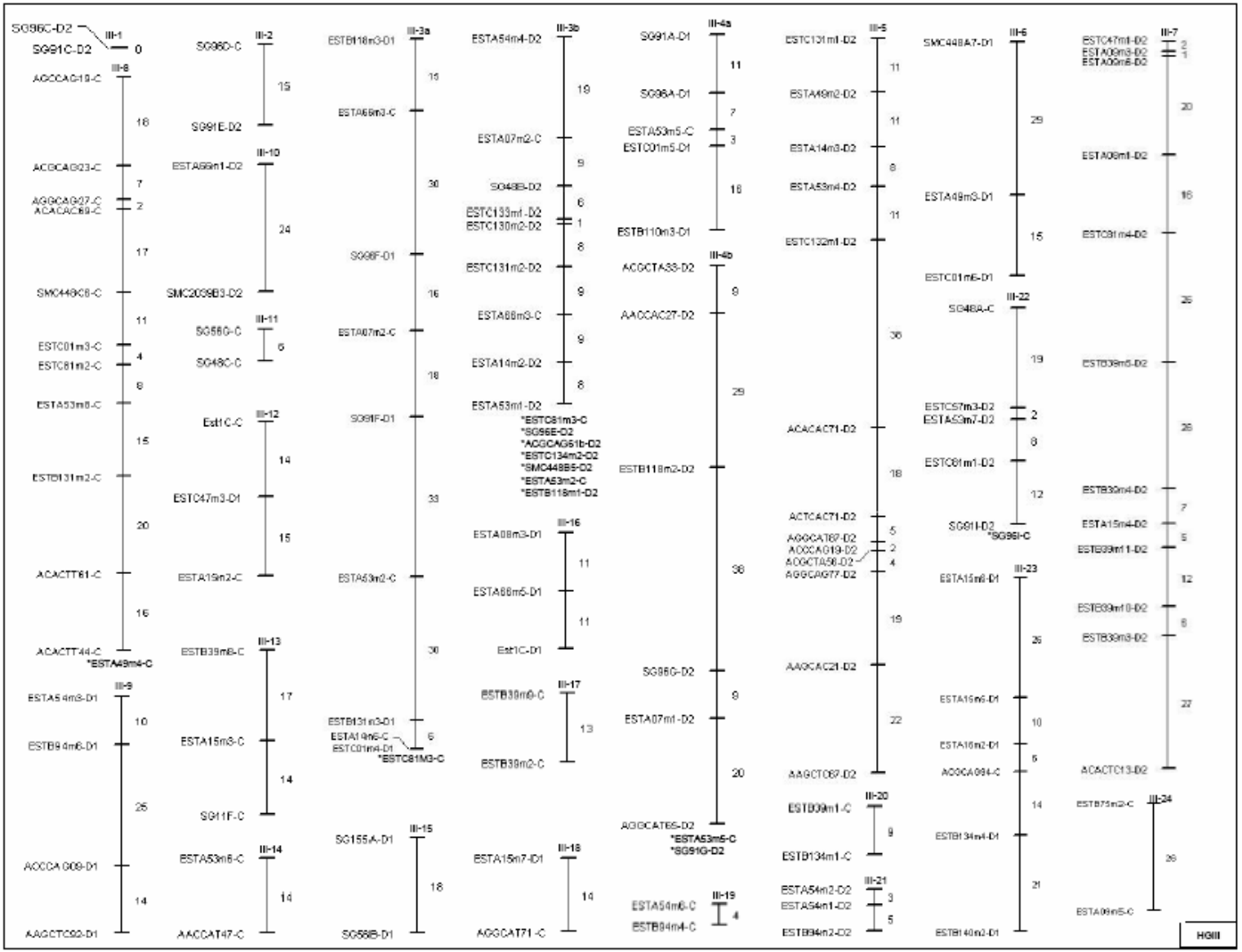
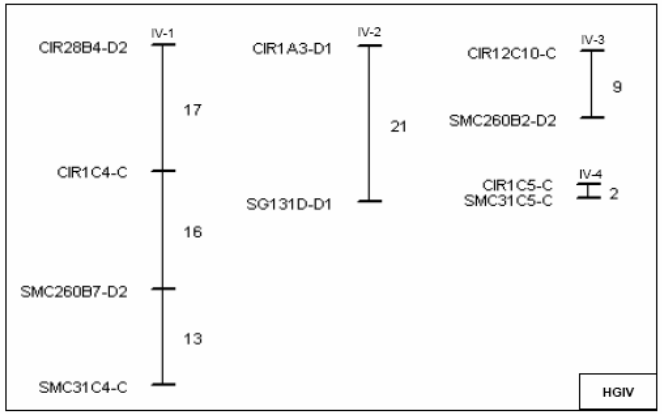
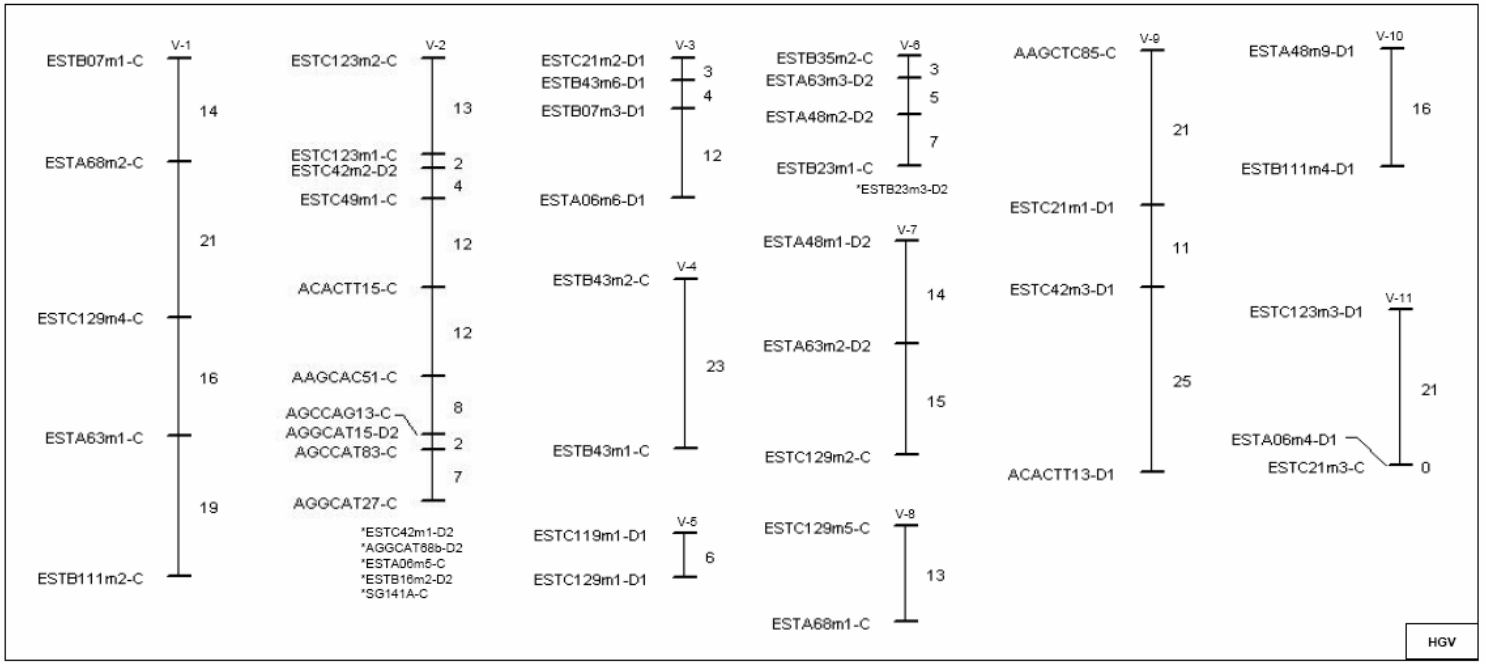
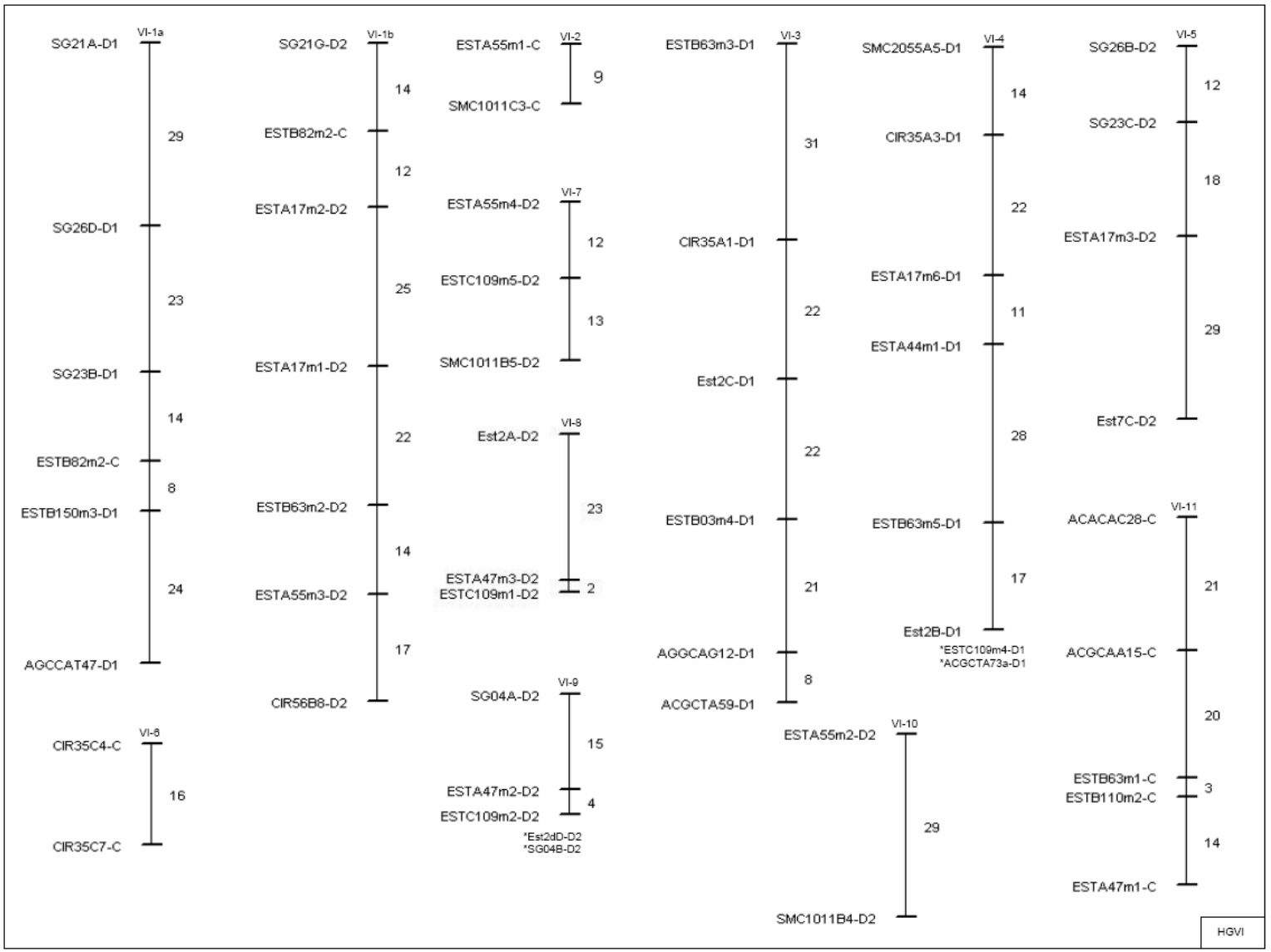


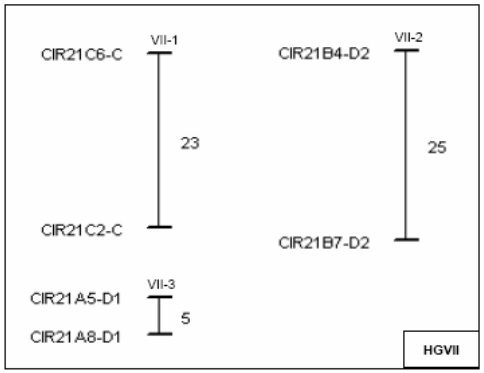
Figure 1 Functional integrated sugarcane map constructed using 100 individuals obtained from a commercial cross SP80-180 x SP80-4966. 664 segregating markers were assembled into 192 co-segregation groups (CGs) and 120 of these groups were arranged by 14 homology groups (HGs). CGs with two markers are listed in Table 3. Positions of loci are given in centiMorgans (Kosambi 1944), on the *right* of each CG and marker names are on the *left*. Accessory markers for some CGs are shown with *asterisks* and below their respective groups. These markers belong to the CG but ordering was not possible.

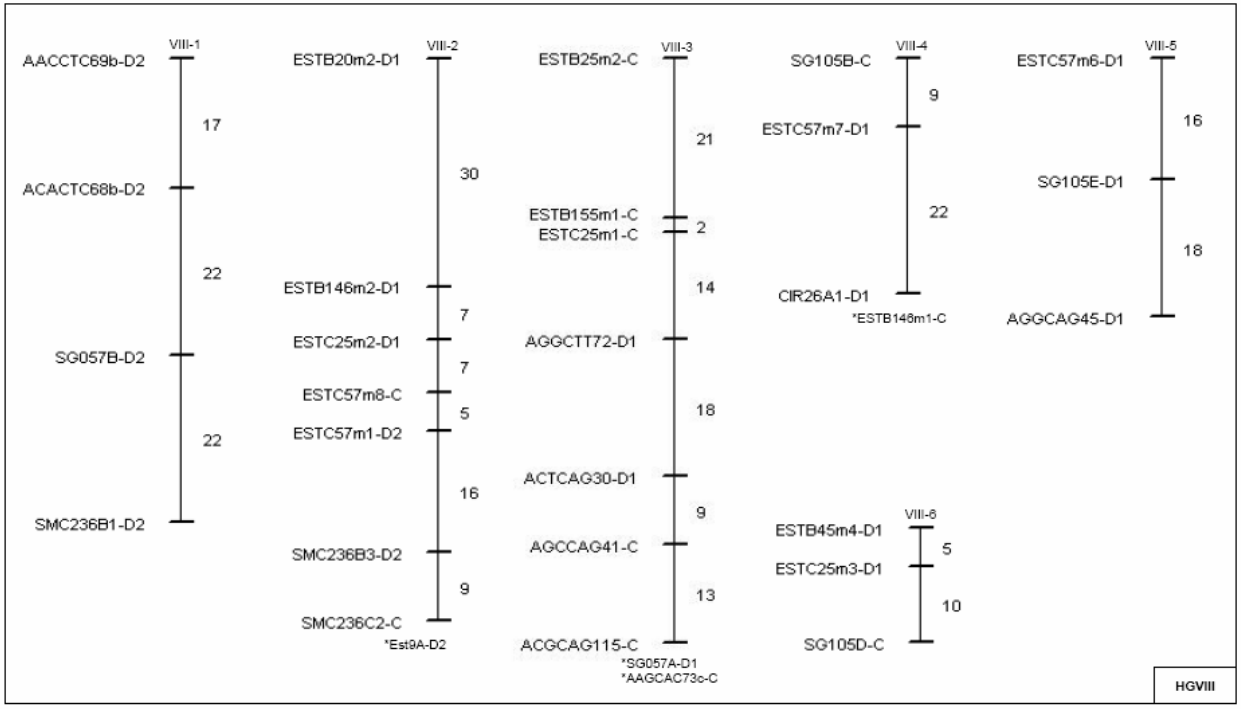

III-1
SG96C-D2
SG91C-D2
0
III-8
AGCCAG19-C
16
ACGCAG23-C
7
AGGCAG27-C
ACACAC69-C
2
17
SMC448C6-C
11
ESTC01m3-C
4
ESTC81m2-C
6
ESTA53m6-C
15
ESTB131m2-C
20
ACACTT61-C
16
ACACTT44-C
*ESTA49m4-C
III-9
ESTA54m3-D1
10
ESTB94m6-D1
25
AGCCAG09-D1
14
AAGCTC02-D1
III-2
SG96D-C
15
SG91E-D2
III-10
ESTA69m1-D2
24
SMC2039B3-D2
III-11
SG56G-C
6
SG48C-C
III-12
Est1C-C
14
ESTC47m3-D1
15
ESTA15m2-C
III-13
ESTB39m8-C
17
ESTA15m3-C
14
SG11F-C
III-14
ESTA53m6-C
14
AACCAT47-C
III-3a
ESTB118m3-D1
16
ESTA68m3-C
30
SG98F-D1
16
ESTA07m2-C
18
SG91F-D1
33
ESTA53m2-C
30
ESTB131m3-D1
ESTA19m6-C
ESTC01m4-D1
6
*ESTC81M3-C
III-15
SG155A-D1
18
SG56B-D1
III-3b
ESTA54m4-D2
19
ESTA07m2-C
9
SG48B-D2
6
ESTC133m1-D2
ESTC130m2-D2
1
8
ESTC131m2-D2
9
ESTA68m3-C
9
ESTA14m2-D2
8
ESTA53m1-D2
*ESTC81m3-C
*SG96E-D2
*ACGCAG51b-D2
*ESTC134m2-D2
*SMC448B5-D2
*ESTA53m2-C
*ESTB118m1-D2
III-16
ESTA08m3-D1
11
ESTA68m5-D1
11
Est1C-D1
III-17
ESTB39m9-C
13
ESTB39m2-C
III-18
ESTA15m7-D1
14
AGGCAT71-C
III-4a
SG91A-D1
11
SG96A-D1
7
ESTA53m5-C
ESTC01m5-D1
3
16
ESTB110m3-D1
III-4b
ACGCTA33-D2
9
AACCAC27-D2
29
ESTB118m2-D2
36
SG96G-D2
9
ESTA07m1-D2
20
AGGCAT65-D2
*ESTA53m5-C
*SG91G-D2
III-19
ESTA54m5-C
ESTB94m4-C
4
III-5
ESTC131m1-D2
11
ESTA49m2-D2
11
ESTA14m3-D2
8
ESTA53m4-D2
11
ESTC132m1-D2
36
ACACAC71-D2
18
ACTCAC71-D2
5
AGGCAT67-D2
ACCCAG19-D2
2
ACGCTA56-D2
4
AGGCAG77-D2
19
AAGCAC21-D2
22
AAGCTC67-D2
III-20
ESTB09m1-C
9
ESTB134m1-C
III-21
ESTA54m2-D2
3
ESTA54m1-D2
5
ESTB94m2-D2
III-6
SMC448A7-D1
29
ESTA49m3-D1
15
ESTC01m6-D1
III-22
SG48A-C
19
ESTC57m3-D2
ESTA53m7-D2
2
8
ESTC81m1-D2
12
SG91I-D2
*SG96I-C
III-23
ESTA15m8-D1
26
ESTA15m5-D1
10
ESTA16m2-D1
5
ACGCA094-C
14
ESTB134m4-D1
21
ESTB140m2-D1
III-7
ESTC47m1-D2
ESTA09m3-D2
ESTA09m8-D2
2
1
20
ESTA08m1-D2
16
ESTC81m4-D2
26
ESTB39m5-D2
26
ESTB39m4-D2
7
ESTA15m4-D2
5
ESTB39m1-D2
12
ESTB39m10-D2
6
ESTB39m3-D2
27
ACACTC13-D2
III-24
ESTB75m2-C
26
ESTA09m5-C
HGIII

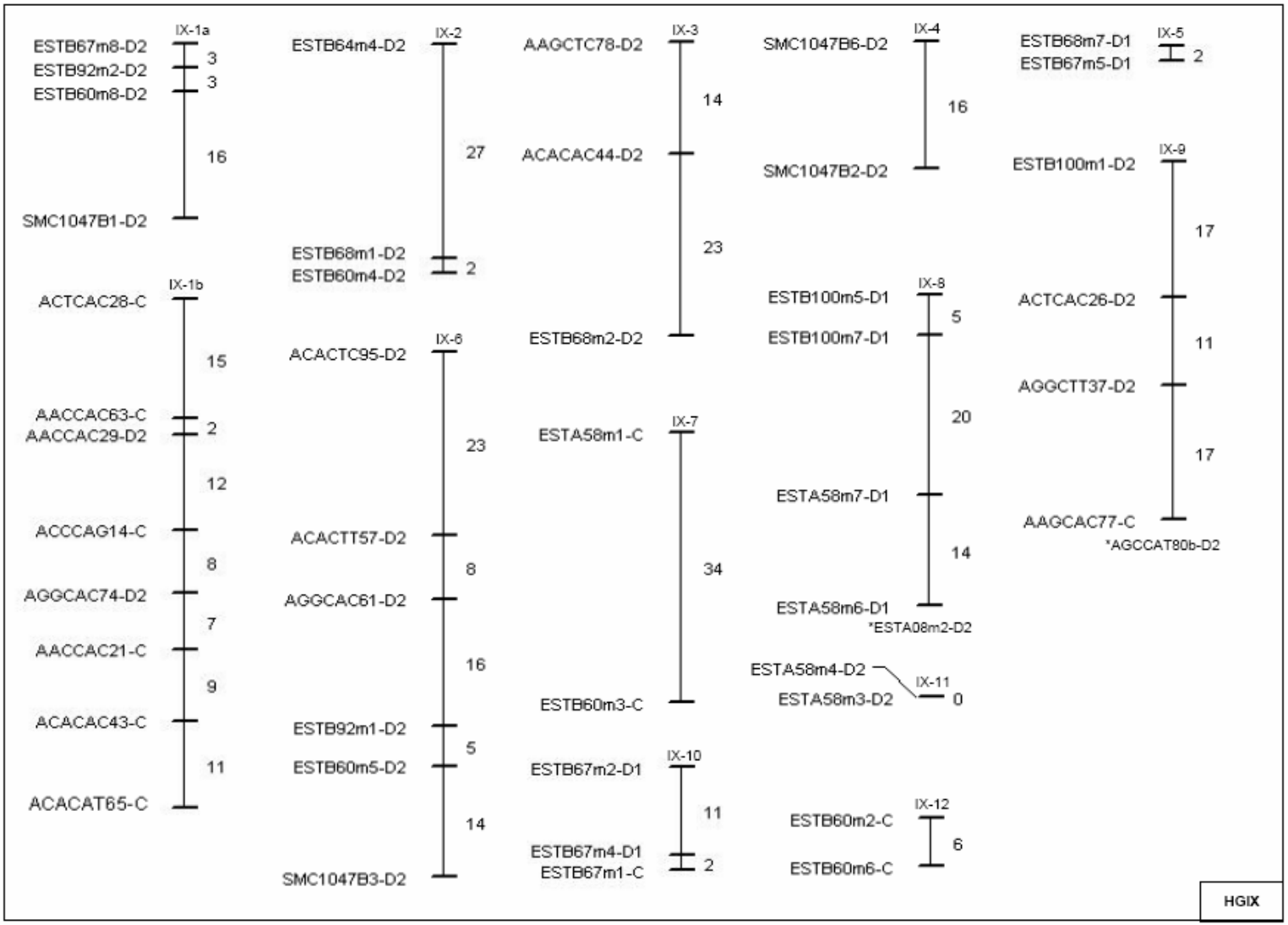

IV-1
CIR28B4-D2
17
CIR1C4-C
16
SMC260B7-D2
13
SMC31C4-C
IV-2
CIR1A3-D1
21
SG131D-D1
IV-3
CIR12C10-C
9
SMC260B2-D2
IV-4
CIR1C5-C
SMC31C5-C
2
HGIV

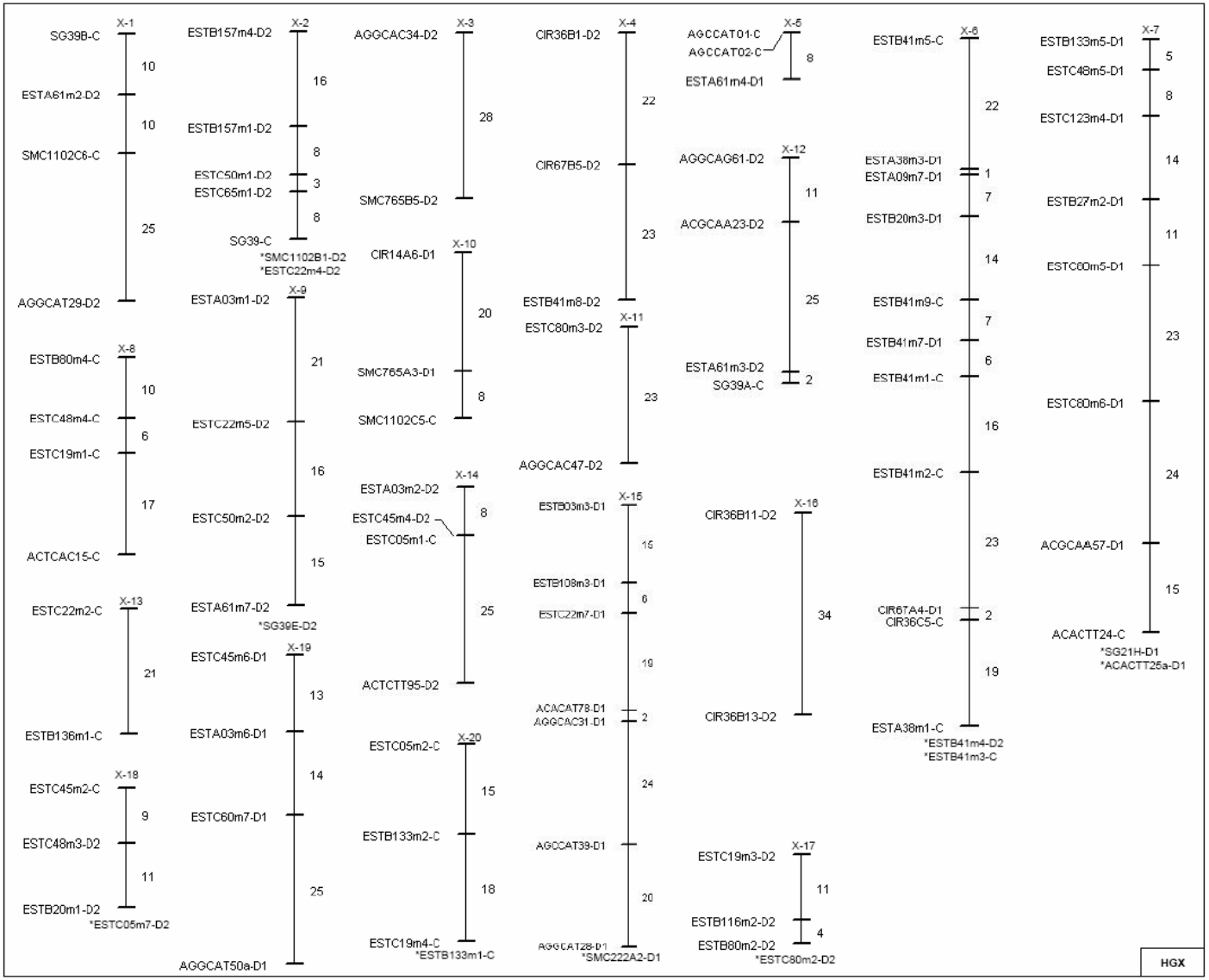

V-1
ESTB07m1-C
14
ESTA68m2-C
21
ESTC129m4-C
16
ESTA63m1-C
19
ESTB111m2-C
V-2
ESTC123m2-C
13
ESTC123m1-C
ESTC42m2-D2
2
ESTC49m1-C
4
12
ACACTT15-C
12
AAGCAC51-C
AGCCAG13-C
AGGCAT15-D2
AGCCAT83-C
8
2
7
AGGCAT27-C
*ESTC42m1-D2
*AGGCAT68b-D2
*ESTA06m5-C
*ESTB16m2-D2
*SG141A-C
V-3
ESTC21m2-D1
ESTB43m6-D1
3
ESTB07m3-D1
4
12
ESTA06m6-D1
V-4
ESTB43m2-C
23
ESTB43m1-C
V-5
ESTC119m1-D1
6
ESTC129m1-D1
V-6
ESTB35m2-C
ESTA63m3-D2
3
ESTA48m2-D2
5
7
ESTB23m1-C
*ESTB23m3-D2
V-7
ESTA48m1-D2
14
ESTA63m2-D2
15
ESTC129m2-C
V-8
ESTC129m5-C
13
ESTA68m1-C
V-9
AAGCTC85-C
21
ESTC21m1-D1
11
ESTC42m3-D1
25
ACACTT13-D1
V-10
ESTA48m9-D1
16
ESTB111m4-D1
V-11
ESTC123m3-D1
21
ESTA06m4-D1
ESTC21m3-C
0
HGV

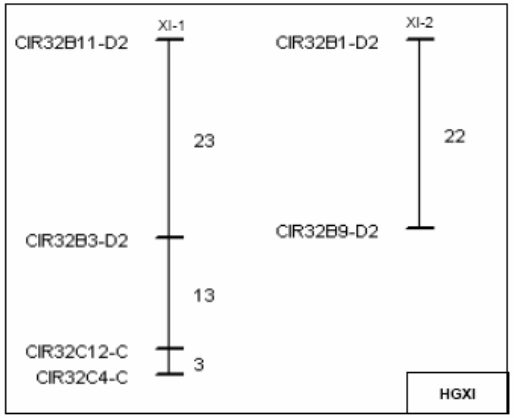

VI-1a
SG21A-D1
29
SG26D-D1
23
SG23B-D1
14
ESTB82m2-C
8
ESTB150m3-D1
24
AGCCAT47-D1
VI-6
CIR35C4-C
16
CIR35C7-C
VI-1b
SG21G-D2
14
ESTB82m2-C
12
ESTA17m2-D2
25
ESTA17m1-D2
22
ESTB63m2-D2
14
ESTA55m3-D2
17
CIR56B8-D2
VI-2
ESTA55m1-C
9
SMC1011C3-C
VI-7
ESTA55m4-D2
12
ESTC109m5-D2
13
SMC1011B5-D2
VI-8
Est2A-D2
23
ESTA47m3-D2
ESTC109m1-D2
2
VI-9
SG04A-D2
15
ESTA47m2-D2
4
ESTC109m2-D2
*Est2dD-D2
*SG04B-D2
VI-3
ESTB63m3-D1
31
CIR35A1-D1
22
Est2C-D1
22
ESTB03m4-D1
21
AGGCAG12-D1
8
ACGCTA59-D1
ESTA55m2-D2
VI-10
29
SMC1011B4-D2
VI-4
SMC2055A5-D1
14
CIR35A3-D1
22
ESTA17m6-D1
11
ESTA44m1-D1
28
ESTB63m5-D1
17
Est2B-D1
*ESTC109m4-D1
*ACGCTA73a-D1
VI-5
SG26B-D2
12
SG23C-D2
18
ESTA17m3-D2
29
Est7C-D2
VI-11
ACACAC28-C
21
ACGCAA15-C
20
ESTB63m1-C
3
ESTB110m2-C
14
ESTA47m1-C
HGVI

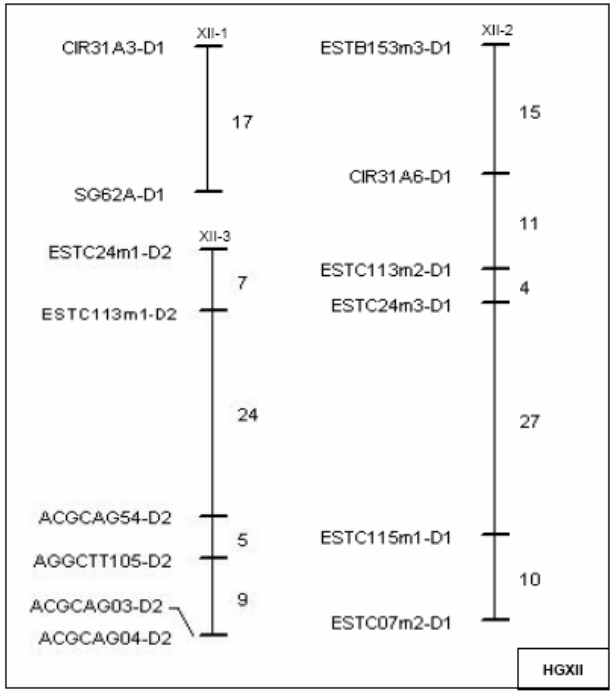

VII-1
CIR21C6-C
23
CIR21C2-C
VII-2
CIR21B4-D2
25
CIR21B7-D2
VII-3
CIR21A5-D1
5
CIR21A8-D1
HGVII

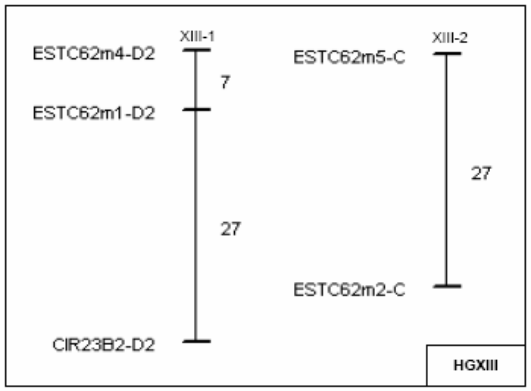

VIII-1
AACCTC69b-D2
17
ACACTC68b-D2
22
SG057B-D2
22
SMC236B1-D2
VIII-2
ESTB20m2-D1
30
ESTB146m2-D1
7
ESTC25m2-D1
7
ESTC57m8-C
5
ESTC57m1-D2
16
SMC236B3-D2
9
SMC236C2-C
*Est9A-D2
VIII-3
ESTB25m2-C
21
ESTB155m1-C
ESTC25m1-C
2
14
AGGCTT72-D1
18
ACTCAG30-D1
9
AGCCAG41-C
13
ACGCAG115-C
*SG057A-D1
*AAGCAC73c-C
VIII-4
SG105B-C
9
ESTC57m7-D1
22
CIR26A1-D1
*ESTB146m1-C
VIII-5
ESTC57m6-D1
16
SG105E-D1
18
AGGCAG45-D1
VIII-6
ESTB45m4-D1
5
ESTC25m3-D1
10
SG105D-C
HGVIII

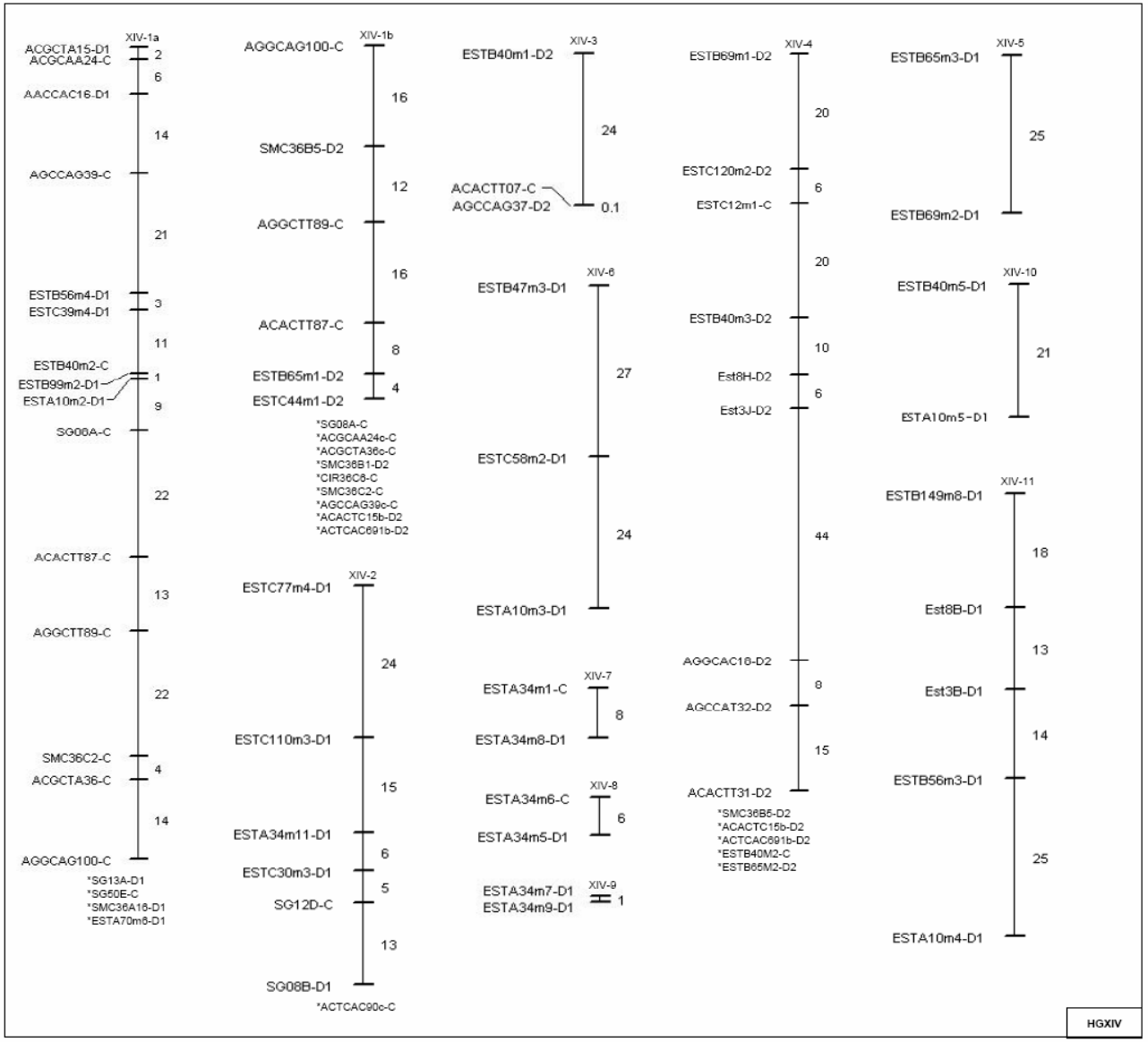

IX-1a
ESTB67m8-D2
3
ESTB92m2-D2
3
ESTB60m8-D2
16
SMC1047B1-D2
IX-1b
ACTCAC28-C
15
AACCAC63-C
2
AACCAC29-D2
12
ACCCAG14-C
8
AGGCAC74-D2
7
AACCAC21-C
9
ACACAC43-C
11
ACACAT65-C
IX-2
ESTB64m4-D2
27
ESTB68m1-D2
2
ESTB60m4-D2
IX-6
ACACTC95-D2
23
ACACTT57-D2
8
AGGCAC61-D2
16
ESTB92m1-D2
5
ESTB60m5-D2
14
SMC1047B3-D2
IX-3
AAGCTC78-D2
14
ACACAC44-D2
23
ESTB68m2-D2
IX-7
ESTA58m1-C
34
ESTB60m3-C
IX-10
ESTB67m2-D1
11
ESTB67m4-D1
2
ESTB67m1-C
IX-4
SMC1047B6-D2
16
SMC1047B2-D2
IX-8
ESTB100m5-D1
5
ESTB100m7-D1
20
ESTA58m7-D1
14
ESTA58m6-D1
*ESTA08m2-D2
ESTA58m4-D2
IX-11
ESTA58m3-D2
0
IX-12
ESTB60m2-C
6
ESTB60m6-C
IX-5
ESTB68m7-D1
ESTB67m5-D1
2
IX-9
ESTB100m1-D2
17
ACTCAC26-D2
11
AGGCTT37-D2
17
AAGCAC77-C
*AGCCAT80b-D2
HGIX


X-1
SG39B-C
10
ESTA61m2-D2
10
SMC1102C6-C
25
AGGCAT29-D2
X-2
ESTB157m4-D2
16
ESTB157m1-D2
8
ESTC50m1-D2
3
ESTC65m1-D2
8
SG39-C
*SMC1102B1-D2
*ESTC22m4-D2
X-3
AGGCAC34-D2
28
SMC765B5-D2
X-4
CIR36B1-D2
22
CIR67B5-D2
23
ESTB41m8-D2
X-5
AGCCAT01-C
AGCCAT02-C
8
ESTA61m4-D1
X-6
ESTB41m5-C
22
ESTA38m3-D1
ESTA09m7-D1
1
7
ESTB20m3-D1
14
ESTB41m9-C
7
ESTB41m7-D1
6
ESTB41m1-C
16
ESTB41m2-C
23
CIR67A4-D1
CIR36C5-C
2
19
ESTA38m1-C
*ESTB41m4-D2
*ESTB41m3-C
X-7
ESTB133m5-D1
5
ESTC48m5-D1
8
ESTC123m4-D1
14
ESTB27m2-D1
11
ESTC60m5-D1
23
ESTC80m6-D1
24
ACGCAA57-D1
15
ACACTT24-C
*SG21H-D1
*ACACTT25a-D1
X-8
ESTB80m4-C
10
ESTC48m4-C
6
ESTC19m1-C
17
ACTCAC15-C
X-9
ESTA03m1-D2
21
ESTC22m5-D2
16
ESTC50m2-D2
15
ESTA61m7-D2
*SG39E-D2
X-10
CIR14A6-D1
20
SMC765A3-D1
8
SMC1102C5-C
X-11
ESTC80m3-D2
23
AGGCAC47-D2
X-12
AGGCAG61-D2
11
ACGCAA23-D2
25
ESTA61m3-D2
SG39A-C
2
X-13
ESTC22m2-C
21
ESTB136m1-C
X-14
ESTA03m2-D2
8
ESTC45m4-D2
ESTC05m1-C
25
ACTCTT95-D2
X-15
ESTB03m3-D1
15
ESTB108m3-D1
6
ESTC22m7-D1
19
ACACAT78-D1
AGGCAC31-D1
2
24
AGCCAT39-D1
20
AGGCAT28-D1
*SMC222A2-D1
X-16
CIR36B11-D2
34
CIR36B13-D2
X-17
ESTC19m3-D2
11
ESTB116m2-D2
4
ESTB80m2-D2
*ESTC80m2-D2
X-18
ESTC45m2-C
9
ESTC48m3-D2
11
ESTB20m1-D2
*ESTC05m7-D2
X-19
ESTC45m6-D1
13
ESTA03m6-D1
14
ESTC60m7-D1
25
AGGCAT50a-D1
X-20
ESTC05m2-C
15
ESTB133m2-C
18
ESTC19m4-C
*ESTB133m1-C
HGX


XI-1
CIR32B11-D2
23
CIR32B3-D2
13
CIR32C12-C
CIR32C4-C
3
XI-2
CIR32B1-D2
22
CIR32B9-D2
HGXI


XII-1
CIR31A3-D1
17
SG62A-D1
XII-3
ESTC24m1-D2
7
ESTC113m1-D2
24
ACGCAG54-D2
5
AGGCTT105-D2
9
ACGCAG03-D2
ACGCAG04-D2
XII-2
ESTB153m3-D1
15
CIR31A6-D1
11
ESTC113m2-D1
4
ESTC24m3-D1
27
ESTC115m1-D1
10
ESTC07m2-D1
HGXII


XIII-1
ESTC62m4-D2
7
ESTC62m1-D2
27
CIR23B2-D2
XIII-2
ESTC62m5-C
27
ESTC62m2-C
HGXIII


XIV-1a
ACGCTA15-D1
ACGCAA24-C
2
6
AACCAC16-D1
14
AGCCAG39-C
21
ESTB56m4-D1
ESTC39m4-D1
3
11
ESTB40m2-C
ESTB99m2-D1
ESTA10m2-D1
1
9
SG08A-C
22
ACACTT87-C
13
AGGCTT89-C
22
SMC36C2-C
4
ACGCTA36-C
14
AGGCAG100-C
*SG13A-D1
*SG50E-C
*SMC36A16-D1
*ESTA70m6-D1
XIV-1b
AGGCAG100-C
16
SMC36B5-D2
12
AGGCTT89-C
16
ACACTT87-C
8
ESTB65m1-D2
4
ESTC44m1-D2
*SG08A-C
*ACGCAA24c-C
*ACGCTA36c-C
*SMC36B1-D2
*CIR36C8-C
*SMC36C2-C
*AGCCAG39c-C
*ACACTC15b-D2
*ACTCAC691b-D2
XIV-2
ESTC77m4-D1
24
ESTC110m3-D1
15
ESTA34m11-D1
6
ESTC30m3-D1
5
SG12D-C
13
SG08B-D1
*ACTCAC90c-C
XIV-3
ESTB40m1-D2
24
ACACTT07-C
AGCCAG37-D2
0.1
XIV-6
ESTB47m3-D1
27
ESTC58m2-D1
24
ESTA10m3-D1
XIV-7
ESTA34m1-C
8
ESTA34m8-D1
XIV-8
ESTA34m6-C
6
ESTA34m5-D1
XIV-9
ESTA34m7-D1
ESTA34m9-D1
1
XIV-4
ESTB69m1-D2
20
ESTC120m2-D2
6
ESTC12m1-C
20
ESTB40m3-D2
10
Est8H-D2
6
Est3J-D2
44
AGGCAC16-D2
8
AGCCAT32-D2
15
ACACTT31-D2
*SMC36B5-D2
*ACACTC15b-D2
*ACTCAC691b-D2
*ESTB40M2-C
*ESTB65M2-D2
XIV-5
ESTB65m3-D1
25
ESTB69m2-D1
XIV-10
ESTB40m5-D1
21
ESTA10m5-D1
XIV-11
ESTB149m8-D1
18
Est8B-D1
13
Est3B-D1
14
ESTB56m3-D1
25
ESTA10m4-D1
HGXIV

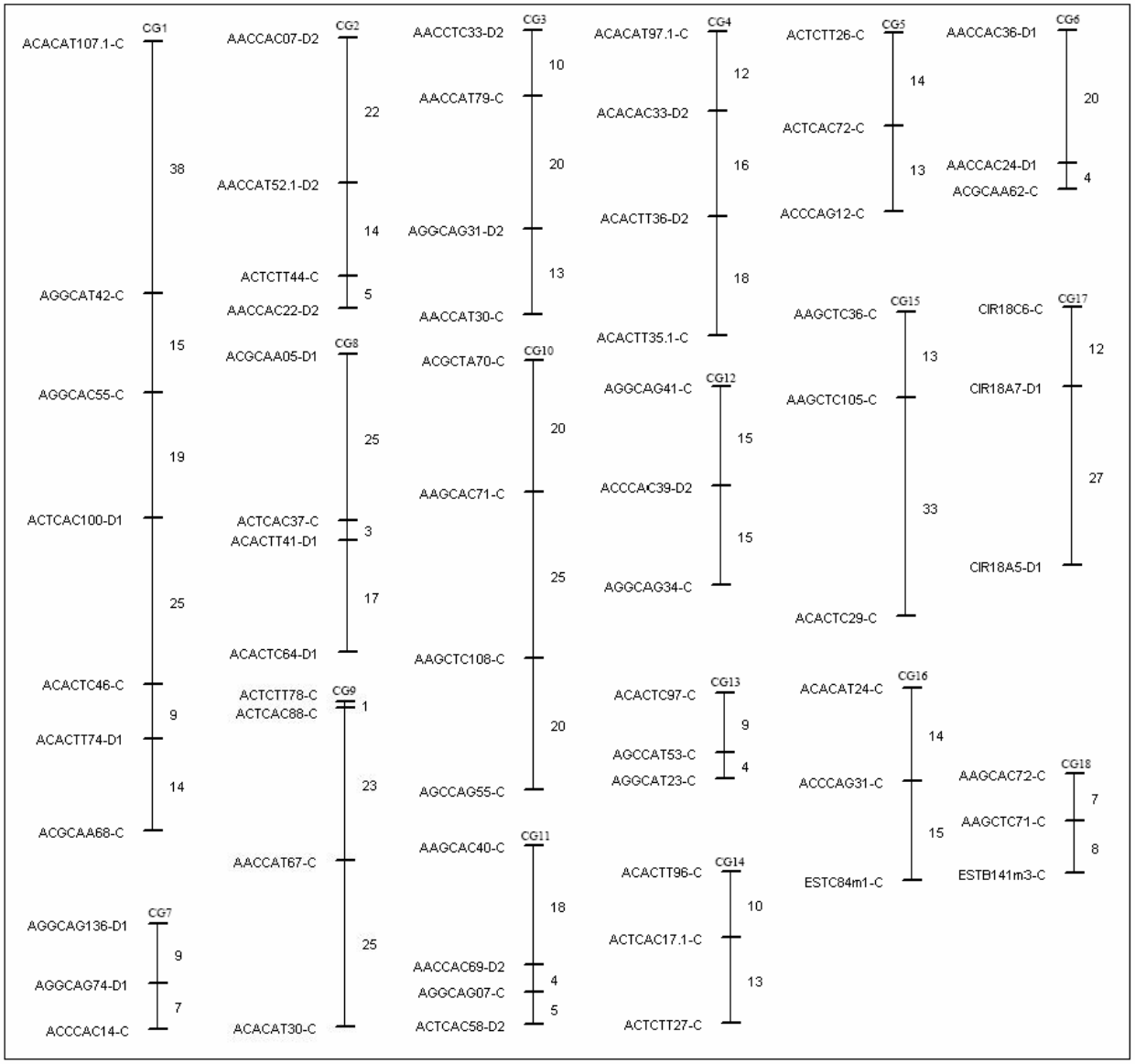

CG1
ACACAT107.1-C
38
AGGCAT42-C
15
AGGCAC55-C
19
ACTCAC100-D1
25
ACACTC46-C
9
ACACTT74-D1
14
ACGCAA68-C
CG7
AGGCAG136-D1
9
AGGCAG74-D1
7
ACCCAC14-C
CG2
AACCAC07-D2
22
AACCAT52.1-D2
14
ACTCTT44-C
5
AACCAC22-D2
CG8
ACGCAA05-D1
25
ACTCAC37-C
3
ACACTT41-D1
17
ACACTC64-D1
CG9
ACTCTT78-C
1
ACTCAC88-C
23
AACCAT67-C
25
ACACAT30-C
CG3
AACCTC33-D2
10
AACCAT79-C
20
AGGCAG31-D2
13
AACCAT30-C
CG10
ACGCTA70-C
20
AAGCAC71-C
25
AAGCTC108-C
20
AGCCAG55-C
CG11
AAGCAC40-C
18
AACCAC69-D2
4
AGGCAG07-C
5
ACTCAC58-D2
CG4
ACACAT97.1-C
12
ACACAC33-D2
16
ACACTT36-D2
18
ACACTT35.1-C
CG12
AGGCAG41-C
15
ACCCAC39-D2
15
AGGCAG34-C
CG13
ACACTC97-C
9
AGCCAT53-C
4
AGGCAT23-C
CG14
ACACTT96-C
10
ACTCAC17.1-C
13
ACTCTT27-C
CG5
ACTCTT26-C
14
ACTCAC72-C
13
ACCCAG12-C
CG15
AAGCTC36-C
13
AAGCTC105-C
33
ACACTC29-C
CG16
ACACAT24-C
14
ACCCAG31-C
15
ESTC84m1-C
CG6
AACCAC36-D1
20
AACCAC24-D1
4
ACGCAA62-C
CG17
CIR18C6-C
12
CIR18A7-D1
27
CIR18A5-D1
CG18
AAGCAC72-C
7
AAGCTC71-C
8
ESTB141m3-C

Table 3 Co-segregation groups (CGs) with two markers in the integrated genetic map of SP80-180 x SP80-4966 mapping population.

Markers	Distance (cM)
SMC119C5-C, SMC119C4-C (CG134)	28
SMC2017A1-D1, SMC41549-D1 (CG117); ACCCAC13a-D1, ESTB37m2-D1 (CG139); CIR55B3-D2, CIR55B17-D2 (CG121)	26
ACGCAA18b-D2, AGCCAT68-D2 (CG155)	25
ACGCAG76a-D1, ACTCAC33a-D1 (CG141); AGCCAT38-D2, AGGCTT21b-D2 (CG159); AGCCAG25a-D1, ESTB58m1-D1 (CG144)	24
AGCCAT19-D2, AGGCAG118b-D2 (CG158); ACACAC68c-C, ACACAC90c-C (CG179); ACACAC56c-C, AGGCAG17c-C (CG178); ACTCTT42c-C, AGCCAT76c-C (CG184)	20
ACACAT45b-D2, AGGCAT13b-D2 (CG150); SG16D-D2, SG16I-D2 (CG143)	19
ACACAC58a-D1, AGGCAT26a-D1 (CG138)	17
AGGCTT90b-D2, ACCCAG17c-C (CG164); ACGCAG59c-C, ACGCAG60c-C (CG183)	16
ACGCAG26b-D2, ACTCAG16b-D2 (CG156)	15
AAGCTC22c-C, ACCCAG53c-C (CG174); AACCTC52c-C, ACGCAA39c-C (CG170)	14
ACGCAA66c-C, ACTCAG33c-C (CG182); AACCAT64.1c-C, AAGCAC50c-C (CG169); ACACTT25.1b-D2, ACTCAC58.1b-D2 (CG151)	13
ACTCAC52a-D1, AAGCAC60c-C (CG140); AGCCAG24c-C, AGCCAG33c-C (CG186); SMC21C2-C, SMC21C5-C (CG131)	12
ACCAC34c-C, ACGCAG70c-C (CG166); ACCCAC72b-D2, AGGCAC21c-C (CG154)	11
AACCAC88a-D1, AGGCAC38a-D1 (CG136)	10
ACACAC48.1b-D2, ACTCAC38b-D2 (CG149)	9
ACCCAG50c-C, ACCCAG67c-C (CG181); AACCAC50b-D2, ACACTT81c-C (CG146); AACCAC83b-D2, AACCTC30b-D2 (CG147)	8
AAGCAC04c-C, AAGCAC05c-C (CG171); ESTB130m2-D2, ESTC78m1-D2 (CG100); ACACAT42c-C, ACACTC20c-C (CG180); ACACAC29c-C, ACACAT47c-C (CG176); AACCAC46c-C, ACTCTT85c-C (CG167)	7
AGGCAC32c-C, AGGCAG42c-C (CG187)	6
ACTCAC35a-D1, AGCCAT82a-D1 (CG142); ACTCTT73c-C, AGGCTT39c-C (CG185)	4
AGGCAG129a-D1, AAGCTC96c-C (CG145); AACCAC08c-C, AACCAT63.1c-C (CG165)	3
AACCAT44.1a-D1, ACACAC100c-C (CG137)	2
SMC280B9-D2, SMC280B8-D2 (CG129); ACACAC53c-C, AGGCAT06c-C (CG177)	1
ACTCTT35b-D2, ACACTT23c-C (CG157); AGGCAC01b-D2, AGGCAC40c-C (CG160); CIR4A1-D1, CIR4C2-C (CG118); ACACTT63b-D2, ACGCAG127c-C (CG152); ACCCAC18b-D2, ACACAT54c-C (CG153); ESTB145m3-C, ESTB145m1-D2 (CG106); AAGCAC02b-D2, AAGCAC03b-D2 (CG148); AAGCAC70.1c-C, ACGCAG25c-C (CG172)	0

Table 4 EST-SSRs and EST-RFLPs linked markers: location on homology groups (HGs) and expected EST homology.

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]
SCCCLR1001A05.g	III (2)	Est1	Est1CD1, Est1CC	Sucrose synthase	34391404
SCCCRZ1001B08.g	VI (4)	Est2	Est2BD1, Est2AD2, Est2DD2, Est2CD1	Triosephosphate isomerase 1	168647
SCCCLR1066B10.g	XIV (2)	Est3	Est3JD2, Est3BD1	Cellulose synthase-4	50935131
SCCCLR1001G06.g	VI (1)	Est7	Est7CD2	S-adenosylmethionine synthetase	51038250
SCCCLR1066B10.g	XIV (2)	Est8	Est8HD2, Est8BD1	Cellulose synthase-4	50935131
SCUTLR1058B04.g	VIII (1)	Est9	Est9A-D2	OSJNBa0072K14.5	50923903
SCCCCL7038A01.g	X (3)	ESTA03	m1D2, m2D2, m6D1	Putative beta-mannan endohydrolase	34909462
SCEZLR1031G07.g	V (3)	ESTA06	m5C, m6D1, m4D1	Lipid transfer protein	25990344
SCRUFL1118A11.b	III (2)	ESTA07	m2C, m1D2	Putative apical-basal pattern formation protein	50917563
SCRLFL4027F03.g	III (3)	ESTA08	m3D1, m1D2, m2D2	Hypothetical protein	55773910
SCCCRZ2C03C03.g	III / X (3)	**/*ESTA09	m3D2, m6D2, m5C, m7D1	No hit	
SCQGSB1143H06.g	XIV (4)	ESTA10	m2D1, m5D1, m3D1, m4D1	No hit	
SCRLAD1139G03.g	III (2)	ESTA14	m6C, m2D2, m3D2	No hit	
SCCCHR1004C07.g	III (5)	**ESTA15	m2C, m7D1, m3C, m4D2, m6D1, m5D1	Caffeoyl CoA 3-O-methyltransferase	32347451
SCCCLR1001C05.g	III (1)	ESTA16	m2D1	Hypothetical protein	50900112
SCEQRT2030G04.g	VI (3)	**ESTA17	m2D2, m1D2, m6D1, m3D2	Hypothetical protein	34894358
SCVPCL6061A01.g	XIV (4)	** ESTA34	m11D1, m1C, m8D1, m6C, m5D1, m7D1, m9D1	Hypothetical protein	12597881
SCJLRZ1025A05.g	X (1)	**ESTA38	m3D1, m1C	No hit	
SCJFLR1035E10.g	VI (1)	ESTA44	m1D1	OSJNBa0086B14.26	50924658
SCRULB2062C01.b	VI (3)	ESTA47	m1C, m3D2, m2D2	Putative branched-chain amino acid aminotransferase protein	34904468
SCACAM2044B11.g	V (3)	ESTA48	m2D2, m1D2, m9D1	Chloroplast phytoene synthase 1	38037628
SCEZHR1087D09.g	III (3)	ESTA49	m4C, m3D1, m2D2	Putative glutathione S-transferase	50918191
SCSFHR1043F12.g	III (6)	ESTA53	m2C, m1D2,m5C, m8C, m4D2, m6C, m7D2	Indeterminate spikelet 1	2944040
SCRULB1059B07.g	III (4)	**ESTA54	m4D2, m6C, m3D1, m2D2, m1D2	No hit	
SCQGST1032G11.g	VI (4)	ESTA55	m3D2, m1C, m4D2, m2D2	No hit	
SCSBSD1029G01.g	IX (3)	**ESTA58	m1C, m7D1, m6D1, m4D2, m3D2	No hit	
SCMCRT2103A12.g	X (4)	ESTA61	m2D2, m4D1, m3D2, m7D2	Lipid transfer protein-like protein	39748092
SCCCLB1002D05.g	V (3)	ESTA63	m1C, m3D2, m2D2	TPA: class III peroxidase 90 precursor	55701047
SCMCRT2088D10.g	III (3)	ESTA66	m3C, m1D2, m5D1	Putative protein kinase	50901454
SCEQRT1024F02.g	I (2)	ESTA70	m6D1, m1C	Peroxidase	51038054
SCCCCL4013D09.g	VI / X (2)	*ESTB03	m4D1, m3D1	12-oxo-phytodienoic acid reductase	63021731
SCCCST1006B01.g	V (2)	ESTB07	m3D1, m1C	Putative diphosphate-fructose-6-phosphate 1-phosphotransferase	55168177
SCSGLV1008C05.g	V (2)	ESTA68	m2C, m1C	Fructose 1,6, bisphosphate aldolase	40362980

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]
SCUTAM2009E01.g	IX (2)	**ESTB100	m5D1, m7D1, m1D2	22 kDa drought-inducible protein	49234816
SCJFRZ2027D03.g	X (1)	ESTB108	m3D1	Hypothetical protein	50939565
SCCCCL4004A10.g	III / VI (2)	*ESTB110	m3D1, m2C	Putative polyprotein	48209910
SCCCCL1001E03.g	V (2)	ESTB111	m2C, m4D1	No hit	
SCCCCL4004B01.g	X (1)	ESTB116	m2D2	P0497A05.15	50902306
SCSBHR1053D01.g	III (2)	ESTB118	m3D1, m1D2, m2D2	Homeobox transcription factor GNARLY1	32351469
SCCCST1002H09.g	I (3)	ESTB122	m9D2, m3D2, m6D2	Hypothetical protein	50931511
SCMCAM1101H03.g	III (2)	ESTB131	m3D1, m2C	Putative helicase	50919223
SCRFAM1025G10.g	X (2)	ESTB133	m2C, m1C, m5D1	HSV-I stimulating-related protein-like	57899937
SCQGST3123C08.g	III (2)	ESTB134	m1C, m4D1	Putative pectin methylesterase	51090795
SCEQRT2025G01.g	X (1)	ESTB136	m1C	Putative AP2 domain containing protein	50920393
SCVPCL6044E08.g	III (1)	ESTB140	m2D1	Hypothetical protein	50910633
SCCCCL4010B11.g	VIII (2)	ESTB146	m2D1, m1C	OSJNBa0071I13.13	50929299
SCSGFL4C06A02.g	XIV / II (2)	*ESTB149	m8D1, m6D1	Cellulose synthase-7	9622886
SCUTLR1058B02.g	VI (1)	ESTB150	m3D1	OSJNBa0011P19.5	34894718
SCEZRZ3100H11.g	XII (1)	ESTB153	m3D1	Multiple stress-associated zinc-finger protein	37548823
SCBGLR1119A10.g	VIII (1)	ESTB155	m1C	OSJNBa0043L09.23	50928999
SCRFLR1034H10.g	X (1)	**ESTB157	m4D2, m1D2	Putative thaumatin-like protein	37531042
SCQGLR1041E11.g	V (1)	ESTB16	m2D2	SbCBF5	60593387
SCQSRT2032H08.g	VIII / X (3)	*ESTB20	m2D1, m1D2, m3D1	Hypothetical protein	56784702
SCRLFL1006B02.g	V (1)	ESTB23	m1C, m3D2	MFS18	22647
SCRUFL1120C04.b	VIII (1)	ESTB25	m2C	Putative benzothiadiazole-induced somatic embryogenesis receptor kinase 1	53792830
SCRUSB1078F07.g	X (1)	ESTB27	m2D1	P0480C01.19	34910656
SCVPLR2027E01.g	V (1)	ESTB35	m2C	MFS18	22647
SCBGLR1002F11.g	III (4)	**ESTB39	m5D2, m4D2, m11D2, m3D2, m10D2,m9C, m2C, m8C, m1C	Alpha 3 subunit of 20S proteasome	34898416
SCBGLR1023A04.g	XIV (3)	ESTB40	m2C, m3D2, m5D1, m1D2	Blue copper-binding protein-like	50882442
SCEQSD2077B12.g	X (2)	**ESTB41	m5C, m9C, m7D1, m1C, m2C, m4D2, m3C, m8D2	Hypothetical protein	50916060
SCCCCL4007E05.g	V (2)	**ESTB43	m6D1, m2C, m1C	Putative diphosphate-fructose-6-phosphate 1-phosphotransferase	55168177
SCCCCL4013C08.g	VIII (1)	ESTB45	m4D1	OSJNBa0032F06.16	50929735
SCCCLR1072G02.g	XIV (1)	ESTB47	m3D1	Putative zinc finger transcription factor	50933107
SCAGLR2011C01.g	XIV (2)	ESTB56	m4D1, m3D1	Glycine-rich RNA-binding protein-like	50934801
SCJFRZ2013F12.g	IX (5)	**ESTB60	m8D2, m4D2, m5D2, m3C, m2C, m6C	OJ000114_01.16	50924506
SCEZRZ1014F04.g	VI (4)	ESTB63	m3D2, m1C, m5D1, m3D1	Mitogen activated protein kinase 6	37594657
SCQSRT2033C08.g	IX (1)	ESTB64	m4D2	RNA-directed DNA polymerase HMG-I and HMG-Y, DNA-binding	87162631

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]
SCCCLR1072G12.g	XIV (2)	ESTB65	m2D2, m1D2, m3D1	Putative ribosomal protein L32	50947773
SCQGSD1045H12.g	IX (3)	**ESTB67	m8D2, m5D1, m2D1, m4D1, m1C	Putative kafirin preprotein	22208459
SCQGSD1045H12.g	IX (3)	ESTB68	m1D2, m2D2, m7D1	Putative kafirin preprotein	22208459
SCQGSD1045H12.g	XIV (2)	ESTB69	m1D2, m2D1	Putative kafirin preprotein	22208459
SCJFRT1059E03.g	III (1)	ESTB75	m2C	Hypothetical protein	34902434
SCJLLR1033A04.g	X (2)	ESTB80	m2D2, m4DC	Ankyrin-like protein	50902216
SCSGFL4C02D07.g	VI (1)	ESTB82	m2C	Putative peroxidase	34913714
SCEZRZ1014H07.g	IX (2)	ESTB92	m2D2, m1D2	Putative sugar transporter	22208506
SCCCFL3002B02.g	III (3)	ESTB94	m4C, m6D1, m2D2	No hit	
SCCCLR1C07C03.g	XIV (1)	ESTB99	m2D1	No hit	
SCEZHR1048C09.g	III (4)	ESTC01	m4D1, m5D1, m3C, m6D1	Putative cysteine protease	50918779
SCUTST3087G12.g	X (3)	ESTC05	m1C, m2C, m7D2	Putative cytochrome B5	34905998
SCEZFL4043H02.g	X (1)	ESTC07	m2D1	Hypothetical protein	9757905
SCUTST3092H12.g	VI (4)	ESTC109	m4D1, m5D2, m1D2, m2D2	Hypothetical protein	34903888
SCRFRZ3058E03.b	XIV (1)	ESTC110	m3D1	ABA 8'-hydroxylase 1	81362266
SCCCST3C02F09.g	XII (1)	ESTC113	m2D1, m1D2	Putative thaumatin-like protein	50726592
SCCCLR2004B05.g	I (1)	ESTC114	m1D2	Phospholipase, putative	77548280
SCBFRZ2017D04.g	XII (1)	ESTC115	m1D1	U2AF small subunit	68036691
SCSBSB1095H03.g	V (1)	ESTC119	m1D1	Putative leucine-rich repeat transmembrane protein kinase	50905839
SCCCLB1C04C09.g	XIV (1)	ESTC12	m1C	No hit	
SCEQLR1007E11.g	XIV (1)	ESTC120	m2D2	Putative uracil phosphoribosyltransferase	50906841
SCBGLR1027H03.g	V / X (3)	**/*ESTC123	m2C, m1C, m3D1, m4D1	mLIP15	1060935
SCSGLR1045E07.g	V (4)	ESTC129	m4C, m2C, m1D1, m5C	Putative leucine-rich repeat transmembrane protein kinase	50905839
SCCCLR2C01B03.g	I (1)	ESTC13	m1D2	Putative actin-depolymerizing factor	34897072
SCMCFL5004A10.g	III (1)	ESTC130	m2D2	Glycine-rich RNA-binding protein	10799202
SCJFRZ2033G09.g	III (2)	ESTC131	m2D2, m1D2	Glycine-rich RNA-binding protein	10799202
SCJLHR1027H06.g	III (1)	ESTC132	m1D2	Glycine-rich RNA-binding protein	10799202
SCQSLB1049C05.g	III (1)	ESTC133	m1D2	Glycine-rich protein	2196542
SCQGAM2026F07.g	III (1)	ESTC134	m2D2	Glycine-rich RNA-binding protein	10799202
SCACLR1057A06.g	X (3)	ESTC19	m4C, m3D2, m1C	P0518C01.24	34906022
SCBGLR1114E07.g	V (3)	ESTC21	m1D1, m2D1, m3C	Lipid transfer protein	311331
SCJFRZ2033H11.g	X (4)	ESTC22	m4D2, m5D2, m2C, m7D1	Hypothetical protein	50939961
SCQGST1032E05.g	XI (2)	ESTC24	m3D1, m1D2	General transcription factor TFIIB	18481632
SCJFRT2060G09.g	VIII (3)	ESTC25	m2D1, m1C, m3D1	CAA303719.1 protein	5777631
SCCCRZ2C01E09.g	XIV (1)	ESTC30	m3D1	Mitochondrial uncoupling protein 4	51860691
SCJLRT3077F09.b	XIV (1)	ESTC39	m4D1	Putative glycosyltransferase	50939609

Cluster[a]	HG[b]	ESTs[c]	Marks[d]	Expected EST homology[e]	GI[f]
SCCCRZ1C01B10.g	V (2)	ESTC42	m1D2, m2D2, m3D1	Putative receptor-like kinase Xa21-binding protein 3	52353644
SCQGST1034A05.g	XIV (1)	ESTC44	m1D2	B1066G12.15	34904540
SCMCST1052A09.g	X (3)	ESTC45	m4D2, m6D1, m2C	SbPCL1	71067066
SCCCCL3005C08.b	III (2)	ESTC47	m3D1, m1D2	Putative acetyl-CoA synthetase	26451642
SCEPFL4177D11.g	X (3)	ESTC48	m3D2, m4C, m5D1	P0518C01.24	34906022
SCEPLR1051G11.g	V (1)	ESTC49	m1C	mLIP15	1060935
SCEZRZ3050D12.g	X (2)	ESTC50	m1D2, m2D2	At1g68060/T23K23_9	23463063
SCACRZ3108F03.g	III / VIII (4)	**/*ESTC57	m3D2, m8C, m1D2, m6D1, m7D1	OSJNBa0072K14.5	50923903
SCVPLR1049C03.g	XIV (1)	ESTC58	m2D1	Methylmalonate semi-aldehyde dehydrogenase	50934827
SCCCLR1C04G09.g	X (1)	ESTC60	m7D1	NADP-specific isocitrate dehydrogenase	34911932
SCMCFL5009H03.g	XIII (1)	**ESTC62	m4D2, m1D2, m5C, m2C	Glyceraldehyde-3-phosphate dehydrogenase	474408
SCJLFL3018F04.g	X (1)	ESTC65	m1D2	Galactosyl transferase GMA12/MNN10 family, putative	77551427
SCRURT2009A09.g	XIV (1)	ESTC77	m4D1	Putative caffeoyl-CoA O-methyltransferase 1	50947279
SCJFST1016B06.g	X (3)	**ESTC80	m2D2, m5D1, m6D1, m3D2	Glycosyltransferase	63087728
SCJLLR1101C01.g	III (4)	ESTC81	m3C, m2C, m1D2, m4D2	Similar to ubiquinol cytochrome c reductase	34898476
SCAGLR1021D10.g	I (2)	ESTC91	m1D2, m2D2	Putative fructose-1,6-bisphosphatase	50931573

[a] EST *clusters* of SUCEST database (http://sucest.lad.ic.unicamp.br/en/)
[b] Homology groups (HGs) – indicated in *parentheses* the number of co-segregation groups (CGs) where each EST-derived markers was assembled.
[c] EST-RFLPs: Est1, Est2, Est3, Est7, Est8 and Est9. EST-SSRs: ESTA, ESTB and ESTC refer, respectively, to di-, tri- and tetranucleotide primer pairs.
[d] For EST-SSRs: mark names present allele number and their type of cross (D1, D2 or C), while for the EST-RFLPs, mark names are the same as described in the *materials and methods* section.
[e] Best hit description
[f] GenBank identification
* EST-SSR duplicated between HGs.
** EST-SSR duplicated in CGs. Other type markers also duplicated in CGs: SG50(HGI), SG94 and SG41(HGII), CIR35(HGVI), CIR21(HGVII), CIR36(HGX), CIR32(HGXI), SMC236(HGVIII) and SMC1047(HGIX).

Discussion

We have exploited the SUCEST database to generate a functionally associated marker-based genetic linkage map for a sugarcane mapping population derived from a cross between two pre-commercial cultivars (SP80-180 and SP80-4966). Expressed sequences from sugarcane that had been classified based on sequence analysis were used to develop EST-RFLP and EST-SSR markers. These markers were located in the previous genetic map performed on this population (Garcia et al. 2006). The coverage and marker density of this genetic map was increased with the integration of these EST-derived markers. The map, generated using the method proposed by Wu et al. (2002), has 307 more markers than the previous one and enhanced genome coverage in 3658.7 cM with the new EST-based co-segregation groups. The total length of the current map is 6251.1 cM with a substantial percentage of unlinked markers (60%). The map is clearly not saturated, given the small size of many co-segregation groups and the numerous unlinked markers. Besides, the number of co-segregation groups (192 CGs) exceeds chromosome number expected for modern cultivars (2n = 100-130). This indicates that gaps remain on most if not all chromosomes, showing that the map is already incomplete. Since we are constrained to discard markers in multiples doses, these gaps are evidently expected.

The number of unlinked markers is smaller than the number obtained in the previous genetic map of this population (68%), however higher than that obtained in other maps for sugarcane (Reffay et al. 2005; Aitken et al. 2005; Raboin et al. 2006). Nevertheless, the current map was generated from a cross between two interespecific modern sugarcane cultivars, aneuploids and with a very complex genetic system, what difficult linkage between markers and, as a result, the loss of informative markers for mapping can be

expected. The genome of modern cultivar needs to be understood in relation to the genome composition of its *S. officinarum* (2n = 80) and *S. spontaneum* (2n = 40-128) progenitors. Structural differences are present between the two genomes (Butterfield *et al*. 2001) as differences in base number are observed (Butterfield et al. 2001). Actually, chromosome-pairing behavior has not been definitively clarified and information on the structure and organization of the genome has been largely speculative. So, difficulties in mapping are expected in this type of population.

Comparison with other sugarcane results (Reffay et al. 2005; Aitken et al 2005; Raboin et al. 2006) has shown that a higher number of double single-dose markers (52% of segregating markers) segregated in the SP80-180 x SP80-4966 population. Indeed, SP80-180 and SP8049-66 have some ancestors in common in the first generations, for instance, POJ2364 [Kassoer x POJ100]. It is known that today's sugarcane cultivars are almost exclusively backcross derivatives involving a very small number of parental clones (Arceneaux 1965; Price 1965). Therefore, a narrow genetic base of modern hybrid varieties is surely one of the principal causes of the present slow rate of sugarcane breeding progress (Berding and Roach 1987). Thus, regarding this, it is expected that populations generated from crosses between modern cultivars present a high number of double single-dose markers.

Although less informative, the C type marker can establish linkage relationship between markers originated from each parent (D1 and D2 types). In fact, to construct an integrated genetic map in sugarcane, a large number of these markers should be available (Garcia et al. 2006). Hence, the high number of C type markers in our population enhanced the construction of the integrated map. However, as in the previous map, linkage between groups with D1 markers (SP80-180 origin) and with D2 markers (SP80-4966 origin),

through C markers, were observed in only a few groups: CGVIII-2 D1 type markers were linked with D2 type markers; the co-segregation groups CGXIV-1, CGIII-1, CGIII-6, CGVI-1 were also made up of all marker types, however, due to insufficient linkage information, they were shared in two groups ('a' and 'b'). Due to the same reason, CGIX-3 was split into two groups (CGIX-3a and CGIX-3b), albeit only D2 and C type markers were assembled.

Previous maps of sugarcane cultivars have identified ten (Grivet et al. 1996), eight (Hoarau et al. 2001; Reffay et al. 2005; Aitken et al 2005) and seven (Rossi et al. 2003) homology groups. But for all of them, homology groups were established separately from each parent. In this study, co-segregation groups of both parents were used to assign the fourteen putative homology groups of this integrated map. Differences in chromosome structure between the progenitor species and pairing behavior in modern varieties suggest that in sugarcane hybrids, the hybrid monoploid number is likely to be greater than 10 (Butterfield et al. 2001). As homology groups represent basic number of chromosomes, the number of HGs achieved in this study is in agreement with the estimated number. The 14HGs showed differences in coverage and in CG number across them. However, due to aneuploidy, an unequal number of chromosomes in each HG is likely to occur. The discrepancy of coverage between HGs highlights the difficulty in mapping large parts of the genome.

Some co-segregation groups contained duplicated loci. Little is known about duplication of genome segments within the monoploid chromosomes of *Saccharum*, since mapping strategy on single-dose markers hinders the identification of the duplicated regions. However, duplicated regions of the genome could be identified within CGs and between HGs, in the current map. These regions remain undetected in relatively low-

density maps, but in many sugarcane maps, they were mentioned. Ming et al. (1998) detected duplications within and across the co-segregation groups. Jannoo et al. (1999) observed 8 cases of duplication to different linkage groups in R570. The proportion of duplicated regions in the monoploid hybrid genome is potentially higher than in ancestral species.

To date, none of the published genetic maps of sugarcane are saturated. No more than one third of the genome is estimated to be tagged on even the most refined maps. As for maps of current cultivars, marker coverage is uneven, with *S. spontaneum* chromosomes being covered more densely than those of *S. officinarum* (Grivet et al. 2001). These maps were constructed for different population types, with anonymous markers. Genetic maps based on expressed sequences have been constructed in a variety of other agronomically important species such as rice (Kurata et al. 1994; Wang et al. 2005), maize (Chao et al. 1994; Falque et al 2005), sugar beet (Schneider et al. 1999), potato (Chen et al. 2001; Feingold et al 2005), barley and wheat (Holton et al. 2002), perennial ryegrass (Faville et al. 2004) and cotton (Han et al. 2006).

Although in terms of SSR discovery, development of functional primer pair, and polymorphism detection, EST-SSR markers are less efficient than DNA-derived SSRs, these caveats are balanced by the relatively low expense with EST-SSR development as a product of a genomic database (Faville et al. 2004). In addition the value of EST-SSRs is enhanced by their superior transferability across taxon boundaries, as demonstrated in grape (Arnold et al. 2002; Decroocq et al. 2003), in white clover (Griffiths et al. 2002) and even in sugarcane (Cordeiro et al. 2001) and their potential as 'perfect markers' for functionally defined genes involved in determining agronomic traits.

The fact that ESTs containing SSRs exhibited sequence similarity to genes with a wide range of functions suggests that there is potential to identify EST-SSRs that may be directly involved in determining agronomically important traits. For example, there are 15 ESTs with strong homology to sugarcane metabolism proteins. The identification of genes that determine economically important plant traits provides important tools to further manipulate plant function and performance, through enhanced conventional breeding using the gene's DNA sequence as a 'perfect' marker for trait selection. The development of a functionally defined gene-based genetic map of sugarcane provides the basis for the correlation of molecular variation associated with functional sequences with the locations of QTLs for putatively related traits. QTL analysis using EST-RFLP and EST-SSR allows the identification of associations between functionally associated marker locations and QTLs. However, a large number of markers are necessary to build a genetic map and to obtain sufficient genome coverage for QTL analysis, due to the large number of chromosomes in sugarcane. Although the current map is not well saturated, the genome coverage of this map will probably facilitate the detection of QTL for important traits in sugarcane, because the addition of known-function gene markers could greatly enhance the utility of a genetic map, as it facilitates the transition from genetic linkage analysis to a candidate gene mapping approach to dissect complex traits.

Acknowledgements

This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 02/01167-1) and from CTC (Centro de Tecnologia Canavieira – Piracicaba / SP). K.M.O. and M.M.P. received doctorate fellowships, respectively, from

FAPESP (02/00197-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); K.M.O. also received PDEE fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, 0181-05-6); L.H.M. received an MS fellowship from FAPESP (03/07960-6); TGM and GRAM received IC fellowships from FAPESP (04/10596-9 and 05/59268-6, respectively); L.R.P. received a Postdoctorate fellowship from FAPESP (01/14656); A.P.S., A.A.F.G. and A.F. received a research fellowship from CNPq. The authors would like to thank Dr. Angélique D'Hont (*Centre de Coopération Internationale en Recherche Agronomique pour le Développement*) for valuable discussions.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110: 789-801.

Al-Janabi SM, Honeycutt RJ, McClelland M, Sobral BWS (1993) A genetic linkage map of Saccharum spontaneum L. 'SES208'. Genetics 134: 1249-1260.

Arceneaux G (1965) Cultivated sugarcanes of the world and their botanical derivation. Proc Int Soc Sugar Cane Technol 12: 844-854.

Arnold C, Rossetto M, McNally J, Henry RJ (2002) The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in Vitaceae. Am J Bot 89: 22-28.

Berding N, Roach BT (1987) Germplasm collection, maintenance and use. In: D.J. Heinz (Ed). Sugarcane improvement through breeding. Elsevier press, Amsterdam, pp 143-210.

Butterfield M, D'Hont A, Berding N (2001) The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc Soc Afr Sugarcane Technol Ass 75: 1-5.

Carson DL, Botha FC (2000) Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769-1779.

Casu R, Dimmock C, Thomas M, Bower N, Knight D, Grof C, McIntyre L, Jackson P, Jordan D, Whan V *et al* (2001) Genetic and expression profiling in sugarcane. Proc Int Soc Sugarcane Technol 24: 626-627.

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102: 296-306.

Chao S, Baysdorfer C, Heredia-Diaz O, usket T, Xu G, Coe EHJr (1994) RFLP mapping of partially sequenced leaf cDNA clones in maize. Theor Appl Genet 88: 717-721.

Chen C, Zhou P, Choi YA, Huang S, Jr-Gmitter FG (2006) Mining and characterizing microsatellites from citrus ESTs.

Chen X, Salamini F, Gerhard C (2001) A potato molecular-function map for carbohydrate metabolism and transport. Theor Appl Genet 102: 284-295.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Sci 160: 1115-1123.

Creste S, Tulmann-Neto A, Figueira A (2002) Detection of single repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. Plant Mol Biol Rep. 19: 229-306.

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterization of the double-genome structure of modern sugarcane cultivars (Saccharum ssp) by molecular cytogenetics. Mol Gen Genet 250: 405-413.

D'Hont A, Lu YH, Gonzalez de Leon D, Grivet L, Feldmann P, Lanaud C, Glaszmann JC (1994) A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome 37: 222-2230.

da Silva JAG, Sorrells ME, Burnquist W, Tanksley SD (1993) RFLP linkage map and genome analysis of Saccharum spontaneum. Genome 36: 782-791.

Decroocq V, Favé M, Hagen L, Bordenhave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106: 912-922.

Doerge RW (1996) Constructing genetic maps by rapid chain delineation. J. Agric. Genomics 2: 6.

Eujayl I, Sledge MK, Wang L, May GD, Chekovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncata* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108: 414-422.

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene Loci. Genetics 170(4): 1957-1966.

Faville MJ, Vecchies AC, Schreiber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster JW (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110: 12-32.

Feinberg AP, Volgelstein B (1983) A technique for radiolabelling DNA restriction fragments to high specific activity. Analitical Biochemistry 132: 6-13.

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (Solanum tuberosum L.). Theor Appl Genet. 111(3): 456-66.

Gao LF, Jing RL, Huo NX, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108: 1392-1400.

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, Souza A (2006) Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112: 298-314.

Griffiths AG, Barrett BA, Ellison N, Sawbridge T, Spangenberg G, Bryan GJ, Schreiber M (2002) Getting out there: transportability of white clover EST and genomic DNA-derived microsatellites. In: Proceedings of the 12th Australasian plant breeding conference, Perth, 15-20 September 2002, pp 69-72.

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

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp): genome organization in a highly polyploid and anueploid interspecific hybrid. Genetics 142: 987-1000.

Guimarães CT, Honeycutt RJ, Sills GR, Sobral BWS (1999) Genetic maps of *S. officinarum* L and *S. robustum* Brandes & Jew. Ex Grassl. Genet Mol Biol 22: 125-132.

Gupta PK, Rutsgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270: 315-323.

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Gen Genomics. 272: 308-327.

Han ZG, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Grossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112: 430-439.

Hoarau J-Y, Offmann B, D'Hont A, Risterucci A-M, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp). I. Genome Mapping with AFLP markers. Theor Appl Genet 103: 84-97.

Hoarau J-Y, Souza Glaucia, D'Hont A, Menossi M, Pinto LR, Souza AP, Grivet L, Menck CFM, Ulian EC, Vincentz M (2006) Sugarcane, a tropical crop with a highly complex genome. In: J. F. Morot-Gaudry, P.Lea and J.F. Briat. (Org.). Plant genomics (*in press*).

Hoisington D, Khairallah M, González-de-Léon D (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. México: CYMMYT, 51p.

Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9: 63-71.

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among sugarcane cultivars. Theor Appl Genet 99: 1053-1060.

Kantley RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48: 501-510.

Kosambi DD (1944) The estimation of map distance from recombination values. Ann. Eug. 12: 172-175.

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin S-Y, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300-kilobase-interval genetic map of rice including 883 expressed sequences. Nat Genet 8: 365-372.

Ma HM, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A (2004) An EST survey of the sugaracane transcriptome. Theor Appl Genet 108: 851-863.

Ming R, Liu SC, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore P, Burnquist W, Sorrells ME, Irvine J, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150: 1663-1682.

Ming R, Liu SC, Bowers JE, Moore PH, Irvine JE, Paterson AH (2002) Construction of a *Saccharum* consensus genetic map from two interspecific crosses. Crop Sci 42: 570-583.

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J3: 175-182.

Mudge J, Andersen WR, Kehrer RL, Fairbanks DJ (1996) A RAPD genetic map of Saccharum officinarum. Crop Science, 36: 1362-1366.

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109: 800-805.

Pinto LR, Oliveira KM, Marconi TG, Garcia AAF, Ulian EC, Souza AP (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breeding (*in press*).

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, Souza AP (2004) Survey in the expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47: 795-804.

Piperidis G, Christopher MJ, Carroll BJ, Berding N, D'Hont A. (2001) Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species Erianthus arundinaceus. Genome 43: 1033-1037.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1: 215–221.

Price S (1965) Interespecific hybridizations in sugarcane breeding. Proc Int Soc Sugar Cane Technol. 12: 1021-1026.

Qureshi SN, Saha S, Kantley RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8: 112-123.

Raboin L-M, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau J-Y, D'Hont A (2006) Genetic mapping in the high polyploid sugarcane using a bi-parental progeny: identification of a gene controlling stalk colour and new rust resistance gene. Theor Appl Genet 112(7): 1382-1391.

Reffay N, Jackson PA, Aitken KS, Hoarau J-Y, D'Hont A, Besse P, McIntyre CL (2005) Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. Mol Breed 15: 367-381.

Roach BT (1969) Cytological studies in *Saccharum*. Chromosome transmission in interspecific and intergeneric crosses. Proc Int Soc Sugar Cane Technol 13: 901-920.

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys M-A, D'Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genomics 269: 406-419.

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy UK, Kantery RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130: 355-364.

Schneider K, Borchardt DC, Schäfer-Pregl R, Nagl N, Glass C, Jeppsson A, Gebhardt C, Salamini F (1999) PCR-based cloning and segregation analysis of functional gene homologues in *Beta vulgaris*. Mol Gen Genet 262: 515-524.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant Journal 3:739-744.

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

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. Genome Res 13: 2725-2735.

Wang XS, Zhu J, Mansueto L, Bruskiewich R (2005) Identification of candidate genes for drought stress tolerance in rice by the integration of a genetic (QTL) map with the rice genome physical map. J Zhejiang Univ Sci B.6(5): 382-388.

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor appl genet 83:294-300

Wu R, Ma CX (2002) Simultaneous maximum likelihood estimation of linkage and linkage phase in outcrossing species. Theor Pop Bio 61: 349-363.

7 Conclusões e Perspectivas

7.1 Desenvolvimento dos marcadores EST-SSRs

(1) A busca no banco de dados do SUCEST revelou um total de 2005 *clusters* contendo seqüências repetitivas (SSRs). Este número representa apenas 4.6% da população total de *clusters* do banco, porém está de acordo com os números encontrados em trabalhos semelhantes.

(2) A abundância de SSRs nas regiões transcritas do genoma da cana-de-açúcar faz deste banco de ESTs uma valiosa fonte no desenvolvimento destes marcadores moleculares, de maneira simples, rápida e econômica.

(3) A avaliação do polimorfismo dos EST-SSRs e a comparação destes com gSSRs estão de acordo com os resultados apresentados na literatura. Os EST-SSRs apresentaram níveis de polimorfismos mais baixos, no entanto, o poder discriminatório dos dois tipos de SSR não diferiram significativamente. Além disso, o fato dos EST-SSRs apresentarem homologia a genes contrabalança o baixo nível de polimorfismo, contribuindo para o conhecimento da diversidade genética na cana-de-açúcar e estendendo-o ao nível funcional.

(4) Trezentos e setenta e dois *clusters* foram selecionados para o desenvolvimento dos EST-SSRs. A quantidade de informação ainda existente no banco de dados possibilita a continuidade do desenvolvimento destes marcadores, que são de grande importância em

estudos de mapeamento genético e de QTL's, uma vez que possuem significativa homologia a proteínas envolvidas na expressão de características de interesse econômico.

7.2 Mapeamento genético dos EST-SSRs

(1) Os 576 marcadores EST-SSRs gerados a partir de 149 locos EST-SSRs, juntamente com marcadores AFLP, gRFLP, gSSR e EST-RFLP previamente mapeados na população 'SP80-180 x SP80-4966', promoveram a elaboração de um mapa genético com uma cobertura do genoma da ordem de 6261.1cM, representando o primeiro mapa funcional em cana-de-açúcar e o primeiro mapa genético integrado construído a partir de uma população resultante de um cruzamento entre duas variedades comerciais de cana-de-açúcar.

(2) A adição dos EST-SSRs possibilitou o aumento da cobertura do genoma, porém o grande número de marcadores não-ligados, aliado ao pequeno tamanho da maioria dos grupos de co-segregação e o número reduzido de marcadores por grupo de co-segregação indicam que o mapa não está saturado.

(3) A população de mapeamento avaliada neste trabalho apresenta um sistema genético muito complexo, por derivar de um cruzamento entre dois híbridos interespecíficos. Sabe-se que este tipo de híbrido poliplóide tem aneuploidia e que seus cromossomos não se pareiam na meiose,

dificultando a ligação dos marcadores. Apesar disso, foi importante a exploração deste tipo de população em um programa de mapeamento genético, pois foi a primeira vez que tal tipo de população foi analisada.

(4) Não é possível acessar o genoma completo desta população utilizando-se poucos marcadores. O elevado número de cromossomos da cana-de-açúcar exige que seja utilizado um grande número de marcadores visando à obtenção de um número de grupos de co-segregação correspondente ao número de cromossomos da cana-de-açúcar. Desta forma, a continuidade de desenvolvimento de EST-SSRs é de grande importância na construção do mapa funcional desta população, de modo que este apresente melhor cobertura do genoma.

(5) A adição de marcadores com homologia a genes de função conhecida aumentou a utilidade do mapa genético, pois estes facilitarão a transição das análises de ligação genética à estratégia de genes candidatos na análise de características complexas.

(6) A população de mapeamento será aumentada e o mapeamento dos QTL's será realizado empregando-se um novo programa em desenvolvimento, o qual reduzirá as dificuldades de mapeamento neste tipo de população, o que proporcionará uma melhor precisão nos resultados.

(7) Pretende-se também realizar estudos de interações epistáticas entre os locos e de pleitotropia, além de analisar interações QTL x ambiente.

8 Referências Bibliográficas

Aitken K.S., Jackson P.A., McIntyre C.L. (2005). A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110: 789-801.

Alix K., Baurens F.C., Paulet F., Glaszmann J.C., D'Hont A. (1998). Isolation and characterisation of a satellite DNA family in *Saccharum* complex. Genome 41(6): 854-864.

Alix K., Paulet F., Glaszmann J.C., D'Hont A. (1999). Inter-Alu like species-specific sequences in the *Saccharum* complex. Theor Appl Genet 6: 962-968.

Al-Janabi S.M., Honeycutt R.J., McClelland M., Sobral B.W.S. (1993). A genetic linkage map of *Saccharum spontaneum* L. 'SES208'. Genetics 134: 1249-1260.

Al-Janabi S.M., Honeycutt R.J., Sobral, B.W.S. (1994a). Chromosome assortment in *Saccharum*. Theor Appl Genet 89: 959-963.

Al-Janabi S.M., McClelland M., Petersen C., Sobral B.W.S. (1994b). Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: Saccharinae. Theor Appl Genet 88: 933-944.

Allard R. W. (1960). Principles of Plant Breeding. John Wily & Sons, New York, N. Y. 485p.

Arumuganathan K., Earl. E.D. (1991). Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208-218.

Asnaghi C., Paulet F., Kaye C., Grivet L., Deu M., Glaszmann J.C., D'Hont A. (2000). Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. Theor Appl Genet 101: 962-969.

Asnaghi C., Roques D., Ruffel S., Kaye C., Hoarau J.Y., Telismart H., Girard J.C, Raboin L.M., Risterucci A.M., Grivet L., D'Hont A. (2004). Targeted mapping of a sugarcane rust resistance gene (Bru1) using bulked segregant analysis and AFLP markers. Theor Appl Genet (108): 759-764.

Bearzoti E. (2000). Mapeamento de QTL. In: 2ª Jornada em Genética e Melhoramento d Plantas, Goiânia, GO, 2000. Análise de QTL no Melhoramento de Plantas. Goiânia: FUNAPE, p.63-223.

Bell C.J., Ecker J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabdopsis*. Genomics 19: 137-144.

Besse P., McIntyre C.L., Berding N. (1996). Ribosomal DNA variations in *Erianthus*, a wild sugarcane relative (Andropogoneae x Saccharinae). Theor Appl Genet 92: 733-743.

Besse P., McIntyre C.L., Berding N. (1997). Characterization of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae x Saccharinae) using RFLP markers. Euphytica 93: 283-292.

Bhat S.R., Gill B.S. (1985). The implication of 2n egg gametes in nobilisation and breeding of sugarcane. Euphytica 34: 377-384.

Brandes E.W., Sartoris G.B., Grassl C.O. (1939). Assembling and evaluating wild forms of sugarcane and closely related plants. Proc. Int. Soc. Sugarcane Technol. 6: 128-154.

Brandes E.W. (1956). Origin, dispersal and use in breeding of the Melanesian garden sugarcanes and their derivatives *Saccharum officinarum* L. Proc. Cong. Int. Soc. Sug. Tecnhol. 9(1): 709-750.

Brandes E.W. (1958). Origin, classification and characteristics. In: Artschwager E. Brandes E.W. (eds) Sugarcanes (*Saccharum officinarum* L.). USDA Agriculture Handbook no. 122, pp 1-35.

Bremer G. (1924). A cytological investigation of some species and species hybrids within the genus *Saccharum*. Genetica 5: 97-148; 273-326.

Bremer G. (1930). The cytology of *Saccharum*. Proc Int Soc Sugar Cane Technol 3: 408-415.

Bremer G. (1961a). Problems in breeding and cytology of sugarcane .I. A short history of sugarcane breeding – the original forms of *Saccharum*. Euphytica 10: 59-78.

Bremer G. (1961b). Problems in breeding and cytology of sugarcane. II. The sugarcane breeding through a cytological viewpoint. Euphytica 10: 121-133.

Bremer G. (1961c). Problems in breeding and cytology of sugarcane. III. The cytological crossing research of sugarcane. Euphytica 10: 229-243.

Brook J.D., McCurrach M.E., Harley H.G., Buckler A.J., Church D., Aburatani H., Hunter K., Stanton V.P., Thirion J.P., Hudson T. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell 68: 799-808.

Burner D.M. (1987). Cytogenetic analyses of sugarcane relatives (Andropogoneae: Saccharinae). Euphytica 54: 125-133.

Burnquist W.L. (1991). Development and the application of restriction fragment length polymorphism technology in sugarcane (*Saccharum ssp*) breeding. PHD thesis. Cornell University, Ithaca, N. Y., USA.

Burnquist W.L., Sorrells M.E., Tanksley S. (1992). Characterisation of genetic variability in *Saccharum* germplasm by means of restriction fragment length polymorphism (RFLP) analysis. Proc. Int. Soc. Sugar Cane Technol 21(2): 355-365.

Burnquist W.L. (2000). Os arquitetos da nova cana. Revista Pesquisa Fapesp 59, p3-4.

Butterfield M., D'Hont A., Berding N. (2001). The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc Soc Afr Sugarcane Technol Ass 75: 1-5.

Camargo L.E.A. (2000). Mapeamento de QRL's: contribuições e perspectivas. In: 2ª Jornada em Genética e Melhoramento de Plantas, Goiânia, GO. Análise de QTL no Melhoramento de Plantas. Goiânia: FUNAPE, p.49-61.

Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., Waugh R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156: 847-854.

Carneiro M.S., Vieira M.L.C. (2002). Mapas genéticos em plantas. Bragantia 61(2): 89-100.

Carson D.L., Botha F.C. (2000). Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 40:1769-1779.

Carson D.L., Botha F.C. (2002). Genes expressed in sugarcane maturing internodal tissue. Plant Cell Rep 20: 1075-1081.

Casu R., Dimmock C., Thomas M., Bower N., Knight D., Grof C., McIntyre L., Jackson P., Jordan D., Whan V., Drenth J., Tao Y., Manners J. (2001) Genetic and expression profiling in sugarcane. Proc Int Soc Sugarcane Technol 24: 626-627.

Casu R., Manners J., Bonnett G., Jackson P., McIntyre C., Dunne R., Chapman S., Rae A., Grof C. (2005). Genomics approaches for the identification of genes determining important traits in sugarcane. Field Crops Research 92: 137-147.

Cato S.A., Gardner R.C., Kent J., Richardson T.E. (2001). A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102: 296-306.

Chao S., Baysdorfer C., Heredia-Diaz O., Usket T., Xu G., Coe E.H.Jr (1994). RFLP mapping of partially sequenced leaf cDNA clones in maize. Theor Appl Genet 88: 717-721.

Chen X., Salamini F., Gerhard C. (2001). A potato molecular-function map for carbohydrate metabolism and transport. Theor Appl Genet 102: 284-295.

Chen C., Zhou P., Choi Y.A., Huang S., Jr-Gmitter F.G. (2006). Mining and characterizing microsatellites from citrus ESTs.

Cho Y.G., Ishii T., Temnykh S., Chen X., Lipovich L., McCouch S.R., Park W.D., Ayres N., Cartinhour S. (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (Oryza sativa L.). Theor Appl Genet 100: 713-722.

Conab (2006). www.conab.gov.br

Cordeiro G.M., Taylor G.O., Henry R.J. (2000). Characterization of microsatellite markers from sugarcane (*Saccharum* spp.), a highly polyploid species. Plant Science 155: 161-168.

Cordeiro G.M., Casu R., McIntyre C.L., Manners J.M., Henry R.J. (2001). Microsatellite markers from sugarcane (*Saccharum* spp) ESTs cross transferable to *Erianthus* and *Sorghum*. Plant Sci 160: 1115-1123.

Cordeiro G.M., Pan Y.B., Henry R.J. (2003). Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. Plant Science 165: 181-189.

D'Hont A., Lu Y.H., Feldmann P., Glaszmann J.C. (1993). Cytoplasmic diversity in sugarcane revealed by heterologous probes. Sugar Cane 1: 12-15.

D'Hont A., Lu Y.H., Gonzalez de Leon D., Grivet L., Feldmann P., Lanaud C., Glaszmann J.C. (1994). A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome 37: 222-2230.

D'Hont A., Rao S., Feldman P., Grivet P., Islam-Faridi L., Berding N., Glaszmann J.C. (1995). Identification and characterization of intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and in situ hybridization. Theor Appl Genet 91: 320-326.

D'Hont A., Grivet L., Feldmann P., Rao S., Berding N., Glaszmann J.C. (1996). Characterization of the double-genome structure of modern sugarcane cultivars (*Saccharum* spp) by molecular cytogenetics. Mol Gen Genet 250: 405-413.

D'Hont A., Ison D., Alix K., Roux C., Glaszmann J.C. (1998). Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41: 221-225.

D'Hont A., Glaszmann J.C. (2001). Sugarcane genome analysis with molecular markers, a first decade of research. Proc. Int. Soc. Sugarcane Technol. 24: 556-559.

D'Hont A., Garsmeur O., Raboin L.M., Paulet F., Begum D., Wing R., Glaszmann J.C. (2001). Chromosome walking toward a major resistance gene for common rust of sugarcane. Proc Int Soc Sugarcane Technol 24: 315-317.

Da Silva J.A.G., Sorrells M.E., Burnquist W., Tanksley S.D. (1993a). RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome 36: 782-791.

Da Silva J.A.G. (1993b). A methodology for genome mapping of autopolyploids and its application to sugarcane (*Saccharum* spp.). Ph.D. dissertation of Cornell University, Ithaca, N.Y. 108p.

Da Silva J.A.G., Honeycutt R.J., Burnquist W.L., Al-Janabi S.M., Sorrells M.E. et al (1995). Saccharum spontaneum L. 'SES 208' genetic linkage map containing RFLP and PCR based markers. Mol Breeding 1: 165-179.

Da Silva J.A.G., Sobral B.W.S. (1996). Polyploid genetics. In: The Impact of Plant Molecular Genetics (Sobral B.W.S., ed.). Birkhauser. Boston. P. 3-35.

Daniels J., Smith P., Paton N., Williams C.A. (1975c). The origin of the genus *Saccharum*. Sugarcane Breed. Newsl. 36: 24-39.

Daniels J., Roach B.T. (1987). Taxonomy and evolution. In: Heinz D.J. (ed). Sugarcane improvement through breeding. Elsevier Press, Amsterdam, pp 7-84.

Daugrois J.H., Grivet L., Roques D., Hoarau J.Y., Lombard H., Glaszmann J.C., D'Hont A. (1996). A putative major gene for rust resistance linked with an RFLP marker in sugarcane cultivar R570. Theor Appl Genet 92: 1059-1064.

Decroocq V., Favé M., Hagen L., Bordenave L., Decroocq S. (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106: 912-922.

Dooner H.K. (1986). Genetic fine structure of the bronze locus in maize. Genetics 86: 1021-1036.

Dufour P., Grivet L., D'Hont A., Deu M., Trouche G., Glaszmann J.C., Hamon P. (1996). Comparative genetic mapping between duplicated segments on maize chromosome 3 and 8 and homoeologous regions in sorghum and sugarcane. Theor Appl Genet 92: 1024-1030.

Dufour P., Deu M., Grivet L., D'Hont A., Paulet F., Bouet A., Lanaud C., Glaszmann J.C., Hamon P. (1997). Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor Appl Genet 94: 409-418.

Eght C.S., Maymarquardt P. (1997). Survey of microsatellite DNA in pine. Genome 40: 9-17.

Edwards K.J., Barker J.H., Daly A., Jones C., Karp A. (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20: 758-760.

Eujayl I., Sledge M.K., Wang L., May G.D., Chekovskiy K., Zwonitzer J.C., Mian M.A.R. (2004). *Medicago truncata* EST-SSRs reveal cross-species genetic markers for Medicago spp. Theor Appl Genet 108: 414-422.

FAPESP. Genoma. São Paulo. 2001. 1v.

Feldmann P., D'Hont A., Guiderdoni E., Grivet L., Glaszmann C. (1997). La canne à sucre. In: L'amélioration des plantes tropicales. Cirad / Orstom, 1997.

Ferreira M.E., Grattapaglia D. (1998). Introdução ao uso de marcadores moleculares em análise genética. 3ª ed. Brasília: Embrapa-Cenargen. 220p.

Fraser L.G., Harvey C.F., Crowhurst R.N., De Silva H.N. (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108: 1010-1016.

Gale M.D., Devos K.M. (1998). Plant comparative genetics after 10 years. Science 282: 656-659.

Ganal M.W., Young N.D., Tanksley S.D. (1989). Pulsed field electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato. Mol Gen Genet 215: 395-400.

Gao L.F., Jing R.L., Huo N.X., Li X.P., Zhou R.H., Chang X.P., Tang J.F., Ma Z.Y., Jia J.Z. (2004). One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108: 1392-1400.

Garcia A.A.F., Kido E.A., Meza A.N., Souza H.M.B., Pinto L.R., Pastina M.M., Leite C.S., da Silva J.A.G., Ulian E.C., Figueira A., Souza A.P. (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112: 298-314.

Gaut B.S., Le Thierry d'Ennequin M., Pekk A.S., Sawkins M.C. (2000). Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 97: 7008-7015.

Glaszmann J.C., Noyer J.L., Fautret A., Feldmann P., Lanaud C. (1989). Biochemical genetic markers in sugarcane. Theor Appl Genet 78: 537-543.

Glaszmann J.C., Dufour P., Grivet L., D'Hont A., Deu M., Paulet F., Hamon P. (1997). Comparative genome analysis between several tropical grasses. Euphytica 96: 13-21.

Glaszmann J.C., Lu Y.H., Lanaud C. (1990). Variation of nuclear ribosomal DNA in sugarcane. J. Genet Breed 44:191-198.

Griffiths A.G., Barrett B.A., Ellison N., Sawbridge T., Spangenberg G., Bryan G.J., Schreiber M. (2002). Getting out there: transportability of white clover EST and genomic DNA-derived microsatellites. In: Proceedings of the 12th Australasian plant breeding conference, Perth, 15-20 September 2002, pp 69-72.

Grivet L., D'Hont A., P. Dufour, P. Hamon, D. Roques D., Glaszmann J.C. (1994). Comparative genome mapping of sugarcane with other species within the Andropogoneae tribe. Heredity 73: 500-508.

Grivet L., D'Hont A., Roques D., Feldmann P., Lanaud C., Glaszmann J.C. (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp): genome organization in a highly polyploid and anueploid interspecific hybrid. Genetics 142: 987-1000.

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

Grivet L., Glaszmann J.C., Arruda P. (2001). Sequence polymorphism from EST data in sugarcane: a fine analysis of 6-phosphogluconate dehydrogenase genes. Genet Mol Biol 24: 161-167.

Grivet L., Glaszmann J.C., Vincentz M., da Silva F., Arruda P. (2003). ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. Theor Appl Genet 106: 190-197.

Grivet L., Daniels, C., Glaszmann J.C., D'Hont A. (2004). A review of recent molecular genetics evidence for sugarcane evolution and domestication. Ethnobotany Research and Applications 2: 9-17.

Guimarães C.T., Sills G., Sobral B. (1997). Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorgum and maize. Proc. Natl. Acad. Sci. 94: 14261-14266.

Guimarães C.T., Moreira M.A. (1999a). Genética molecular aplicada ao melhoramento de plantas. In: Borém, A (ed). Melhoramento de espécies cultivadas. Editora UFV Viçosa-Minas Gerais, p715-740.

Guimarães C.T., Honeycutt R.J., Sills G.R., Sobral B.W.S. (1999b). Genetic maps of *S. officinarum* L and *S. robustum* Brandes & Jew. Ex Grassl. Genet Mol Biol 22: 125-132.

Guimarães, C.T. (1999). Mapeamento comparativo e detecção de QTL's em cana-de-açúcar utilizando marcadores moleculares. Viçosa, 70p. Tese (Doutorado) – Universidade Federal de Viçosa.

Gupta P.K., Rutsgi S., Sharma S., Singh R., Kumar N., Balyan H.S. (2003). Transferable EST-SSR markers for study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270: 315-323.

Gur-Arie R., Cohen C.J., Eitan Y., Shelef L., Hallerman E.M., Kashi Y. (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res 10: 62-7.

Gustafson J.P., Dillé J.E. (1992). Chromosome location of *Orysa sativa* recombination linkage groups. Proc. Natl. Acad. Sci. USA 89: 8646-8650.

Ha S., Moore P., Heinz D., Kato S., Ohmido N., Fukui K. (1999). Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescent in situ hybridisation and imaging methods. Plant Molecular Biology 39: 1165-1173.

Haldane J.B.S. (1919). The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8: 299-309.

Han Z.G., Wang C., Song X., Guo W., Gou J., Li C., Chen X., Zhang T. (2006). Characteristics, development and mapping of *Grossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112: 430-439.

Harvey M., D'Hont A., Alix K., Huckett B. (1998). Use PCR-based markers for identification of *Erianthus* genetic material in putative intergeneric hybrids (*Saccharum* x *Erianthus*). Proc. S. Afr. Sug. Technol. Ass. 72: 218-320.

Heinz D.J., Tew T.L. (1987). Hybridization procedures. In: Heinz DJ ed Sugarcane Improvement Through Breedding. Elsevier, Amsterdam, p 313-342.

Hoarau J.Y. (2000). Cartographie génétique d'un cultivar de canne-à-sucre (*Saccharum*) aux moyens d'AFLPs et marquage de caractères agronomique utiles en sélection. Le titre de Docteur en Biologie et Agronomie de Lénsar, France.

Hoarau J-Y., Offmann B., D'Hont A., Risterucci A-M., Roques D., Glaszmann J.C., Grivet L. (2001). Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp). I. Genome Mapping with AFLP markers. Theor Appl Genet 103: 84-97.

Hoarau J.Y., Grivet, L., Offmann B., Raboin L.M., Diorflar J.P., Payet J. Hellmann M., D'Hont A., Glaszmann J.C. (2002). Genetic dissection of a modern sugarcane cultivar (Saccharum ssp): II. Detection of QTLs for yield components. Theor Appl Genet (105): 1027-1037.

Holton T.A., Christopher J.T., McClure L., Harker N., Henry R.J. (2002). Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9: 63-71.

Jackson R.C., Casey J. (1980). Cytogenetics of polyploids. In: Polyploidy: Biological Relevance, Lewis, W.H., ed. Plenum Press, New York.

Jackson R.C. (1982). Polyploidy and diploidy: New perspectives in chromosome pairing and its evolutionary implications. Amer. J. Bot. 69: 1512-1523.

Jagathesan D., Sreenivasan T.V., Ratnambal M.J. (1970). Chromosome numbers in noble sugarcane. Indian J. Hered. 2: 89-96.

Janaki Amnal E.K. (1936). Cytogenetic analysis of *Saccharum spontaneum* L. 1. Chromosome studies in Indian forms. Indian J. Agric. Sci. 6: 1-8.

Jannoo N., Grivet L., Seguin M., Paulet F., Domaingue R., Rao P.S., Dookun A., D'Hont A., Glaszmann J.C. (1999a). Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99: 171-184.

Jannoo N. (1998). Structure et dynamique du genome des cultivars modernes de canne a sucre (Saccharum spp): apports du marquage moleculaire (RFLP). PhD. Diss Institut National Agronomique Paris-Grignon, Paris.

Jordan D.R., Casu R.E., Besse P., Carroll B.C., Berding N., McIntyre C.L. (2004). Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. Genome (47): 988-993.

Jornal da Cana (2006). www.jornalcana.com.br

Kantety R.V., La Rota M., Matthews D.E., Sorrells M.E. (2002). Data mining for simple sequence repeats in expressed tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48: 501-510.

Keith C.S., Hoang D.O., Barrett B.M., Feigelman B., Nelson M.C., Thai H., Baysdorfer C. (1993). Partial sequence analysis of 130 randomly selected maize cDNA clones. Plant Physiol 101: 329-332.

Kellogg E.A. (2001). Evolutionary history of the grasses. Plant Physiol 125: 1198-1205.

Kosambi D.D. (1944). The estimation of map distance from recombination values. Ann. Eug. 12: 172-175.

Kumar L.S. (1999). DNA markers in plant improvement: an overview. Biotechnology Advances 17: 143–182.

Lagergrantz U., Ellegren H., Anderson L. (1993). The abundance of various polymorphic SSR motifs differs between plants and vertebrates. Nucleic Acid Res 21: 1111-1115.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Lavi U., Akkaya M., Bhagwat A., Lahav E., Cregan P.B. (1994). Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea Americana* M.). Euphytica 80: 171-177.

Li H.W., Price S. (1967). Chromosome numbers of noble sugarcane clones. Proc Int Soc Sugar Cane Technol 12: 884-886.

Li Y.C., Korol A.B., Fahima T., Beiles A., Nevo E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465.

Lima M.L.A., Garcia A.A.F., Oliveira K.M., Matsuoka S., Souza Jr C.L., Souza A.P. (2002). Analysis of genetics similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.). Theor Appl. Genet. 104: 30-38.

Liu B.H. (1998). Statistical Genomics: Linkage, Mapping and QTL Analysis. Boca Raton, FL: CRC Press.

Lu Y.H., D'Hont A., Paulet F., Grivet L., Arnaud M., Glaszmann J.C. (1994a). Molecular diversity and genome structure in modern sugarcane varieties. Euphytica 78: 217-226.

Lu Y.H., D'Hont A., Walker D.I.T., Rao P.S., Feldmann P., Glaszmann J.C. (1994b). Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. Euphytica 78: 7-18.

Ma H.M., Schulze S., Lee S., Yang M., Mirkov E., Irvine J., Moore P., Paterson A. (2004). An EST survey of the sugarcane transcriptome. Theor Appl Genet 108: 851-863.

Mac Key J. (1987). Implications of polyploidy breeding. Biol. Zent. Bl. 106: 257-266.

Machado F.B.P. (2003). Brasil, a doce terra – História do Setor. (www.jornalcana.com.br)

Mather K. (1957). The measurement of linkage in heredity. Methuen & Co. LTD., London.

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

Ming R., Liu S.C., da Silva J., Wilson W., Braga D., van Deynze A., Wenslaff T.F., Wu K.K., Moore P., Burnquist W., Sorrells M.E., Irvine J., Paterson A.H. (1998). Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150: 1663-1682.

Ming R., Liu S.C., Moore P., Irvine J.E., Paterson A.H. (2001). QTL analysis in a complex autopolyploid: genetic control of sugarcane content in sugarcane. Genome 11: 2075-2084.

Ming R., Liu S.C., Bowers J.E., Moore P.H., Irvine J.E., Paterson A.H. (2002a). Construction of a *Saccharum* consensus genetic map from two interspecific crosses. Crop Sci 42: 570-583.

Ming R., Del Monte T.A., Hernandez E., Moore P.H., Irvine J.E., Paterson A.H. (2002b). Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome (45): 794-803.

Morgante M., Olivieri A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. Plant J3: 175-182.

Morgante M., Hanafey M., Powell W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30: 194-200.

Morton N. (1955). Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7: 277-318.

Mudge J., Andersen W.R., Kehrer R.L., Fairbanks D.J. (1996). A RAPD genetic map of *Saccharum officinarum*. Crop Science, 36: 1362-1366.

Murkherjee S.K. (1954). Revision of the genus *Saccharum* Linn. Bull. Bot. Soc. Bengal 8: 143-148.

Murkherjee S.K. (1957). Origin and distribution of *Saccharum*. Bot. Gaz. 119: 55-61.

Naidu K.M., Sreenivasan T.V. (1987). Conservation of sugarcane germplasm. In: Copersucar Int Sugarcane Breed Wkshp, Copersucar, São Paulo, Brazil, pp 33-53.

Nair N., Nair S., Sreenivasan T.V., Mohan M. (1999). Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. Genet. Res. Crop Evol. 46: 73-79.

Newman T., de Bruijn F.J., Green P., Keegstra K., Kende H., McIntoch L., Ohlrogge J., Raikhel N., Somerville S., Thomashow M., Retzel E., Sommerville C. (1994). Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabdopsis cDNA clones. Plant Physiol 106: 1241-1255.

Nicot N., Chiquet V., Gandon B., Amilhat L., Legeai F., Leroy P., Bernard M., Sourdille P. (2004). Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109: 800-805.

Nishiyama I. (1956). Basic number in the polyploidy of *Saccharum*. Journal of Heredity 47: 91-99.

Panaud O., Chen X.L., McCouch S.R. (1996). Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252: 597-607.

Panje R.R., Babu C.N. (1960). Studies in *Saccharum spontaneum*. Distribution and geographical association of the chromosome number. Cytologia 25: 152-172.

Petren K., Rosemary Grant B., Grant P.R. (1999). A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proc R Soc London B 266: 321-329.

Piperidis G, Christopher MJ, Carroll BJ, Berding N, D'Hont A. (2000). Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species *Erianthus arundinaceus*. Genome 43: 1033-1037.

Piperidis G., D'Hont A. (2001). Chromosome composition analysis of various *Saccharum* interespecific hybrids by genomic in situ hybridization (GISH). Proc. Int. Soc. Technol. 24: 565-566.

Powell W., Machray G.C., Provan J. (1996). Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1: 215–221.

Price S. (1957a). Cytological studies in *Saccharum* and allied genera. II. Geographic distribution and chromosome numbers in *S. robustum*. Cytologia 22: 40-52.

Price S. (1965a). Cytology of *Saccharum robustum* and related sympatric species and natural hybrids. U.S. Dep. Agric. Res. Serv., Tech. Bull. 133, 47p.

Price S. (19965b). Interespecific hybridisation in sugarcane breeding. Proc. Int. Soc. Sugarcane Technol. 12: 1021-1026.

Price S. (1968). Cytology of Chinese and North Indian sugarcanes. Econ Bot 22: 155-164.

Price S., Daniels J. (1968). Cytology of south Pacific sugarcane and related grasses: with special reference to Fiji. J Hered 59: 141-145.

Qu L., Hancock J.F. (2001). Detecting and mapping repulsion-phase linkage in polyploids with polysomic inheritance. Theor Appl Genet 103: 136-143.

Qureshi S.N., Saha S., Kantley R.V., Jenkins J.N. (2004). EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8: 112-123.

Raboin L-M., Oliveira K.M., Lecunff L., Telismart H., Roques D., Butterfield M., Hoarau J-Y., D'Hont A. (2006). Genetic mapping in the high polyploid sugarcane using a bi-parental progeny: identification of a gene controlling stalk colour and new rust resistance gene. Theor Appl Genet 112(7): 1382-1391.

Reffay N., Jackson P.A., Aitken K.S., Hoarau J-Y., D'Hont A., Besse P., McIntyre C.L. (2005). Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. Mol Breed 15: 367-381.

Rice W.R. (1989). Analysing tables of statistical tests. Evolution 43: 223-225.

Roach B.T. (1986). Evaluation and breeding use of sugarcane germplasm. Proc. Int. Soc. Sugarcane Technol. 19: 492-501.

Roach B.T. (1972). Nobilisation of sugarcane. Proc Int Soc Sugar Cane Tech 14: 206-216.

Roach B.T., Daniels J. (1987). A review of the origin and improvement of sugarcane. In: Copersucar Int Sugarcane Breed Wkshp, Copersucar, São Paulo, Brazil, pp 1-31.

Röder M.S., Plaschke J., König S.U., Börner A., Sorrells M.E., Tanksley S.D., Ganal M.W. (1995). Abundance, variability and chromosomal location of microsatellite in wheat. Mol Gen Genet 246: 327-333.

Rossi M., Araujo P.G., Paulet F., Garsmeur O., Dias V.M., Chen H., Van Sluys M-A., D'Hont A. (2003). Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genomics 269: 406-419.

Russel J.R., Fuller J.D., Young G., Thomas B., Taramino G., Macaulay M., Waugh R., Powell W. (1997). Discrimination between barley genotypes using microsatellite markers. Genome 40: 442-450.

Saghai Maroof M.A., Biyashev R.M., Yang G.P., Zhang Q., Allard R.W. (1994). Extraordinarily polymorphism microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91: 466-470.

Saha S., Karaca M., Jenkins J.N., Zipf A.E., Reddy U.K., Kantety R.V. (2003). Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130: 355-364.

Sasaki T., song J.Y., Kogaban Y., Matsui E., Fang F., Higo H., Nagasaki H., Hori M., Miya M., Muraya-makayano E., Takiguchi T., Takasuga A., Niki T., Ishimaru K., Ikeda H., Yamamoto Y., Mukai Y., Ohta I., Miyadera N., Havukkala I., Minobe Y. (1994). Toward cataloguing all rice genes – large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J. 6: 615-624.

Scott K.D., Eggler P., Seaton G., Rossetto M., Ablett E.M., Lee L.S., Henry R.J. (2000). Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723-726.

Secex (2005). www.desenvolvimento.gov.br/sitio/secex

Senior M.L., Chin E.C.L., Lee M., Smith J.S.C., Stuber C.W. (1996). Simple sequence repeat markers developed from maize sequences found in the Genbank database: map construction. Crop Sci 36: 1676-1683.

Sharapova N., McMullen M.D., Schultz L., Schroeder, S., Sanchez-Villeda H., Gardiner J., Bergstrom D., Pouching K., Melia-Hancock S., Musket T., Duru N., Polacco M., Edwards K., Ruff T., Register J.C., Brouwer C., Thompson R., Velasco, R., Chin E., Lee M., Woodman-Clikeman W., Long M.J., Liscum E., Cone K., Davis G., Coe E.H. (2002). Development and mapping o SSR markers for maize. Plant Mol Bio 48: 463-481.

Shen B., Carneiro N., Torres-Jerez I., Stevenson B., McCreery T., Helentjaris T., Baysdorfer C., Almira E., Ferl R.J., Habben J.E., Larkins B. (1994). Partial sequencing and mapping of clones from two maize cDNA libraries. Plant Mol Biol 26: 1085-1101.

Sills G., Bridges W., Al-Janabi, S., Sobral B. (1995). Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. Ex. Grassl). Molecular Breeding (1): 355-363.

Silvarolla M.B., Aguiar-Perecin M.L.R. (1994). Evaluation of chromosome number stability in two sugarcane varities. Rev Brasil Genet 17(2): 237-242.

Sobral B.W.S., Braga D.P.V., LaHood E.S., Keim P. (1994). Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. Theor Appl Genet 87: 843-853.

Soltis D.E., Soltis P.S. (1993). Molecular data and the dynamic nature of polyploidy. Critical reviews in Plant Science 12: 243.

Sreenivasan T.V. (1975). Cytogenetical studies in *Saccharum spontaneum* L. Proc. Indian Acad. Sci. 81: 131-144.

Sreenivasan T.V., Jagathesan D. (1975). Meiotic abnormalities *Saccharum spontaneum* L. Euphytica 24: 543-549.

Sreenivasan T.V., Ahloowalia B.S. and Heinz D.J. (1987). Cytogenetics. In Sugarcane Improvement Through Breeding. Edited by Heinz D.J. Amsterdam: Elsevier Press. pp211-253.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. The Plant Journal 3:739-744.

Stebbins G.L (1950). Variation and evolution in plants. Oxford University Press: Oxford.

Sterky F., Lundeberg J. (2000). Sequence analysis of genes and genomes. Journal of Biotechnology 76:1–31.

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

Szewc-McFadden A.K., Lamboy W.F., Hokanson S.C., McFerson J.R. (1996). Utilization of identified simple sequence repeats (SSRs) in *Malus* x *Domestica* (apple) for germplasm characterization. Hort Sci 31: 619.

Taylor C., Madsen K., Borg S., Moller M.G., Boelt B., Holm P.B. (2001). The development of sequence-tagged sites (STSs) in *Lolium perenne* L.: the application of primer sets derived from other genera. Theor Appl Genet 103: 648-658.

Teixeira L.H.M. (2006). Mapeamento funcional em cana-de-açúcar utilizando ESTs como marcadores moleculares. Campinas, SP, 109p. Dissertação (Mestrado) – Universidade Estadual de Campinas, 2006.

Temnykh S., DeClerk G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441-1452.

Thiel T., Michalek W., Varshney R.K., Graner A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106: 411-422.

Uhl C.H. (1992). Polyploidy, dysploidy, and chromosome pairing in *Echeveria* (Crassulaceae) and its hybrids. Am. J. Bot. 79: 556-566.

Van de Loo F.J., Turner S., Somerville C. (1995). Expressed sequence tags from developing castor seeds. Plant Physiol 108: 1141-1150.

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

Vigouroux Y., Jaqueth J.S., Matsuoka Y., Smith O.S., Beavis W.D., Smith J.S.C., Doebley J. (2002). Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19: 1251-1260.

Vincent J.B., Patterson A.D., Strong E., Petronis A., Kennedy J.L. (2000). The unstable trinucleotide repeat story of major psychosis. Am J Med Genet 97: 77-97.

Wang X., Bowen B. (1998). A progress report on corn genome projects at pioneer hi-bred. In: Plant and Animal genome VI Conference, San Diego, California.

Wu K.K., Burnquist W., Sorrells M.E., Tew T.L., Moore P.H., Tanksley S.D. (1992). The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl genet 83:294-300

Wu K.S., Tanksley S.D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225-235.

Yamamoto K., Sasaki T. (1997). Large-scale EST sequencing in rice. Plant Mol Biol 35: 135-144.

Yang G.P., Saghai Maroof M.A., Xu Q.F., Zhang C.G., Biyashev R.M. (1994). Comparative analysis of microsatellite polymorphism in landraces and cultivars of rice. Mol Gen Genet 245: 187-194.

Yu J.K., Dake T.M., Singh S., Benscher D., Li W., Gill B., Sorrells E. (2004). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47: 805-818.

9 Resultados Complementares

Análises citogenéticas

Análises citológicas dos parentais (SP80-180 e SP80-4966) e de 5F1 (28, 34, 44, 56, 65), escolhidos ao acaso, foram realizadas com o intuito de determinar seus números cromossômicos e assim entender o comportamento dos cromossomos destes híbridos, oriundos do cruzamento entre as duas variedades pré-comerciais em estudo. Imaginou-se que o número cromossômico entre os parentais pudesse ser bem diferente e que este fato resultaria muitas anomalias cromossômicas na progênie, sendo esta a provável razão pela qual estaria havendo dificuldades na ligação e ordenação dos marcadores no mapa genético.

Assim, dado o grande número de cromossomos esperado, tornou-se imprescindível o estabelecimento de um pré-tratamento que auxiliasse na obtenção de preparações, nas quais os cromossomos se apresentassem bem espalhados, com boa morfologia e em células intactas. As metodologias empregadas foram as seguintes.

a) Pré-tratamentos

Os pré-tratamentos sugeridos por Silvarolla e Perecin (1994) utilizam inibidores de fuso e de síntese protéica para a determinação do número de cromossomos somáticos em cana-de-açúcar, a partir do esmagamento das pontas das raízes germinadas. Com o objetivo de recuperar as células ainda em divisão mitótica, as pontas das raízes deveriam ser cortadas quando atingissem cerca de 1-3cms. Assim, após 24hs de crescimento radicular, já foi possível coletar as primeiras raízes para os pré-tratamentos.

O princípio de ação dos inibidores de fuso baseia-se na alteração da viscosidade do citoplasma que leva à destruição do mecanismo de formação do fuso, de modo a permitir que os cromossomos fiquem livres e passíveis de dispersão durante o processo de preparo da lâmina, além de conferir aos cromossomos um grau de condensação adequado à manutenção de sua morfologia. Neste estudo o inibidor 8-hidroxiquinolina (8HQ) foi empregado.

Alguns testes foram realizados apenas com o 8HQ, enquanto que outros foram feitos através de uma combinação do 8HQ com um inibidor de síntese protéica, denominado cicloheximida. No caso da cana-de-açúcar este é um tratamento particularmente útil, pois provoca contração dos cromossomos tanto em prófase quanto em metáfase, facilitando a contagem dos cromossomos em ambas fases.

Desta forma foram elaborados cinco pré-tratamentos: (1) 8HQ(0.2mM) isolado, à 28ºC, durante 5hs; (2) 8HQ(0.2mM) isolado, à 14ºC, durante 5hs; (3) 8HQ(0.2mM) combinado com a cicloheximida (0.009%), à 28ºC, durante 5hs; (4) 8HQ(0.2mM) combinado com a cicloheximida (0.009%), à 14ºC, durante 5hs; (5) 8HQ(0.2mM) combinado com a cicloheximida (0.009%), à 4ºC, durante 24hs. As pontas de raízes de cada indivíduo foram coletadas, pré-tratadas e fixadas com Carnoy (álcool-acético 3:1), por 24hs.

b) Preparo e análise das lâminas

As raízes fixadas foram hidrolisadas em HCl 1N, à 60º, durante 13 minutos. Em seguida, as raízes foram coradas com o reativo de Schiff, durante 2 horas e lavadas com água destilada por 30 minutos ou até retirar o excesso de corante. Após a lavagem do corante, as raízes foram submetidas à hidrólise enzimática (Pectinase 2.5% + celulase 1%),

à 35ºC, por 5 minutos, visando torná-las mais propícias ao esmagamento no momento do preparo das lâminas, produzindo mais células intactas. Com o auxílio de uma lupa e materiais de corte de material, a coifa foi retirada da ponta da raiz e o meristema foi coletado. O material foi esmagado e as lâminas montadas. As lâminas satisfatórias tiveram as lamínulas removidas em nitrogênio líquido, foram secas ao ar livre e montadas em Entellan.

c) Contagem dos cromossomos

As lâminas foram analisadas com o auxílio de um microscópio ótico Olympkus BX51. Os pré-tratamentos foram avaliados segundo o número de metáfases, a integridade das células e o grau de condensação e morfologia dos cromossomos. Na metáfase os cromossomos apresentam alta condensação, clara morfologia e maior facilidade de contagem, sendo esta a fase de maior vantagem para tal tipo de estudo. A integridade das células foi checada com o contraste de fase do microscópio. Foram descartadas as lâminas que apresentaram poucas ou nenhuma célula em divisão, bem como aquelas cujas células apresentavam cromossomos muito condensados ou com pouco espalhamento, o que dificultava a contagem dos mesmos. As fotomicrografias das células selecionadas foram obtidas com o auxílio de uma câmera digital acoplada ao microscópio Olympkus BX51. Os cromossomos foram contados diretamente das fotos impressas.

d) Resultados e discussão

Os tratamentos combinados de 8 hidroxiquinolina (8HQ) (0.2mM) com cicloheximida (0.009%), em todas as temperaturas testadas, foram mais eficientes que os tratamentos apenas com o inibidor de fuso 8HQ. Os tratamentos isolados praticamente não apresentaram células em divisão, desta forma, não puderam ser empregados na contagem dos cromossomos. Dos tratamentos combinados, a mistura do 8HQ com cicloheximida, a 14°C, durante 5hs resultou em preparações com o maior número de metáfases disponíveis para a contagem cromossômica. Para os parentais e 2 F1s (indivíduos 28 e 44) da população de mapeamento, este pré-tratamento possibilitou a contagem dos cromossomos, enquanto que os 3 F1s restantes (indivíduos 34, 56, 65) foram descartados, pois nenhum dos tratamentos testados resultou em células em divisões e passíveis de contagem.

Silvarolla e Perecin (1994) também observaram que os tratamentos combinados resultaram em melhores preparações. No entanto, para a variedade NA56-79, o pré-tratamento selecionado combinou 8HQ (0.025%) com cicloheximida (0.009%), a 28°C, durante 5 horas, enquanto que para a variedade Co419 a combinação de α-bromonaftaleno (2%), cicloheximida (0.007%) e DMSO, durante 3 horas, mostrou-se mais eficaz.

O principal desafio da investigação citológica em cana-de-açúcar, aliado a grande quantidade de cromossomos do gênero *Saccharum*, é a necessidade do ajuste dos tratamentos para cada variedade. No nosso estudo, os cromossomos encontraram-se ainda pouco condensados e alguns com aspecto difuso, características observadas principalmente devido à ação da 8HQ. . Poderíamos alterar a concentração deste inibidor para melhorar o aspecto das células, porém, para nosso objetivo no presente estudo as preparações se mostraram adequadas.

O esmagamento do material na lâmina foi uma etapa crucial no preparo das lâminas. Uma vez que o número de cromossomos esperados para estas variedades modernas de cana-de-açúcar varia de 100 a 130, as células devem ser bem espalhadas na lâmina, para evitar sobreposição dos cromossomos e para não acarretar erros no processo de contagem. O tratamento das pontas de raízes com as enzimas pectinase e celulase auxiliou neste procedimento de esmagamento, tornando o material mais propício ao esmagamento.

Para cada variedade, as células em metáfase foram selecionadas, seguindo todos os critérios citados. Em seguida, efetuou-se a contagem, sendo estabelecido como número cromossômico o número mais freqüente observado nas células de cada indivíduo. Porém, números precisos não puderam ser estipulados para cada uma das variedades, uma vez que as células não ficaram completamente separadas a ponto de não deixar margem de dúvida. No parental feminino (SP80-180) cerca de 105-106 cromossomos foram identificados na maioria das células contadas. No parental masculino (SP80-4966) o número de cromossomos mais freqüente foi de 107, havendo uma variação de 1-2 cromossomos para mais ou para menos. No indivíduo 28 da progênie F1, cerca de 105-107 cromossomos foram contados, enquanto que no indivíduo 44, o número variou de 102 a 104 (Figura 1). As variações entre as células, para cada indivíduo, foram pequenas e podem ser consideradas como erros de interpretação, uma vez que alguns ajustes no pré-tratamento deveriam ter sido feitos para obtenção de preparações com células mais nítidas e sem sobreposição. Apesar das sobreposições de cromossomos observadas nas células, conseguimos fazer a contagem dos mesmos. Apesar de não podermos estabelecer os números cromossômicos exatos de cada indivíduo, pudemos observar que apresentam números muito próximos entre si (2n = cerca de 106).

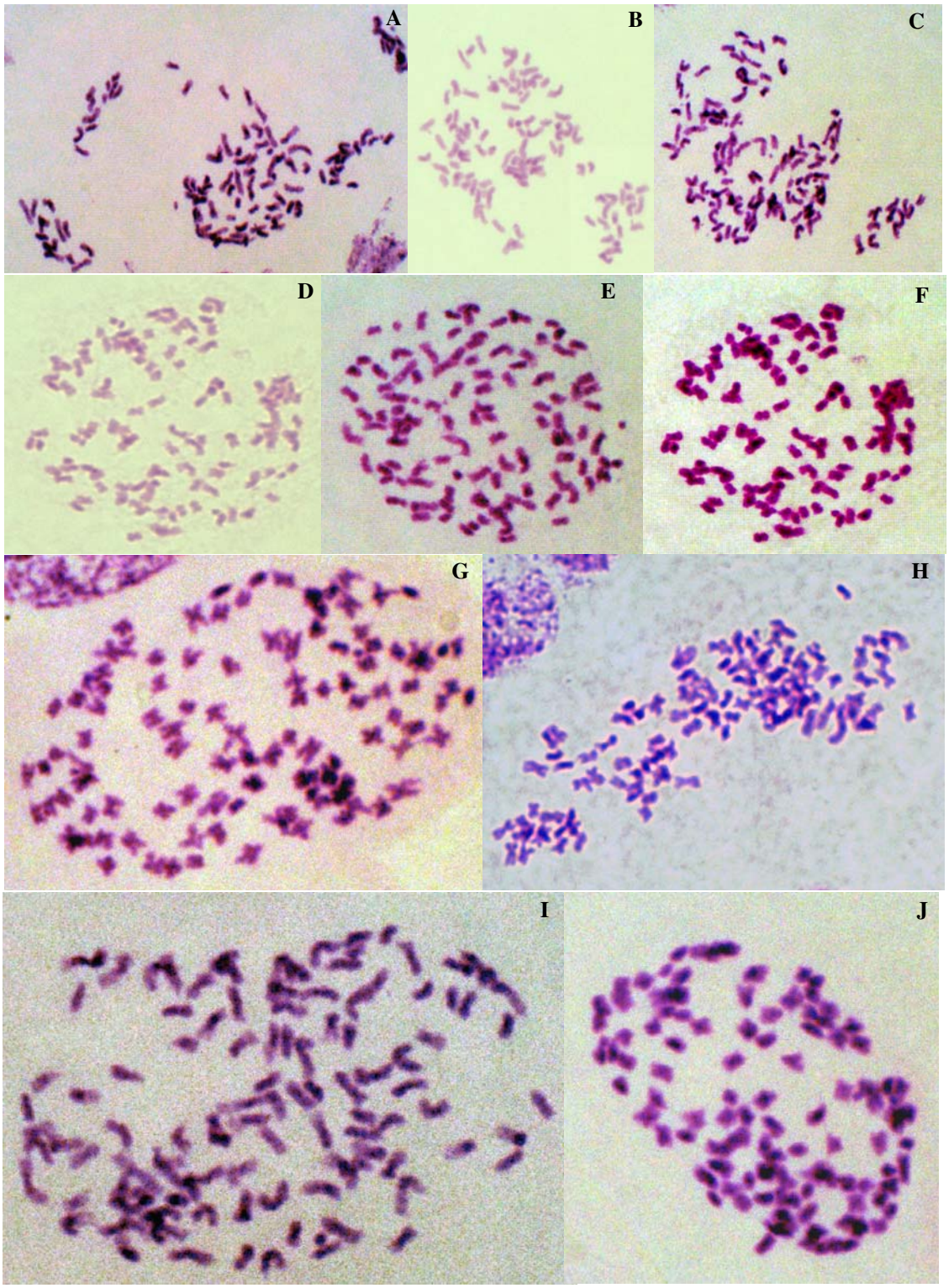


Figura 1 Metáfase mitótica em ponta de raiz do parental feminino (SP80-180 – A, B, C), do parental masculino (SP80-4966 – D, E, F), dos indivíduos 28 (G, H) e 44 (J, L) da progênie F1. A contagem cromossômica para cada célula foi aproximadamente de: (A, B, C) 2n=105-106; (D, E, F) 2n=106-108; (G, H) 2n=105-107; (I, J) 2n=104**.**

10 Anexos

10.1 Anexo I

Artigo: "Genetic mapping in sugarcane, a high polyploidy, using bi-parental progeny: identification of a gene controlling stalk colour and a new resistance gene"

L.M. Raboin, K.M. Oliveira, L. Lecunff, H. Telismart, D. Roques, M. Butterfield, J-Y Hoarau, A. D'Hont

Publicado na revista

Theoretical and Applied Genetics (112: 1382-1391)

Theor Appl Genet (2006) 112: 1382–1391
DOI 10.1007/s00122-006-0240-3

ORIGINAL PAPER

L. -M. Raboin · K. M. Oliveira · L. Lecunff
H. Telismart · D. Roques · M. Butterfield
J. -Y. Hoarau · A. D'Hont

Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene

Received: 11 March 2005 / Accepted: 13 February 2006 / Published online: 22 March 2006

Abstract Modern sugarcane cultivars (*Saccharum* spp) are highly polyploïd and aneuploid interspecific hybrids ($2n = 100–130$). Two genetic maps were constructed using a population of 198 progeny from a cross between R570, a modern cultivar, and MQ76-53, an old Australian clone derived from a cross between Trojan (a modern cultivar) and SES528 (a wild *Saccharum spontaneum* clone). A total of 1,666 polymorphic markers were produced using 37 AFLP primer combinations, 46 SSRs and 9 RFLP probes. Linkage analysis led to the construction of 86 cosegregation groups for R570 and 105 cosegregation groups for MQ76-53 encompassing 424 and 536 single dose markers, respectively. The cumulative length of the R570 map was 3,144 cM, while that of the MQ76-53 map was 4,329 cM. Here, we integrated mapping information obtained on R570 in this study with that derived from a previous map based on a selfed R570 population. Two new genes controlling Mendelian traits were localized on the MQ76-53 map: a gene controlling the red stalk colour was linked at 6.5 cM to an AFLP marker and a new brown rust resistance gene was linked at 23 cM to an AFLP marker. Besides another previously identified brown rust resistance gene (*Bru*1), these two genes are the only other major genes to be identified in sugarcane so far.

Keywords Sugarcane · Genetic mapping · AFLP · Rust resistance gene · Colour gene · Polyploid

Communicated by S. J. Knapp

L. -M. Raboin (✉) · H. Telismart · D. Roques
CIRAD (Centre de coopération internationale en recherche agronomique pour le développement) - U-MR PVBMT,
Station de Ligne Paradis 7 chemin de l'IRAT,
97410, Saint Pierre Réunion, France
E-mail: raboin@cirad.fr
Tel.: +33-262-499223
Fax: +33-262-499293

K. M. Oliveira
Universidade Estadual de Campinas,
13083–970, Campinas, SP, Brazil

L. Lecunff · A. D'Hont
CIRAD, UMR 1096, TA 40/03, Avenue Agropolis,
34 398, Montpellier cedex 5, France

M. Butterfield
SASRI (South African Sugarcane Research Institute),
Private bag X02, Mount Edgecombe, 4300, South Africa

J. -Y. Hoarau
CERF (Centre d'essai, de recherche et de formation),
BP 315 97494, Ste-Clotilde Réunion, France

Introduction

Modern sugarcane cultivars have one of the most complex genomes among important crops (Grivet and Arruda 2001). They are polyploid and aneuploid hybrid derivatives from two highly polyploid species, i.e. the domesticated sugar-producing species *Saccharum officinarum* ($x = 10$, $2n = 8x = 80$) and the wild species *Saccharum spontaneum* ($x = 8$, $2n = 40–128$). The first interspecific hybrids produced in the early twentieth century were backcrossed with *S. officinarum*. In both interspecific F1 and BC1 crosses, *S. officinarum* transmitted its somatic chromosome number ($2n$) to the progeny (Bhat and Gill 1985). Modern cultivars thus have chromosome numbers in the 100–130 range, around 15 to 25% of which are derived from *S. spontaneum* (D'Hont 1993). Despite this complicated picture, the meiosis of modern sugarcane cultivars is fairly regular, mainly involving bivalent pairing (Price 1963; Burner and Legendre 1993, 1994), and the chromosome assortment is the result of a combination of preferential and random pairing (Jannoo et al. 2004).

Genetic maps have been produced, based on single dose markers (Wu et al. (1992), for the two ancestral species *S. spontaneum* (Al Janabi et al. 1993; Da Silva et al. 1993, 1995; Ming et al. 1998, 2000b) and *S. officinarum* (Mudge et al. 1996; Guimaraes et al. 1997;

Ming et al. 1998, 2000b). Genetic maps have also been constructed for two modern sugarcane cultivars, i.e. Q165 (Aitken et al. 2005) and R570. For the latter cultivar, a first RFLP map was constructed using a selfed population derived from R570 (Grivet et al. 1996). AFLPs (Hoarau et al. 2001), SSRs and resistance gene analogs (Rossi et al. 2003) were later mapped using a larger number of selfed R570 progeny. This AFLP-based map encompassed more than 1,100 markers and reached a total length of 7,800 cM, which represents a coverage of around 46% of the anticipated genome size (17,000 cM). The species origins of the markers and the results of molecular cytogenetic studies (D'Hont et al. 1996) revealed that about 10% of the R570 chromosomes were inherited from *S. spontaneum* and that another 10% were recombinants between *S. spontaneum* and *S. officinarum*. The marker coverage on the map of the interspecific cv. R570 is uneven, with *S. spontaneum* chromosomes being covered more densely than those of *S. officinarum*. This discrepancy results from the lower polymorphism rate in the highly polyploid *S. officinarum*—the main component of the modern sugarcane cultivar genome—as compared to *S. spontaneum* (D'Hont et al. 1994; Grivet et al. 1996).

A saturated map must be developed to be able to efficiently localize major genes or Mendelian factors involved in quantitative trait loci (QTL). Moreover, QTL mapping in sugarcane is a real challenge, since many alleles coexist at each locus due to the high polyploidy. At a particular locus, the effect of an allele should be perceptible only if it exceeds the average effect of all other segregating alleles in the background but not, as in diploids, if its effect simply exceeds that of a single alternative allele (D'Hont and Glaszmann 2001; Hoarau et al. 2002).

Ming et al. (2001, 2000a) investigated the genetic basis of traits related to sugar content, plant height and flowering in interspecific crosses between *S. officinarum* and *S. spontaneum*. Numerous QTLs were detected that could be localized in a few genomic regions, suggesting that substantially fewer genes may actually be involved in the genetic control of these traits. Within an interspecific cross, wide phenotype segregation can provide a favourable setting for QTL detection. By comparison, a study of yield components (plant height, stalk diameter, stalk number and Brix) in the selfed progeny of the modern cultivar R570 revealed numerous QTLs with smaller individual effects (Hoarau et al. 2002). Similarly, Jordan et al. (2004) detected numerous small QTLs for stalk number in sugarcane and found sorghum QTLs for tillering in syntenic positions. The only major gene that has been localized so far in the sugarcane genome is a rust resistance gene (Daugrois et al. 1996). This gene (called *Bru*1 for brown rust) is currently the focus of a map-based cloning project (D'Hont et al. 2001; Asnaghi et al. 2004).

In this paper, we describe the identification and mapping of two other major genes, including a new potential rust resistance gene and a gene controlling stalk colour. These genes were identified in the sugarcane clone MQ76-53 through an extensive genetic mapping study of a bi-parental R570 × MQ76-53 cross. We discuss the constraints and advantages of working in a bi-parental cross in comparison with a selfed progeny population.

Materials and methods

Plant and DNA material

The mapping population consisted of 198 progeny derived from a R570 × MQ76-53 cross. R570 is a modern cultivar that was developed at CERF [Centre d'Essai de Recherche et de Formation, Réunion] and is derived from a cross between two modern cultivars H32-8560 and R445. MQ76-53 is an old Australian sugarcane clone that comes from a cross between the old cultivar Trojan (Co 270 × *S. officinarum*) and the *S. spontaneum* clone SES528. Its genetic structure should therefore be close to that of an interspecific F1 hybrid. R570 is rust resistant and has green stalks. MQ76-53 is also rust resistant but has red stalks. Total genomic DNA of mapping progenies was extracted from fresh leaves according to the method described by Hoisington (1992).

In addition, 133 progeny from a cross between the cultivars B63-758 (rust susceptible) and MQ76–53 (rust resistant) were evaluated for rust resistance.

Field evaluation and statistical analysis of rust resistance

A subset of 166 out of the 198 mapping progenies was evaluated for rust resistance, under natural infestation conditions, in a randomized complete block design with three replicates at the Ligne Paradis CIRAD research station (Saint Pierre, Reunion). Each plot included four distinct stools. The trial was planted in mid-November 2002 and evaluated for rust resistance in early September 2003, i.e. at the end of the winter season which is the most favourable period for rust development in Reunion (cool humid conditions). Rust resistance was scored on each plot on a 1 (the most resistant:) to 9 (the most susceptible) scale according to Tai et al. (1981). A score of 1 indicated the absence of sporulating pustules (uredospores). Susceptible plants could have scores ranging from two (a few sporulating pustules) to nine (many pustules even on younger leaves and necrosis of older ones). Natural rust infestation during the experiment was heavy. Indeed, B34-104, a moderately susceptible clone that is usually rated between 3 and 4, planted as control, was rated 6. The two other controls, R570 and MQ76-53, were rated 1. Analysis of variance of the rust score variable was performed using the SAS mixed procedure (SAS version 6.12, SAS Institute, Inc., NC, USA). The genotype (clone) factor was considered as random and replication as fixed. Broad-sense heritabilities were calculated at the experimental design

level from the ratio between genetic variance (σ_g^2) and phenotypic variance (σ_p^2), with $\sigma_p^2 = \sigma_g^2 + \sigma_e^2/j$, where σ_e^2 is the error variance and j the number of replications. Segregation analysis was performed on mean scores computed for all replications for each progeny. One way ANOVA was performed to calculate the proportion of phenotypic variance (R^2) explained by markers associated with rust resistance. These analyses were performed using all available markers and mean rust scores for each genotype (clone). The proportion of total phenotypic variance explained by a marker (R^2) was calculated using the sums of squares obtained by ANOVA.

Progeny from the B63-758 × MQ76-53 cross were studied in the field, with seedlings planted randomly 50 cm apart. In this cross, rust resistance was scored for each individual seedling on the basis of the presence/absence of sporulations. Seedlings bearing sporulating pustules were classified as susceptible, and otherwise they were classified as resistant.

Field evaluation and statistical analysis of stalk colour

Stalk colour of the 198 mapping progenies was scored in three different environments on the basis of the discrete classification red versus non-red colour (mostly green or yellow stalks);(1) in the collection used to conserve the population colour was scored in August 2002, a few weeks after leaves had been stripped from the first stool in each conservation plot; (2) in a duplicate of this collection, colour was scored in June 2003, without previous stripping of the stalks (3) finally, colour was scored in a separate trial where two cuttings of each of the 198 clones were planted in 10 L pots that were well separated from each other in order to favour sunlight exposure. Stalks were regularly stripped before colour evaluation in July 2004. The consistency of the three different stalk colour evaluations was checked using Fisher's exact test (Mehta and Patel 1983; SAS Institute 1990, FREQ procedure) between scoring of environments (1) and (2), environments (2) and (3), and environments (1) and (3).

AFLP markers

AFLP analysis (Vos et al. 1995) was performed using the Gibco BRL genome I kit according to the manufacturer's instructions, except for slight modifications as described in Hoarau et al. (2001). Each AFLP marker was identified by the primer combination consisting of six letters plus a band number indicated as a suffix. The first three letters represent *Eco*RI selective nucleotides and the last three *Mse*I selective nucleotides.We used the same 37 combinations as those used to build the R570 map of Hoarau et al. (2001) and Rossi et al. (2003), except for the three *Eco*RI/*Mse*I primer combinations aag/caa, aag/cat and aag/cta, which were replaced by the combinations aag/ctg, aac/ctg and acc/cta. Most of the bands inherited from R570 could thus be labelled as in the previously published maps.

Microsatellite markers

Thirty-three SSRs developed at CIRAD in collaboration with Génoscope (Evry, France), in addition to 13 SSRs mined from the Brazilian sugarcane EST database (Pinto et al. 2004) were analysed using the protocol described in Rossi et al. (2003). CIRAD SSR bands inherited from R570 were coded according to this previously published map using the following nomenclature: m (microsatellite) followed by the number of the SSR, and then the letter m (as for marker) followed by a number identifying the band. CIRAD SSRs (complete nomenclature = mSSCIRxx) are described at http://www.tropgenedb.cirad.fr/en/sugarcane.html. The nomenclature of EST SSR was m (microsatellite) followed by a letter and a two digit code number and then the letter m (as for marker) followed by a number identifying the band. SSR bands were scored as dominant markers (presence vs absence).

RFLP markers

Nine candidate genes differentially expressed in response to challenge by smut (Heinze et al. 2001) were kindly provided by the South African Sugarcane Research Institute (SASRI) and used as RFLP probes (Table 1). The nomenclature used for these RFLP markers was 'ADS' followed by a figure between 1 and 9, and then a code for the restriction enzyme used (Eco for EcoRV and Hin for HindIII) followed by a number identifying the band. DNA preparation, Southern blotting and hybridization were performed as described previously by Grivet et al. (1996). RFLP bands were scored as dominant markers (presence vs. absence).

Map construction

The genetic maps were constructed using linkages in coupling phase between single dose markers only

Table 1 Description of candidate genes used as RFLP probes provided by SASRI

Probe name	Putative gene function
ADS1	Thaumatin
ADS2	Flavonoid pathway transcription factor
ADS3	Pto ser/thr protein kinase
ADS4	Pathogen induced nucleotide binding site
ADS5	Phosphoprotein phosphatase
ADS6	Isoflavone reductase
ADS7	G protein receptor
ADS8	Cellulose synthase
ADS9	Cell wall associated kinase

Table 2 Synthesis of mapping information on R570: alignment of the present map with maps of Hoarau et al. (2001) and Rossi et al. (2003). Locus-specific markers (SSRs and RFLPs) highlighted in bold characters represent newly localised markers. CGs preferentially pairing at meiosis are associated within boxes (bold boxes indicate newly detected repulsion linkages) When a CG is involved in repulsion linkage with two CGs a vertical line is drawn between each pair of CGs in repulsion). When part of a former CG was presumably wrongly linked in coupling phase to its remaining part, it is shown as crossed out in bold characters. When two former CGs could be linked together to form a new CG, they are shown in bold characters with a + sign. Detailed information on the different genetic maps of R570 is available at http://www.tropgenedb.cirad.fr/en/sugarcane.html

HG[4]	Cons[3]	Rossi[1]	JYH[2]	LMR[3]	SSR loci	RFLP loci
I					m2, m13, m14, m19, m27 m42*, m52, m53	R024Sst[5], R057hin, R083Sst, R125Sst, R183Sst, R313eco, R371Dra
I	I a	I a	51	54	m52m8	R057Hinc, R024Sstf
I	I b	I b+U 62	52+38	30	m14m3, 27ml	R057Hinf
I	I c	I c	105	29	m27m2	R083Sstc
I	I 4	I 4	1	43	m19m1, m42m7, m52m3, m53m2	R083Sste, R125Ssta
I	I 5	I 5	3	154	m2m2, ml4m6	R024Sstc, R024Sstg,R183Sstg
I	I 6	I 6+U 3	37+72	111	m13m2, m27m3, m52m6	R024Ssta, R183Ssta, R313Ecob, R313Ecoe,R371Draa
I	I 7	I 7	74	40	m19m5, m53m9	R024Sste, R083Sstd
I	I 8	I 8	68	-	m52m1	R125Sstc
I	I 9	I 9	2	108	m13m1, m14m1	R057Hine, R057Hind, R125Sstb, R313Ecoa, R313Ecod
I	I 10	I 10	10	42	m19m2, m53m1	
I	I 17	U 8	82	114+143	m52m2	
II					mA03, mA16*,mB01, m25, m34, m35, m39, m41, m48, m50, m56, m69*	R012Eco, R145Eco, R169Hin, R258Dra, R275Eco, R366Hin, R386Dra
II	II a	II a	90	128		R012Ecof, R145Ecoe, R275Ecoe
II	II b	II b	71	166		R012Ecoc, R275Ecoc
II	II 4	U 35	84	-	m39m1	
II	II u21	U 21	85	107		
II	II 6	II 6	92+5	21	mA03m5, m25m2, m34m1	R012Ecoa, R386Drac
II	II u29	U 29	88	-	m25m3	
II	II 7	II 7	62	-	m39m2	R275Ecof
II	II 8	II 8	59	23	m25ml, m34m3, m48m2, m50ml ,m69m2	R169Hina, R386Draa
II	II 9	II 9	4	69	mA03m6, mB01m1, m34m2, m35m1, m41m1	R169Hinc, R258Draa, R275Ecod, R366Hina
II	II 10	II 10	7	148	mA16m1, m35m2, m41m2	R169Hinf, R258Drab, R366Hinc
II	II 11	II 11	-	-	m48m3, m56m4	
II	II 12	II 12	-	-	m35m3	R366Hind
II	II 14	U 61	6	115		
III					mC03, m33, m40, m42*, m55, m64, m70	R149Eco, R523Eco
III	III a	III a	-	189	m40m6, m64m2	R523Ecoc
III	III b	III b	-	-	m64m6, m70m4	R523Ecoa
III	III 2a	U 57	~~24~~	-		
III	III 3	III 3	23	53	m42m4, m55m2, m70m2	
III	III 2	III 2	24	126	m55m8, m70m6	
III	III 4	III 4	22	68+191	m40m1	R149Ecoa
III	III 7	III 7	83	82	m40m4, m65m9, m64m8, m70m7	
III	III 8	III 8	107	27	m33m5, m55m5, m64m5, m70m3	
III	III u	~~VI 3~~	~~29~~	59	mC03m3, m40m2	R149Ecoe
III	III u54	U54	34	-	m33m2, m42m1	
III	III 10	III 10	114	57	m64m1, m70m1	
IV					mA06, m74	R142Eco*, R184Eco, R372Eco, R441Sst, R526Eco*, R533Hin
IV	IV 1	IV 1	8	81+109	mA06m4	R142Ecod, R184Ecog, R184Ecof, R372Ecoe, R372Ecoa, R441Sstd, R441Ssta, R526Ecob, R533Hinb
IV	IV 2	IV 2	9	106	m74m2, mA06m2	R441Sstf, R441Sste, R533Hina
VI					m7, m9, m10, m23 , m31, m37, m47, m54, m57, m58, m60, m68, m73*	ADS1Eco, ADS3Hin, ADS4Eco, ADS6Hin, ADS8Hin, R087Hin, R152Hin, R162Eco, R196Eco, R272Eco, R367Dra, R396Hin, R482Eco
VI	VI 1a	VI 1a	26	-	m10m1	R196Ecoc, R196Ecod
VI	VI 1b	VI 1b	33	-	m47m4, m60m4	
VI	VI 2	VI 2	28	8+173	m10m7, m54m7	ADS3Hin10, ADS6Hin4 , ADS8Hin3b, R272Ecoa, R482Ecod
VI	VI 3	VI 3+VI 8	29+70	7	m10m3, m31m6, m37m1, m54m6, m60m3, m73m3	ADS6Hin8, ADS8Hin3h, R272Ecod
VI	VI 4	VI 4	31	125	m9m2, m68m1	
VI	VI 9	VI 9	32	32	m37m5,m47m5, m57m3	
VI	VI 5	VI 5	94	-	m10m2, m58m2	
VI	VI 10	VI 10	25	66	m37m4, m57m1, m60m2	
VI	VI 11	VI 11	27	-	m60ml, m73m1	R367Draa
VI	VI 12	VI 12	30	113	m37m3, m57m2	
VI	VI 15	VI 15	109	169	m10m5	

Table 2 (Contd.)

HG[4]	Con[5]	Rossi[1]	JYP[2]	LMR[3]	SSR	RFLP
VI	VI 20	U 11	77+104	164+178	m54m8	R087Hinb, R152 Hinb, R162Ecoa, R162Ecod, R162Ecoc
VI	VI a	VI a	-	-	m10m6	R196Ecof, R272Ecob
VI	VI b	VI b	-	-	m58m1	R196Ecoa, R272Ecog
VI	VI c	VI c	57	11		ADS4Eco6, R396Hinb
VI	VI u52	U 52	56	12	m23m2, m54m2, m7m1	ADS4Eco5
VI	VI d	VI d	61	65+110		R367Drac, R396Hinc
VI	VI u19	U 19	73	19		ADS3Hin4, ADS8hin7
VI	VI u36	U 36	95	16		ADS1Eco4, R272Ecoc
VI	VI u60	U 60	75	72	m23m1, m54m1	
VI	VI u51	U 51	54	-		R152Hinc, R396Hina
VI	VI u45	U 45	117	-	m3l m1	
VII					mB41, m4, m8, m21, m29, m36, m43, m67, m69*	R019Sst, R137St
VII	VII a	VII a	80		m 4m4	R019Sstc, R137Ssta, R 137Sstb, R137Sstf
VII	VII 1a	VII 1a	60	146	m4m5, m43m1	R137Sstd, (Rust resistance gene)
VII	VII 1b	VII 1b	13	48	m8m3, m67 m1, m69m1	
VII	VII 2	U 12	53	-		
VII	VII 3	U 14	11	-		
VII	VII 4	U 50	67	49	m8m4	
VII	VII 5	U 13	12	-		
VII	VII 8	VII 8	63	-	m21 m1, m36m2, m67m8	
VII	VII 9	VII 9	-	-	m4m8, m43m3	
VII	VII 14	VII 14	-	-	m4m3	R019Sste
VII	VII 18	U 2	-	-	m21m3, m29m3	
VII	VII 20	U 40	39	129	mB41m2, m29m1	
VIII					mA16*, m1, m12, m17, m18, m26, m28, m38, m44, m46, m61, m65, m73*	ADS7Hin, ADS9Eco, R039Sst, R082Dra, R118Eco, R131Eco, R142Eco*, R 173Sst, R185Sst, R267Sst, R281Hin, R326Sst, R335Eco, R342Sst,
VIII						R406Sst, R488Hin, R526Eco*, R542Hin
VIII	VIII a	VIII a	108	184	mA16m3	R118Ecoa, R185Sstb, R281Hind, R281Hing, R326Ssta, R326Sstf
VIII	VIII b	VIII b	65	55		R185Sstb, R267Sste, R326Sstb
VIII	VIII d	VIII d	55	156	m61m2	R542Hinb
VIII	VIIIu16	U 16	58	-		
VIII	VIII e	VIII e	78	71		R542Hina
VIII	VIII f	VIII f	45	-	m17m4, m46m1	
VIII	VIII g	VIII g	102	-		R039Sstf, R488Hinc
VIII	VIIIu58	U 58	101	-		R488Hind
VIII	VIII 1	VIII 1	14	67	m1m2, m12m4, m18m1, m38m1	R082Draa, R082Drab, R185Ssti, R281Hinb, R326Ssti
VIII	VIII 2	VIII 2	15	5	m1m1, m12m5, m17m3, m28m6, m44m1, m46m2, m65m3	ADS7Hin2, R039Sstd, R118Ecoc, R131Ecoe, R173Sstc, R185Sstg, R281Hine, R326Sste, R335Ecog, R342Sstb, R406Ssta
VIII	VIII 11	VIII 11	16	56	m17m2, m26m3, m65m2	R039Ssta, R335Ecoa, R335Ecob, R342Sstc, R342Sstd, R488Hinb
VIII	VIIIu49	U 49	18	-		
VIII	VIII 3	VIII 3	-	-	m12m1	R142Ecoe, R526Ecoc
VIII	VIII 4	VIII 4	43	155	m12m8, m28m2	R142Ecoc, R526Ecoa
VIII	VIII 5	VIII 5	20	1	m38m2, m73m2	ADS9Eco2, R118Ecob, R 173Sstf, R173Sstd, R173Ssta, R185Sstf, R267Sstc, R281Hina, R326Sstd
VIII	VIII 16	VIII 16	21	-	m61m3	R173Sste
VIII	VIII 8	U 53	17	-		
VIII	VIII 10	VIII 10	19	-	m18m3	
VIII	VIII 15	VIII 15	106	-	m38m4	R118Ecoe, R118Ecod, R173Sstb, R173Sstg, R281Hinf
-	U 1	U 1	99+110	-	m49m6	
-	U 4	U 4	69	172	m23m3	R016Drae
-	U 5	U 5	-	-		R482Ecob
-	U 15	U 15	-	-		R149Ecod
-	U 27	U 27	81	176		R012Ecoe
-	U 42	U 42	120	181	m4m2	
-	U 43	U 43	-	-		R031Ecoa
-	U 46	U 46	-	-		R129Ecod, R371Drab
-	U 47	U 47	100	112		R088Ecoa
-	U 55	U 55	118	-	R129Ecoc	
-	-	-	-	175	m54m4	
-	-	-	-	186	m54m5	
-	-	-	-	46	m16m1	
-	-	-	-	58	m4m6	

[1] Rossi *et al.* (2003) ; [2] Hoarau *et al.* (2001), [3] This study, [4] proposed homology groups, [5] Consensus map integrating information from [1,2,3]
* indicates probe/enzyme combinations or SSR primers producing alleles (bands) attributed to CGs belonging to different HGs
$ Resistance gene analogs mapped by Rossi et al. (2003)

(Wu et al. 1992). In order to distinguish single dose markers (1:1) from bi-parental single dose markers (3:1), we used a segregation ratio threshold of 1.73:1 since this ratio gives equal χ^2 values for both 1:1 and 3:1 hypotheses (Mather 1957). Two-point analyses between single dose markers were performed at a LOD score threshold of five and a recombination fraction threshold of 0.35 using Mapmaker 3.0 (Lander et al. 1987). Markers within each cosegregation group (CG) were then ordered by multipoint analysis using the Haldane mapping function. CGs were pooled in the same homology groups (HG) when (1) they had at least two RFLP probes or SSR in common, or when (2) they were linked in repulsion. Moreover, some R570 CGs could be assigned to HG on the basis of anchor markers (AFLP markers or SSR markers) they had in common with CGs from a previous R570 RFLP-based map (Grivet et al. 1996 and unpublished results). Chromosome pairing behaviour was investigated using the procedure described in Hoarau et al. (2001).

Results

Parental linkage maps

A total of 1,666 polymorphic markers were produced in the progeny using 37 AFLP primer pair combinations, 46 SSRs and 9 RFLP probes. Among these markers, 1,057 were single dose markers and used to build the map, with 584 (55%) specific to MQ76-53 and 473 (45%) specific to R570. Linkage analysis of all of these single dose markers resulted in 191 CGs encompassing a total of 960 markers, while 97 single dose markers remained unlinked: 424 R570-specific markers formed 86 CGs and 536 MQ76-53-specific markers formed 105 CGs. CG lengths ranged from 1.6 to 179.4 cM in MQ76-53 and from 1.1 to 158.3 cM in R570. The cumulative length of the MQ76-53 map was 4,329 cM. The cumulative length of the R570 map was 3,144 cM. The detailed maps are available at http://www.tropgenedb.cirad.fr/en/sugarcane.html (Ruiz et al. 2004). Only 17 out of the 105 MQ76-53 CGs could be assigned to the homology group framework defined for the R570 map on the basis of at least two common RFLP probes or SSR loci. Out of the 86 CGs on the R570 map, 60 CGs (70%) could be assigned to an HG. No preferential pairing between MQ76-53 CGs was detected at LOD = 3. In contrast, preferential pairing was observed between 18 pairs of CGs in R570, 14 of which had been previously detected by Hoarau et al. (2001).

Alignment of different R570 AFLP-based maps

The present R570 map was aligned with the former AFLP-based map of the same cultivar (Hoarau et al. 2001, Rossi et al. 2003). This was possible because most of the single dose AFLP markers specific to R570 in the present map (based on a bi-parental progeny derived from the R570 × MQ76-53 cross) were common to the former map (based on a selfed R570 progeny). Localisation of 32 new locus-specific markers (19 SSRs and 13 RFLPs) and detection of four new pairs of CGs in repulsion improved the assignment of the CGs of the previously published map to the homology group (HG) framework. With these new data, 72% of the CG could be assigned to a HG as compared to 52% in the R570 map of Rossi et al. (2003). When we integrated the information of all three studies, we obtained a tentative consensus map organized in seven HGs encompassing a total of 90 assigned CGs. The number of CGs per HG ranged from 2 (HG IV) to 22 (HG VI). The present mapping data also allowed reorganizing several previous published CGs (Table 2). In particular, the basal part of CG VI3 in Rossi et al. (2003) may have been erroneously attached to the upper part of this CG, assigned to HG VI. This basal part, which corresponds to CG 59 in this study, carries two locus-specific markers (m40 and R149Eco), thus suggesting that it belongs to HG III. In three cases, based on the new data, two formerly separate CGs could be pooled into a single CG, i.e. CG30 = Ib + U62 (Rossi et al. 2003); CG7 = VI3 + VI8 (Rossi et al. 2003); and CG111 = I6 + U3 (Rossi et al. 2003). In addition to the CGs presented in Table 2, the R570 genetic map encompassed 30 small CGs (average size = 21.8 cM) bearing no locus-specific markers. These CGs thus cannot yet be assigned to homology groups.

Identification and mapping of a major rust resistance gene in MQ76-53

The distribution of rust resistance scores for 166 progeny clones is presented Fig. 1. Since the broad sense heritability of rust resistance scores at the experimental design level was very high ($h^2 = 0.96$), the distribution was based on mean scores for three replications. The results showed clear segregation between resistant clones (mean score < 2) and susceptible clones (mean score ≥ 2) with a segregation ratio of between 1:1 and 1:3 (Table 3). We already know that R570 has one copy of a rust resistance gene (*Bru*1) that has been mapped, and flanked with AFLP markers (Asnaghi et al. 2004). Two of these AFLP markers, i.e. aaccac6 mapped at 1 cM on the distal side of the gene and actctg9R at 2.2 cM on the proximal side of the gene, were also mapped in the present study. This gave us the opportunity to select clones that do not carry the *Bru*1 gene among the 166 clones evaluated for rust. We thus only selected clones without the two AFLP markers flanking *Bru1*. This left us with a subpopulation of 90 clones that did not bear the *Bru*1 gene—note that we assumed that the probability of having a double recombination (one recombination on each side of *Bru*1) within the 3.2 cM defined by the flanking AFLP markers was almost nil. The distribution of the rust resistance scores of these 90 clones presented in Fig. 1 shows a clear segregation

Table 3 Observed ratios of resistant and susceptible clones in three different populations and the Chi-square test of two hypotheses: one major gene is segregating in the population (1:1 ratio is expected) or two major genes are segregating (3:1 ratio is expected)

Studied populations	Parental rust resistance status		Number of progenies			χ^2	
	Female (R/S)	Male (R/S)	Total	Resistant	Susceptible	2 doses[a]	1 dose[rm a]
Full population of clones evaluated for rust resistance	R570 (R)	MQ76-53 (R)	166	105	61	11.66*** ↔	12.22***
Subpopulation of clones without markers flanking the R570 resistance gene *Bru*1	R570 (R)	MQ76-53 (R)	90	36	54	36.75***	3.60 NS
Control cross	B63-758 (S)	MQ76–53 (R)	133	68	65	6.61***	0.07 NS

[a] $\chi^2 = 3.84$ at the 5% level

between resistant clones (mean score <2) and susceptible clones (mean score ≥ 2). The segregation ratio was skewed but not significantly different from 1:1 (Table 3). This led us to hypothesize that a second rust resistance gene inherited from MQ76-53 was present. This hypothesis was confirmed by the 1:1 segregation ratio (68 resistant:65 susceptible clones) observed within 133 progenies derived from a cross between cultivars B63-758 (susceptible to rust) and MQ76-53 (Table 3).

For mapping purposes, rust resistance, in the subpopulation of 90 clones not bearing *Bru*1, was used as a morphological marker that was ranked 1 for resistant clones and 0 for susceptible clones. This marker was localized on CG3 of the MQ76-53 map (Fig. 2), which was assigned to HGVIII on the basis of two locus-specific markers (ADS7 and m38). A Fisher's exact test performed to assess associations between rust resistance and each of the 1666 segregating markers at the conservative threshold of $P = 3\times10^{-5}$ ($P = 0.05/1666$) revealed that the only markers significantly associated to the rust resistance gene belonged to CG3, except for an unmapped bi-parental single dose marker (Table 4). This test confirmed the localisation of the rust resistance gene on CG3 and showed that the most tightly associated marker was acgcta16 (Table 4), which was located 23.1 cM from the gene (Fig. 2).

The QTL analysis revealed two strong QTLs corresponding to the two genes described above. Marker aaccac6, which was closest to *Bru1*, explained 26% of the phenotypic variance whereas marker acgcta16, which was closest to the new rust resistance gene from MQ76-53, only explained 13% of the phenotypic variance. Both markers explain 38% of the phenotypic variation after multiple regression. However, in the subset of 90 clones not carrying *Bru1*, marker acgcta16 explained 35% of the phenotypic variance. Both genes probably have an effect of the same magnitude, but the closest marker associated with the new rust resistance gene detected in MQ76-53 was still located far from the gene. Moreover, as already observed by Daugrois et al. (1996) for *Bru1*, an important rust susceptibility level variation still exist in the susceptibility progeny class (Fig. 1) that can not be explained by the segregation of the two resistance genes.

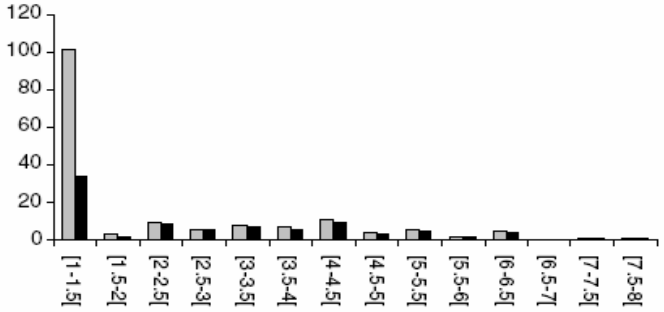


Fig. 1 Distribution of mean rust resistance scores. In *gray* (166 unselected clones of the progeny) and in *black* (90 clones of the progeny selected because they did not have any AFLP markers flanking the R570 rust resistance gene)

Identification and mapping of a major gene controlling stalk colour in MQ76-53

Stalk colour of 198 progeny clones was scored in three different environments (see Materials and methods). The

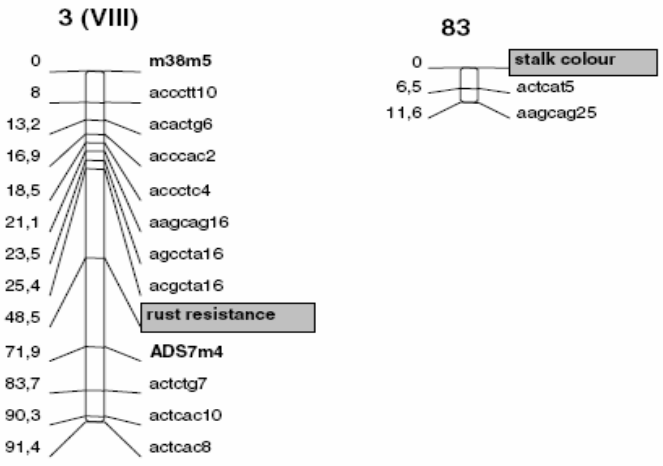


Fig. 2 Representation of cosegregation groups of the MQ76-53 genetic map bearing the putative new rust resistance gene (CG 3, HG VIII) and the putative gene controlling stalk colour (CG 83, unassigned to an HG). The detailed genetic map of MQ76-53 is available at http://www.tropgenedb.cirad.fr/en/sugarcane.html

Table 4 Significant associations between markers and rust resistance according to a Fisher exact test at $P < 3\times10^{-5}$, proportion of phenotypic variance (R^2) explained by these markers in the subset of 90 clones not carrying *Bru1*

Marker[a]	$P < F$	HG	CG	Position	R^2 (%)
actctg7	1.08×10^{-05}	VIII	3	83.7	29
ADS7_r4[a]	NS[a]	VIII	3	71.9[a]	37[a]
Putative gene		**VIII**	**3**	**48.5**	–
acgcta16	6.98×10^{-09}	VIII	3	25.4	35
agccta16	5.37×10^{-08}	VIII	3	23.5	29
aagcag16	3.77×10^{-07}	VIII	3	21.1	24
accctc4	3.77×10^{-07}	VIII	3	18.5	24
acccac2	2.36×10^{-06}	VIII	3	16.9	22
acactg6	8.30×10^{-06}	VIII	3	13.2	18
accctt10	1.47×10^{-05}	VIII	3	8	25
actctt26	1.73×10-05	Unlinked	–	–	7

[a]A lot of missing data for this RFLP marker

three evaluations were highly consistent, as indicated by the Fisher's exact test probabilities [$P = 1.15\times10^{-34}$ between environments (1) and (2); $P = 1.6\ 10^{-34}$ between (2) and (3); $P = 1.27\ 10^{-39}$ between (1) and (3)]. The evaluation was more difficult in environment (2) since the leaves had not been stripped off and since sunlight could not readily penetrate through the closed canopy in this field. Eighteen clones were not classified because of difficulties in deciding to which of the two categories they belonged, e.g. some clones appeared to be slightly red in some environments and not red in others, and some clones presented a peculiar striped colour pattern. These few ambiguous clones were not considered in the further analysis. Finally, a clear segregation was observed between 91 red stalk clones and 89 non-red stalk clones (this subpopulation of clones had green or yellow stalks). The resulting observed segregation ratio was not significantly different from 1:1 ($\chi^2 = 0.02$), suggesting that red stalk colour is under the control of a major gene inherited from MQ76-53. Stalk colour was therefore used as a morphological marker that, for mapping purposes, was ranked 1 for red stalk clones and 0 for non-red stalk clones. The major gene was localized on cosegregation group 83 on the MQ76-53 map, closely linked to an AFLP marker (actcat5) at 6.5 cM (see Fig. 2).

Discussion

The cumulative lengths of the genetic maps developed from our bi-parental population were 3,144 and 4,329 cM for R570 and MQ76-53, respectively. For R570, the map coverage (3,144 cM) was less than that of the map built from selfed R570 progeny (5,849 cM in Hoarau et al. 2001), although the number of AFLP primer combinations used was nearly the same. As expected, the same genotyping effort yielded less informative markers for mapping a given parent when comparing a bi-parental population to a selfed population. This is because it is not possible to use markers common to both parents for mapping. In our study, 168 bi-parental single dose markers common to R570 and MQ76-53 (identical alleles or homoplasic bands) could not be used for mapping, whereas most of them had been mapped in the selfed R570 progeny. These non-mappable, bi-parental single dose markers represented at least 26% of the single dose markers in the R570 genome in this study. Moreover, this percentage marker loss is underestimated since it does not take into account single dose R570 markers not coded due their presence in multiple doses in MQ76-53.

MQ76-53 is an F1 hybrid between an old cultivar and a wild *S. spontaneum* clone (SES528) that is not found in the recorded pedigree of any international cultivars (Machado 2001). Therefore this clone is presumably quite genetically distant from R570 and from other present-day cultivars. This was confirmed by an AFLP-based diversity study that included R570, MQ76-53 and 72 modern cultivars from various breeding programs. This study revealed that 40 to 65% of the single dose AFLP bands mapped in R570 are common with the other cultivars (unpublished data). These figures provide an interesting estimation of the extent of informative (=single dose) marker loss that could occur in a bi-parental mapping project. In the present study, we loosed 40% of single dose markers that were common to R570 and MQ76-53. However, in a cross between two standard modern cultivars this proportion should be higher. Despite the lower yield of mappable markers, a bi-parental mapping population seems more suitable than a selfed population when QTL detection is the ultimate objective, especially if the two parents have highly contrasted phenotypes for the trait of interest. Indeed, in bi-parental progeny: (1) the two populations to be compared (the one with the marker to be tested and the one without it) are of much more balanced size (1:1) than in a selfed progeny context (3:1); (2) the buffering effect of the background of alternative alleles should be less strong since half of this background is inherited from the "contrasted parent". QTL detection power is therefore likely to be greater in a bi-parental context, with other parameters being equal (population size, type 1 error, etc.), and should help to more accurately detect the most favourable alleles.

After alignment of the different R570 AFLP-based maps, cosegregation groups of R570 were grouped into 7 HGs containing between 2 (HG IV) and 22 (HG VI) CGs. Structural differences between *S. spontaneum* and *S. officinarum* are expected since they have different basic numbers, i.e. $x = 8$ and $x = 10$, respectively (D'Hont et al. 1998). These differences may result from simple fusion or fission events, as suggested for HG VIII in which two sets of homologous *S. officinarum* chromosomes are assigned to the same HG (VIII) because of their homology to the same *S. spontaneum* CG (Grivet et al. 1996; D'Hont et al. 1996, this study). Considering the basic chromosome number of *S. spontaneum*, we would have expected 8 HGs instead of the seven found. Tentative explanations could be proposed: (1) genome

coverage with locus-specific markers is still insufficient for identifying all HGs (some unassigned CGs may belong to missing HGs); (2) duplication of some SSR and RFLP locus within the basic chromosome set may lead to misassembly of sets of homologous CGs into the same HG (Butterfield et al. 2001); (3) complex structural differences between the basic chromosome sets of *S. officinarum* and *S. spontaneum* may also lead to misassembly of sets of homologous or homeologous CGs. Mapping additional locus-specific markers (SSRs and RFLPs) should help to resolve this question.

No linkage in repulsion was detected between CGs of MQ76-53, while preferential pairing was observed for 18 CG pairs in R570. This contrasting situation may in part be explained by the differences in genome constitution between these two clones. The MQ76-53 clone is the result of a cross between a cultivar (Trojan) and a *S. spontaneum* clone (SES528) with $2n=64$ chromosomes (Panje and Babu 1960). The *S. spontaneum* component of the MQ76-53 genome is therefore composed of the *S. spontaneum* chromosome inherited from SES528 plus the *S. spontaneum* chromosomes inherited from Trojan. By comparison, R570 is composed of only one or two *S. spontaneum* chromosomes per homologous class. In MQ76-53 no preferential pairing could be detected within *S. spontaneum* at LOD = 3. This picture is in agreement with the lack of preferential pairing noted in *S. spontaneum* SES208, which suggests that there is polysomic inheritance and autopolyploidy in this species (Al Janabi et al. 1993; Ming et al. 1998). However, the absence of preferential pairing among MQ76-53 *S. officinarum* chromosomes is out of line with the incomplete polysomy observed in *S. officinarum* (Mudge et al. 1996; Guimaraes et al. 1997). This could be due to a dramatic deficiency in the coverage of the *S. officinarum* component of the MQ76-53 genome which would make it almost invisible. Chromosome pairing behaviour is complex in R570. Preferential pairing involves CGs of *S. officinarum* origin, CGs of *S. spontaneum* origin, as well as *S. spontaneum* × *S.officinarum* recombinant CGs (Hoarau et al. 2001; Grivet et al. 1996). In-depth analysis of all pairing frequencies between homologous and homeologous chromosomes of a single homology group (HGI) in R570 revealed pairing affinities between chromosomes ranging from 0 to 40%. These chromosome affinities are only partly explained by the species origin of the chromosomes (Jannoo et al. 2004).

A Mendelian factor governing stalk colour has been identified in the MQ76-53 genome. This gene is linked to two AFLP markers in a cosegregation group (CG 83). Although cane colour varies considerably according to age of the stalk and to the amount of sunlight received (Stevenson 1965), a rather clear 1:1 segregation ratio between red stalk cane and non-red stalk cane was noted. This finding may not be relevant from an agronomic standpoint but is another indication that even in the complex polyploid background of sugarcane it is possible to identify traits controlled by a single gene, and that genetic mapping in sugarcane should be continued to identify new major genes. In addition, a new rust resistance gene has been located in the MQ76-53 genome. This gene clearly differs from the R570 rust resistance gene *Bru*1 (Asnaghi et al. 2004). Indeed, none of the AFLP markers flanking *Bru1* in R570 is present in MQ76-53, indicating that this new resistance gene belongs to a different haplotype. Moreover, this gene is located on CG3, which should belong to HG VIII, whereas *Bru*1 is located on a R570 CG that belongs to HG VII. Rossi et al. (2003) identified three resistance gene analog clusters, including RGA of the NBS-LRR and S/T kinase types, in HG VIII. These RGAs could be used as candidate genes and mapped in our population to determine whether they are located in the vicinity of this new rust resistance gene. This approach could help to more rapidly pinpoint the exact location of the gene (Pflieger et al. 2001). In addition, it would be interesting to trace the origin of this resistance gene. It could represent a yet untapped alternative source of rust resistance if it turns out to have been inherited from the *S. spontaneum* SES528 clone.

Acknowledgements We thank Iréné Promi for his help in the fieldwork. The experiments presented in this paper were conducted in compliance with current French legislation.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110:789–801

Al-Janabi SM, Honeycutt RJ, McClelland M, Sobral BWS (1993) A genetic linkage map of *Saccharum spontaneum* L 'SES 208'. Genetics 134:1249–1260

Asnaghi C, Roques D, Ruffel S, Kaye C, Hoarau JY, Telismart H, Girard JC, Raboin LM, Risterucci AM, Grivet L, D'Hont A (2004) Targeted mapping of a sugarcane rust resistance gene (*Bru*1) using bulked segregant analysis and AFLP markers. Theor Appl Genet 108:759–764

Bhat SR, Gill BS (1985) The implication of 2*n* egg gametes in nobilisation and breeding of sugarcane. Euphytica 34:377–384

Burner DM, Legendre BL (1993) Chromosome transmission and meiotic stability of sugarcane (Saccharum spp) hybrid derivatives. Crop Sci 33:600–606

Burner DM, Legendre BL (1994) Cytogenetic and fertility characteristics of elite sugarcane clones. Sugar Cane 1:6–10

Butterfield MK, D'Hont A, Berding N (2001) The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. Proc S Afr Sug Technol Ass 75:1–5

Da Silva JAG, Burnquist WL, Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome 36:782–791

Da Silva J, Honeycutt RJ, Burnquist W, Al-Janabi SM, Sorrells ME, Tanksley SD, Sobral BWS (1995) *Saccharum spontaneum* L 'SES 208'genetic linkage map combining RFLP- and PCR-based markers. Mol Breed 1:165–179

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC, D'Hont A (1996) A putative major gene for rust resistance linked with an RFLP marker in sugarcane cultivar R570. Theor Appl Genet 92:1059–1064

D'Hont A (2005) Unravelling 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, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers, a first decade of research. Proc Int Soc Sugarcane Technol 24: 556–559

D'Hont A, Lu YH, Gonzàlez de Leòn D, Grivet L, Feldmann P, Lanaud C, Glaszmann JC (1994) A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the *Andropogoneae* tribe. Genome 37:222–230

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp) by molecular cytogenetics. Mol Gen Genet 250:405–413

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus Saccharum by physical mapping of ribosomal RNA genes. Genome 41:221–225

D'Hont A, Garsmeur O, Raboin LM, Paulet F, Begum D, Wing R, Glaszmann JC (2001) Chromosome walking toward a major resistance gene for common rust of sugarcane. Proc Int Soc Sugarcane Technol 24: 315–317

Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122–127

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC (1996) RFLP mapping in a highly polyploid and aneuploid interspecific hybrid. Genetics 142:987–1000

Guimaraes CT, Sills GR, Sobral BWS (1997) Comparative mapping of Andropogoneae: *Saccharum* L (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci USA 94:14261–14266

Heinze BS, Thokoane LN, Williams NJ, Barnes JM, Rutherford RS (2001) The smut–sugarcane interaction as a model system for the integration of marker discovery and gene isolation. Proc S Afr Sug Technol Ass 75:88–93

Hoarau JY, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp) I Genome mapping with AFLP markers. Theor Appl Genet 103:84–97

Hoarau JY, Grivet L, Offmann B, Raboin LM, Diorflar JP, Payet J, Hellmann M, D'Hont A, Glaszmann JC (2002) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp) II Detection of QTLs for yield components. Theor Appl Genet 105:1027–1037

Hoisington D (1992) Laboratory protocols. CIMMYT applied molecular genetics laboratory, Mexico

Jannoo N, Grivet L, D'Hont A, Glaszmann JC (2004) Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. Heredity 93:460–467

Jordan DR, Casu RE, Besse P, Carroll BC, Berding N, McIntyre CL (2004) Markers associated with stalk number and suckering in sugarcane colocate with tillering and rhizomatousness QTLs in sorghum. Genome 47:988–993

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Machado GR (2001) Sugarcane variety notes—an international directory seventh revision, Piracicaba, Int. Soc. Sugar Cane Techn. Februray 2001, 132p

Mather K (1957) The measurement of linkage in heredity. Methuen, London

Mehta CR, Patel NR (1983) A network algorithm for performing Fisher's exact test in rxc contingency tables. J Am Statist Assoc 78:427–434

Ming R, Liu S-C, Lin Y-R, Da Silva J, Wilson W, Braga D,Van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH (2002a) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome 45:794–803

Ming R, Liu SC, Bowers JE, Moore PH, Irvine JE, Paterson AH (2002b) Construction of a Saccharum consensus genetic map from two interspecific crosses. Crop Sci 42:570–583

Mudge J, Andersen WR, Kehrer RL, Fairbanks DJ (1996) A RAPD genetic map of *Saccharum officinarum*. Crop Sci 36: 1362–1363

Panje RR, Babu CN (1960) Studies in *Saccharum spontaneum* distribution and geographical association of chromosome numbers. Cytologia 25:152–172

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291

Pinto LR, Oliveira KM, Ulian EC, Garcia AA, de Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47(5):795–804

Price S (1963) Cytogenetics of modern sugar canes. Econ Bot 17:97–105

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genet 269:406–419

Ruiz M, Rouard M, Raboin LM, Lartaud M, Lagoda P, Courtois B (2004) Tropgene-DB, a multitropical crop information system. Nucleic Acids Res 32:D364-D367

SAS Institute (1990)SAS procedures guide, ver 6, 3rd edn. SAS Institute Inc, Cary

Stevenson GC (1965) Genetics and breeding of Sugar cane. Longmans, London, pp 284

Tai PYP, Miller JD, Dean JL (1981) Inheritance of resistance to rust in the sugarcane field. Crops Res 4:261–268

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wu KK, Burnquist W, Sorrels ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genet 83:294–300

10.2 Anexo II

Artigo: "Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*"

Louis-Marie Raboin, Athiappan Selvi, Karine Miranda Oliveira, Florence Paulet, Caroline Calatayud, Marie-Françoise Zapater, Philippe Brottier, Rosalyn Luzaran, Olivier Garsmeur, Jean Carlier, Angélique D'Hont

Aceito para publicação na revista

Fungal Genetics and Biology

ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology xxx (2006) xxx–xxx
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*

Louis-Marie Raboin [a,1], Athiappan Selvi [b,1], Karine Miranda Oliveira [c], Florence Paulet [d], Caroline Calatayud [d], Marie-Françoise Zapater [e], Philippe Brottier [f], Rosalyn Luzaran [g], Olivier Garsmeur [d], Jean Carlier [e], Angélique D'Hont [d,*]

[a] *CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), UMR PVBMT, Pôle de Protection des Plantes, Ligne Paradis, 97410 Saint-Pierre, Réunion, France*
[b] *Sugarcane Breeding Institute, Coimbatore, India*
[c] *Universidade Estadual de Campinas, Brazil*
[d] *CIRAD, UMR 1096, PIA, TA40/03, Avenue Agropolis, 34398 Montpellier cedex 5, France*
[e] *CIRAD, UMR BGPI, TA 41 / K, Campus International de Baillarguet 34398 Montpellier Cedex 5, France*
[f] *Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, 91057 Evry, France*
[g] *PHILSURIN, 6776 Ayala Avenue, Makati City 1226, Philippines*

Received 4 April 2005; accepted 17 July 2006

Abstract

The basidiomycete *Ustilago scitaminea* Sydow, which causes sugarcane smut disease, has been spreading throughout Africa and America since the 1940s. The genetic diversity and structure of different populations of this fungus worldwide was investigated using microsatellites. A total of 142 single-teliospore were isolated from 77 distinct whips (sori) collected in 15 countries worldwide. Mycelium culture derived from on generation of selfing of these single teliospores were analysed for their polymorphisms at 17 microsatellite loci. All these strains but one were homozygous at all loci, indicating that selfing is likely the predominant reproductive mode of *U. scitaminea*. The genetic diversity of either American or African *U. scitaminea* populations was found to be extremely low and all strains belong to a single lineage. This lineage was also found in some populations of Asia, where most *U. scitaminea* genetic diversity was detected, suggesting that this fungal species originated from this region. The strong founder effect observed in *U. scitaminea* African and American populations suggests that the fungus migrated from Asia to other continents on rare occasions through movement of infected plant material.

Keywords: Sugarcane; Smut; *Ustilago scitaminea*; Population genetic structure; Selfing; Outcrossing; Fungus dispersal; Microsatellites

1. Introduction

Smut caused by the fungus *Ustilago scitaminea* Sydow is a major sugarcane disease that can cause considerable yield losses (Lee-Lovick, 1978). It is easily identified by the characteristic whip-like sorus that forms from the apex of infected stalks spreading huge quantities of spores. The use of resistant cultivars is the only efficient way to control this disease. Sugarcane resistance to smut was demonstrated to be a fairly heritable trait (Chao et al., 1990; Wu et al., 1988) although its genetic determinism is still unknown. Reports on possible race–cultivar interactions in Hawaii (Comstock and Heinz, 1977), Brazil (Toffano, 1976), Pakistan (Muhammad and Kausar, 1962) and Taiwan (Leu and Teng, 1972) suggest the possible involvement of major resistance genes.

Before the 1940s, sugarcane smut was recorded in Asia and a few countries from South and East Africa (Natal, Mozambique, Zimbabwe) and Mascarenes (Madagascar, Mauritius and Reunion) (Antoine, 1961). Since then, this

* Corresponding author. Fax: +33 0 4 67 61 56 05.
[1] These authors contributed equally to this work.

1087-1845/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2006.07.004

disease has spread to nearly all sugarcane producing countries (Comstock, 2000). It was first recorded in America in 1940 in Tucuman province of Argentina (Hirshhorn, 1949) and spread to Paraguay in 1944 (James, 1978) and to Brazil in 1948 (Veiga, 1972) but no further extension of the epidemic was observed. In West Africa, the disease was first detected in Tchad in 1969. From there, cuttings were sent to Burkina Faso triggering a new epidemic (Barat, 1971). Later smut was found in Mali, Ivory Coast in 1972 and Niger in 1973 (Delassus, 1975). Smut was recorded in 1971 on Hawaii's island of Oahu (Ferreira and Comstock, 1989) and in the Caribbean state of Guyana in 1974 (Bates, 1975). It then spread to all the Caribbean islands, central America and mainland USA. In Australia, the disease was first observed in 1998 in a marginal production area on the west coast (Riley et al., 1999) and was detected in the main producing area (east coast) in 2006. Only Papua New Guinea which is considered as the center of origin of *Saccharum officinarum*, the main species involved in modern cultivars, and Fiji Islands are still free of this disease.

Sugarcane smut is a basidiomycete belonging to the Ustilaginales order. These fungi produce diploid spores called teliospores (Martinez-Espinoza et al., 2002). When it germinates, the teliospore undergoes meiosis and gives rise to a septate promycelium bearing four haploid sporidia (basidiospores). *Ustilago scitaminea*, like most parasitic heterobasidiomycetes (Bakkeren and Kronstad, 1994), has a diallelic bipolar mating system (Alexander and Srinivasan, 1966; Leu, 1978) in which only sporidia of opposite mating types conjugate. Of the four initial sporidia or basidiospores from each teliospore, two have a positive mating allele and two have a negative mating allele. *U. scitaminea* can thus both self and outcross, but the frequency of natural selfing *versus* outcrossing is unknown. A dikaryotic mycelium develops after fusion of compatible sporidia. This dikaryotic mycelium is infectious, penetrates behind bud scales and invades the meristematic zone of the bud. The mycelium systematically colonizes the plant by growing in association with each developing bud primordium (Ferreira and Comstock, 1989). Finally, the apical meristem of smut-infected cane produces a long whip-like structure bearing billions of teliospores (i.e. sorus). The sorus is covered by a thin silvery membrane which soon ruptures and the spores are dispersed by the wind.

Knowledge on the genetic structure and evolutionary potential of pathogens are essential for breeding and management of plant resistance (Leung et al., 1993; McDonald and Linde, 2002). Population structure analysis could also help identify potential sources of resistance, which are expected in areas where genetic diversity of both pathogen and host is maximal. Finally, these analyses may be used to evaluate the relative importance of factors such as genetic recombination and gene flow in the evolution of these pathogens. These factors, related to the reproduction system and dispersal processes, may condition the choice of resistance breeding and deployment strategies (McDonald and Linde, 2002). Ustilaginales have been poorly investigated in this respect, although they are important plant pests and considered as a model system (Martinez-Espinoza et al., 2002).

In *Ustilago scitaminea*, the lack of reliable morphological and physiological markers has hampered population genetics studies, despite a few attempts (Lambat et al., 1968; Amire et al., 1982). Only recently, a preliminary report (Braithwaite et al., 2004) suggested that *U. scitaminea* populations have a very low level of genetic diversity except in Asia. However, this study was limited to a small number of isolates and the use of AFLP markers as well as the sampling methodology prevented a rigorous description of genetic relationships between genotypes and inferences on the reproduction system. In the present study, we have developed microsatellite markers for the analysis of *U. scitaminea* populations. These markers have the advantage of being codominant and highly polymorphic and were used to type strains derived from single-teliospore isolates from worldwide sugarcane-producing countries. The objectives of this study were (1) to better describe the genetic diversity of different populations of *U. scitaminea* worldwide using a larger number of isolates than previous studies, (2) to describe the genetic relationship between populations from the different geographical origins, and (3) to infer the reproduction mode of these populations.

2. Material and methods

2.1. Fungal strains

A total of 77 single whips (sori) were collected on various sugarcane cultivars grown in 15 countries worldwide (Table 1). For four regions, a significant number of samples were collected in a limited number of localities: Reunion island (11 sori/7 localities), the Philippines (15 sori/8 localities), Colombia (10 sori/3 localities) and Burkina Faso (11 sori from various plots at a single location, the SOSUCO plantation). For Colombia, the Philippines and Reunion, samples were collected from the same region in fields several tens of kilometers apart. This collection therefore represents the diversity of *Ustilago scitaminea* populations at a regional scale. The Burkina samples were collected from various plots within the same commercial plantation.

Teliospores from each sorus were plated on solid agar medium for germination. One to three single germinating teliospores per sori, corresponding in total to 142 single spores, were isolated and grown on potato dextrose agar at 26 °C for 1 week. The resulting white mycelial colonies were further multiplied in order to produce sufficient fungal material. The mycelium was dried overnight at 60 °C, stored at −20 °C and later used for DNA extraction. The white mycelial-type colonies correspond to dikaryotic hyphae (Trione, 1990). Dikaryotic hyphae arise from the fusion of haploid sporidia produced by the diploid teliospore after meiosis and thus represent selfed dikaryotic progenies issued from isolated teliospores. Although karyogamy has not occurred DNA extracts correspond to two haploid

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 1
Origin of the142 *Ustilago scitaminea* isolates studied

Isolate whip	Number code[a]	Country	Locality	Variety	Contact	Year	nb
		America					19
031	a	Brazil	Piracicaba	IAC 66/6	Copersucar Technologie Center/M.V. Casagrande	2002	1
033	a	Brazil	Piracicaba	NA 56/79	Copersucar Technologie Center/M.V. Casagrande	2002	1
030	b	Brazil	Piracicaba	SP 70/11343	Copersucar Technologie Center/M.V. Casagrande	2002	1
032	a	Brazil	Piracicaba	SP 80/185	Copersucar Technologie Center/M.V. Casagrande	2002	1
034	a	Brazil	Jau	Pool	Copersucar Technologie Center/M.V. Casagrande	1997	1
046	a	Colombia	Cenicana	CC 84/75	CENICANA/J.C. Angel	2002	1
047	a	Colombia	Cenicana	CO 421	CENICANA/J.C. Angel	2002	1
048	a	Colombia	Cenicana	CP 57/603	CENICANA/J.C. Angel	2002	1
050	a	Colombia	Central sicarare	B 61	CENICANA/J.C. Angel	2002	1
049	a	Colombia	Central sicarare	CC 89/452	CENICANA/J.C. Angel	2002	1
052	b	Colombia	Central sicarare	CC89/43	CENICANA/J.C. Angel	2002	1
051	a	Colombia	Central sicarare	PH 5333	CENICANA/J.C. Angel	2002	1
054	a	Colombia	Cucuta	B 49/191	CENICANA/J.C. Angel	2002	1
055	b	Colombia	Cucuta	CO 421	CENICANA/J.C. Angel	2002	1
053	b	Colombia	Cucuta	RD 75/11	CENICANA/J.C. Angel	2002	1
007	d	Guadeloupe	Roujol	HJ 57/41	CIRAD/J.H. Daugrois	2002	1
004	a	Hawai	Kauai	H 77/4643	Hawaii agriculture Research Center/S. Schenck	2002	1
003	a	Hawai	Maui Island	H 78/7750	Hawaii agriculture Research Center/S. Schenck	2002	1
001	e	Texas	Weslaco	NCo310	Texas Agricultural Experiment Station/J. Da Silva	2002	1
		Asia					64
041	a	India	Gorakhpur	Unknown	U.P. Council of Sugarcane Research/G.P Rao	2002	1
079	a,b,c	Indonesia	3rd division	RGM 94/137	GPM Sugar group/S. Lamadji	2002	3
082	a,b,c	Indonesia	5th division	GM 23	GPM Sugar group/S. Lamadji	2002	3
078	a,b,c	Indonesia	5th division	RGM 94/122	GPM Sugar group/S. Lamadji	2002	3
015	a,b	Indonesia	ILP factory	Berat	GPM Sugar group/S. Lamadji	2002	2
021	a,b	Japan	Nago	Ni 9	JIRCAS/M. Matsuoka	2002	2
020	a,b	Japan	Naha	RK 97/35	JIRCAS/M. Matsuoka	2002	2
017	a,b	Japan	Naha	RK 97/7020	JIRCAS/M. Matsuoka	2002	2
019	a,b	Japan	Naha	RK 97/7032	JIRCAS/M. Matsuoka	2002	2
018	a,b	Japan	Naha	RK 97/7065	JIRCAS/M. Matsuoka	2002	2
045	a,b	Philippines	Luy-a	84524	Philippine Sugar Research Inst. Foundation/R. Cu	2002	2
042	a,b	Philippines	Luy-a	86550	Philippine Sugar Research Inst. Foundation/R. Cu	2002	2
097	a,b,c	Philippines	Luy-a	HOCP 91/555	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
099	a,b,c	Philippines	Luy-a	VMC 95/100	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
098	a,b	Philippines	Luy-a	VMC 95/104	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	2
096	a,b,c	Philippines	Luy-a	VMC 95/262	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
093	a,b,c	Philippines	Hda Cansilayan	Phil 74/64	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
085	a,b,c	Philippines	Hda Florencia	VMC 84/524	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
084	b,c	Philippines	Hda Florencia	VMC 86/551	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	2
090	a,b	Philippines	Hda Linay	Phil 74/64	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	2
092	c	Philippines	Hda Linay	VMC 86/550	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	1
087	b,c	Philippines	Hda Santa Cruz	VMC 88/354	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	2
088	a,b,c	Philippines	Hda Tinihaban	Phil 74/64	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	3
091	a,b	Philippines	La Granja	Phil 80/13	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	2
094	b	Philippines	MDDC Nursery	VMC 86/550	Philippine Sugar Research Inst. Foundation/R. Luzaran	2003	1
027	a,b	Thailand	Ban Bua	K88-92	Mitr Phol Sugarcane Research Center/M.S. Rao	2002	2
025	a,b	Thailand	Ban Bua	Marcos	Mitr Phol Sugarcane Research Center/M.S. Rao	2002	2
028	a,b	Thailand	Ban Kud Jork	K86-161	Mitr Phol Sugarcane Research Center/M.S. Rao	2002	2
026	a,b	Thailand	Nong Kon thai	K84-200	Mitr Phol Sugarcane Research Center/M.S. Rao	2002	2
		Africa					59
114	a,b,c	Burkina Faso	plot BN2	SP 70/1006	SN SOSUCO/Balaya	2003	3
112	a,b,c	Burkina Faso	plot BS2	Co 64/15	SN SOSUCO/Balaya	2003	3
115	a,b,c	Burkina Faso	plot BS5	PR 1007	SN SOSUCO/Balaya	2003	3
104	b	Burkina Faso	plot diaraba	Nco 310	SN SOSUCO/Balaya	2003	1
110	a,b,c	Burkina Faso	plot NN12	SP 70/1006	SN SOSUCO/Balaya	2003	3
113	a,b,c	Burkina Faso	plot NS16	R 570	SN SOSUCO/Balaya	2003	3
111	a,b,c	Burkina Faso	plot NS28	Co 64/15	SN SOSUCO/Balaya	2003	3
109	b	Burkina Faso	plot PK1 SNSo	Co 64/15	SN SOSUCO/Balaya	2003	1
108	a,b,c	Burkina Faso	plot PL4	R 570	SN SOSUCO/Balaya	2003	3
105	a,b,c	Burkina Faso	plot PL4 SnSo	PR 1007	SN SOSUCO/Balaya	2003	3
106	a,b,c	Burkina Faso	plot PS1	PR 1007	SN SOSUCO/Balaya	2003	3

(continued on next page)

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 1 (*continued*)

Isolate whip	Number code[a]	Country	Locality	Variety	Contact	Year	nb
118	a	Mauritius	Belle vue	M 1030/71	MSIRI/A. Dookun-Saumtally	2003	1
119	a,b,c	Mauritius	Pamplemouse	M 376/64	MSIRI/A. Dookun-Saumtally	2003	3
117	a,b,c	Mauritius	Petite rivière	M 1030/71	MSIRI/A. Dookun-Saumtally	2002	3
056	a	Réunion	La Mare	Unknown	CERF/J.Y. Hoarau	1998	1
057	a	Réunion	La Mare	Unknown	CERF/J.Y. Hoarau	1998	1
070	a,b	Réunion	Gol 343	R 00/122	CERF/J.Y. Hoarau	2002	2
065	a,b	Réunion	La Bretagne	R 85/1238	CERF/J.Y. Hoarau	2002	2
058	a,b	Réunion	Saint Benoit 111	R 88/215	CERF/J.Y. Hoarau	2002	2
075	a,b,c	Réunion	Ligne Paradis	Unknown	CERF/J.Y. Hoarau	2002	3
067	a,b	Réunion	Saint Benoit 112	R 90/4143	CERF/J.Y. Hoarau	2002	2
074	a,b	Réunion	ST Phillippe 415	R 83/267	CERF/J.Y. Hoarau	2002	2
076	a,c	Réunion	ST Phillipe 415	R 83/644	CERF/J.Y. Hoarau	2002	2
060	b,c	Réunion	Vue Belle 205	R 91/4297	CERF/J.Y. Hoarau	2002	2
013	c	Senegal	CSS A4	N 14	Compagnie Sucrière Sénégalaise/R. Toll	2002	1
014	a	Senegal	CSS A7	N 14	Compagnie Sucrière Sénégalaise/R. Toll	2002	1
011	a	Senegal	CSS A5a	N 14	Compagnie Sucrière Sénégalaise/R. Toll	2002	1
024	b	South Africa	Pongola	Unknown	SASEX/I. Antwerp	2002	1

[a] Letter code corresponding to the distinct single spore strains from the same whip that have been genotyped.

nuclei co-existing in the same cell, which is equivalent to the DNA content of the diploid phase.

Finally, a total of 142 mycelium cultures derived from 142 single spore isolates after one generation of selfing were obtained, they will be further refereed as "strains". A larger number of strains (38) were sampled in the Philippines since a high genetic diversity among *Ustilago scitaminea* strains was revealed in this region in our preliminary analyses.

2.2. DNA extraction

Genomic DNA was extracted using a modification of the method described by Rogers and Bendich (1988). Approximately 350 mg of dried mycelium was ground with sand, dispersed in 800 µl of cetylmethylammonium bromide (CTAB) extraction buffer (1% CTAB, 0.7 M NaCl, 50 mM Tris–HCl, pH 8.0, 10 mM EDTA, pH 8.0, 1% *B*-mercaptoethanol, 100 µg/ml RNase) and incubated at 65 °C for 60 min with continuous gentle rocking. Two successive chloroform:isoamyl alcohol (24:1) purifications were performed. The DNA was precipitated by adding 0.8 vol isopropanol and the DNA pellet was washed with 2 ml of 70% ethanol. The DNA pellet was air dried for 20 min and then resuspended in 200 µl of TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, pH 8.0). The extractions are yielded from 3–10 µg of DNA.

2.2.1. Construction of a microsatellite enriched genomic library and definition of microsatellite loci specific primers

A genomic library enriched with GA and GT microsatellite sequences was produced according to the protocol previously described by Billotte et al. (1999) with the following modifications. Total DNA of *U. scitaminea* was restricted with the endonuclease *Rsa*I. The microsatellite sequences were selected using two biotin-labelled oligoprobes in a mix for simultaneous selection of CA and GA repeats: 5′-I∗IIIITCTCTCTCTCTCTC-3′ and 5′-I∗IIIITGTGTGTGTGTGTG-3′, where I∗ is a biotinylated inosine. After ligation, transformation was made in the Epicurian XL1-blue *Escherichia coli* strain. A total of 480 clones were transferred to microplates containing 150 µl of LB/ampicillin solution. Insert sizes were estimated by agarose gel electrophoresis of insert PCR products using M13 universal primers. Gels were alkaline Southern transferred to Hybond N^+ nylon membranes (Amersham) and hybridized with γ-^{32}P-radiolabeled $(GA)_{15}$ and $(GT)_{15}$ probes. Inserts of 200 clones that gave a satisfactory positive signal with a size over 500 bp were sequenced by the Centre National de Séquençage (CNS, Génoscope, Evry, France). Primer pairs were designed using the online Primer 3 package (Whitehead Institute for Biomedical Research) for 38 inserts containing more than five repeated dinucleotides motifs. A total of 22 microsatellite loci were used for the present study (Table 2).

2.3. Microsatellite analysis

The forward primers were end-labeled with [γ-^{33}P]ATP, and amplification was performed in an MJ Research PTC 100 Thermal Cycler in 20 µl reaction mixtures containing 10 ng of template DNA, 200 µM dNTP mix, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris–HCl (pH 8.3), each primer at 0.2 µM, and 1 U of *Taq* polymerase. The samples were denatured at 94 °C for 4 min and subjected to 35 cycles of 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1 min. After the addition of 20 µl of loading buffer (98% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol), the amplified products were denatured at 94 °C for 3 min, and 5 µl of each sample was loaded onto a 5% polyacrylamide gel with 7.5 M urea and electrophoresed in 1% TBE buffer at 60 W for 1 h 40 min. The 30–330 bp AFLP DNA ladder was used in order to estimate the size of AFLP fragments. The gels were dried for 30 min at 80 °C and exposed for 4 days to X-ray film (Fuji RX).

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 2
Characteristics of the 22 *Ustilago scitaminea* SSR loci analysed (T_a, annealing temperature, *n*, number of alleles detected in this study)

Embl accession number	Name	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Repeat motif	T_a (°C)	Product size	*n*
AM397247	mUsCIR1	GTCGCTCGTTTCACTT	CTACCTCGCTATCCCTT	$(GT)_{18}$	50	175	4
AM397248	mUsCIR2	ATAGAAATGGGCTGAAGA	GATTGATACCGAGAGGG	$(CA)_3A_5(CA)_6$	51	192	1
AM397249	mUsCIR3	ACTGTAGAACTGGTGCCT	TTTCTTTTTCCTCTCCC	$(GT)_6$	50	309	2
AM397250	mUsCIR4	ACAAACCATCCCCTCT	CGTGGGTATTCTACTGCT	$(GT)_6(GA)_2$	51	196	1
AM397251	mUsCIR5	CGATACCATCCACACC	AGTGCGTTTAGGTTTTTC	(TGC)5	50	309	1
AM397252	mUsCIR6	ATCTTGCGTCTTCTGTATC	CGGTCAAGGTGTTTTC	$(CA)_7$	50	183	1
AM397253	mUsCIR7	TCGTTCATCCCTCTATG	GAGAAAGAGCAGTAACCC	$(CT)_2G(CT)_7$	49	287	3
AM397254	mUsCIR8	TCTCAGGTTTATTCGAT	AACATACCCTCTTTCAC	$(GT)_6GC(GT)_3GC(GT)_4AAA(G\ T)_3GC(GT)_3GC(GT)_4AAA(GT)_3GC(G\ T)_2$	45	204	2
AM397255	mUsCIR9	CTCTTGGGTTTTCTTTTC	ACATAGACCCACACCTTC	$(CT)_{23}$	50	269	9
AM397256	mUsCIR10	AGAGCTTAGGAGGCAAC	GCTACTCCCCACATCA	$(GT)_7$	50	164	2
AM397257	mUsCIR12	AGCCACCACTACTCATTT	ACGGTCTTACCCATTCT	$(CA)_{28}$	50	205	4
AM397258	mUsCIR13	ATTGGAAAGAAAGGAGG	TTTCCGACTGATTACTGA	$(GA)_{18}$	50	168	4
AM397259	mUsCIR15	CACCACTTCTCTACTTGCT	GCTCGTTTGAGGCTAC	$(GA)_6(GT)_7$	50	121	2
AM397260	mUsCIR18	GCTTCTCCTTTTTGCTT	ATCCAGCGTTCTCTCTC	$(GT)_{15}$	50	270	3
AM397261	mUsCIR19	AACCTCGCACTTGATT	TGAAAGGTTGTATTGGG	$(CA)_4CG(CA)_6AA(CA)_4$	49	158	1
AM397262	mUsCIR21	ACTACTTTAGGCGAGGG	TCCGTGAATGCTTATTT	$(GT)_{17}$	50	144	4
AM397263	mUsCIR29	CATTTCACTCGTCTTCTTT	ATTTGGTCTGTGTGGTG	$(CA)_{14}$	50	203	2
AM397264	mUsCIR30	AAGGTGGCGAACAGAG	AATGTAGGGCGAAAGG	$(GA)_7(GT)_8(GA)_{12}$	52	109	3
AM397265	mUsCIR32	TCGTTAGGGAAAGGTCT	CGGTTAGGTTGAGTGTG	$(GT)_{11}$	50	265	2
AM397266	mUsCIR33	GCTCGTTTGAGGCTAC	CACCACTTCTCTACTTGCT	$(CA)_6(CT)_7$	50	121	2
AM397267	mUsCIR34	CGATTATTTTGTCTATGCTC	TGAGGTGTTTGTGGTTT	$(CA)_{11}$	50	278	3
AM397268	mUsCIR38	TTCTTGACTGGCGATT	GATGACTTACCTCTGGCT	$(CA)_5TACATA(CA)_9$	50	273	2

2.4. Data analysis

Genetic dissimilarities were computed from the microsatellite data using the shared allele distance (Chakraborty and Jin, 1993) in the Darwin software package (Perrier et al., 2003). A pair-wise dissimilarity matrix was used to construct a neighbour joining (N-J) tree. A bootstrapping procedure was performed over 1000 replications. Allele frequencies, unbiased estimates of expected heterozygosity under Hardy Weinberg assumptions, observed heterozygosity, percentage of polymorphic loci when most frequent allele did not exceed 95% and mean number of alleles per locus were computed using Genetix version 4.03 software (Belkhir et al., 2004).

3. Results

Genetic diversity and heterozygosity of *U. scitaminea* strains from sugarcane smut whips

A total of 142 single-spore *U. scitaminea* isolates were isolated from 77 sugarcane smut whips (sori) collected in 15 different countries from America, Africa and Asia (Table 1). DNAs extracted from strains corresponding to mycelium culture derived from one generation of selfing of single spore isolates were analysed with 22 SSR primers pairs. Five SSR loci were not polymorphic, while the other 17 SSR loci amplified 53 distinct alleles with a mean of 3.1 alleles per locus (Fig. 1, Table 2). These 53 alleles are distributed into 22 distinct genotypes (Table 3).

Although DNA samples were extracted from dikaryotic mycelia, in most cases a single allele was amplified at each loci. Among the 2114 possible combinations (142 single-spore strains × 17 SSR), we only identified one strain, with genotype 42b, that displayed a heterozygous pattern at one locus (Usc8) (Table 3). These results indicate that *U. scitaminea* teliospores are homozygous.

For 45 sori, from seven of the 15 countries sampled, two to three single-spore strains per sorus were analysed. No polymorphism between single-spore strains from the same sorus (within sorus) was observed (Fig. 1) except in two cases. The first exception observed is two single-spore strains issued from sorus 42 that presented very different genotypes at 13 out of 17 loci (Table 3). Genotype 42a was homozygous and genotype 42b was homozygous at all loci except Usc8 (Table 3). This situation suggests that whip 42 was infected by two distinct *U. scitaminea* genotypes. The second exception is two single-spore strains from sorus 119 of Mauritius that presented two distinct genotypes (119a and 119b) differing at only one locus (Usc9) out of 17 (Table 3). This latter situation could result from a residual heterozygosity. With the exceptions of whips 42 and 119, the absence of variations between single-spore strains issued from a same sorus suggests that sugarcane is infected by a single genotype or that whips are colonized by a single genotype.

3.1. Global genetic structure of U. scitaminea populations

Further analyses were performed with one strain per homogenous sorus (75 strains) and the four strains described previously, resulting in a total of 79 strains. A neighbour joining (N-J) tree was constructed using pairwise

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

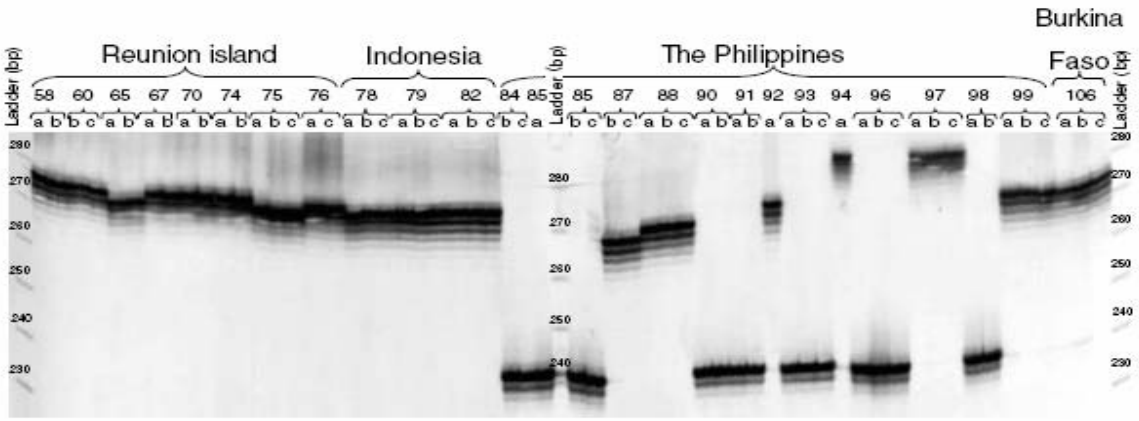


Fig. 1. Example of microsatellite profiles (Usc9). No polymorphism is detected among *U. scitaminea* dikaryotic progenies (a–c) from teliospores collected on the same sorus. A high polymorphism is observed among strains from The Philippines by comparison to samples from Reunion Island, Burkina Faso and Indonesia. One allelic variation (one-motif slippage) is observed among Reunion Island strains.

genetic distances to estimate genetic relatedness between genotypes (Fig. 2). Genetic diversity statistics were computed for countries where the sample size was large enough (Colombia, the Philippines, Reunion Island and Burkina Faso, Table 4). A contrasted situation was observed between *U. scitaminea* populations, with almost no genetic diversity in populations from America and Africa and high genetic diversity in populations from Asia.

In Africa and America (including Hawaii), 69% (33 out of 48) of the strains displayed exactly the same genotype (Worldwide Genotype:WG). The remaining 15 *U. scitaminea* strains corresponded to five different genotypes (labelled WL) that differed from WG by only one or two alleles out of 17 (Table 3 and Fig. 2). Eight of them, carried allele 3 at Usc9 instead of allele 4 in WG (Tables 3 and 5). Strains Mauritius 119 and Burkina Faso 104 carried allele 2 and 5 at Usc9, respectively (Tables 3 and 5). Strains 106, 113, 115, 105 from Burkina Faso carried allele 4 at Usc1 instead of allele 3 in WG (Tables 3 and 5). One last strain from Reunion (76) differed at two loci from WG, carrying allele 3 at Usc9 instead of allele 4 and carrying allele 1 at Usc21 instead of allele 2 (Tables 3 and 5). In total, five genotypes slightly differing from WG have been observed in Africa and America (Table 3, Fig. 2).

Asian strains from India, Indonesia, Thailand, Japan and the Philippines corresponded to 17 different genotypes including WG. Four strains from Indonesia and two strains from Thaïland displayed the WG. The two other Thai strains displayed new genotypes related to WG (Fig. 2) as shown for genotype 27 from Thailand that differed from WG at a single locus (Table 3). Strains from Japan, although collected in two locations, were identical and differed from all other genotypes (Table 3). Indian strain differed from all other genotypes (Table 3, Fig. 2). In the Philippines, the 16 strains analysed displayed 12 distinct new genotypes (Table 3, Fig. 2). Some variability was detected at some locations in the Philippines. For example, the six strains from 'Luy-a' (Nos. 042, 045, 096, 097, 098 and 099; Table 1) collected from 6 sori, displayed 6 distinct genotypes. Conversely, we also found similar strains at different localities in the Philippines such as strains 091, 093 and 096 collected from 'La Granja', 'Hacienda Cansilayan' and 'Luy-a' localities, respectively (Table 1).

3.2. Diversification from WG and relationships between WG and Asian populations

Seventeen out of the 22 alleles detected in *U. scitaminea* African and American populations, corresponding to all 17 alleles of WG, were also detected in Asian populations. WG was present in Indonesia and Thailand. Nevertheless, in the Philippines, where most diversity was detected, the WG was not observed although all the alleles of WG were detected individually. Five alleles were specific to *U. scitaminea* African and American populations (alleles 2, 3 and 5 of Usc9 ; allele 4 of Usc1 and allele 1 of Usc21; Table 5). These alleles differ from WG alleles at corresponding loci by an increment of plus or minus one or two di-nucleotide repeat (Table 5). These results suggest that these isolates have diverged recently from isolates corresponding to the main WG through the accumulation of mutations. Thus, we hypothesize that all the genotypes related to WG in Africa and America are genetic variants issued from WG and therefore belong to the same Worldwide Lineage (WL, Fig. 2). In Thaïland, two strains displayed WG and two other displayed new genotypes closely related to WG (Fig. 2, Table 3). Since all the alleles found in Thaïland strains are present in the Philippines, the diversity observed among Thaï strains may result from a limited gene flow between nearby countries where a high diversity exists (as observed in the Philippines).

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 3
Distinct genotypes observed in this study at 17 polymorphic loci

n°	Reference strain (Fig. 2)	n	Usc1	Usc3	Usc7	Usc8	Usc9	Usc10	Usc12	Usc13	Usc15	Usc18	Usc21	Usc29	Usc30	Usc32	Usc33	Usc34	Usc38	Present in
1 WL(WG)	3	39	3	2	3	1	4	1	1	4	1	2	2	1	2	2	1	3	2	−1
2 WL	119b	8	3	2	3	1	3	1	1	4	1	2	2	1	2	2	1	3	2	−2
3 WL	115	4	4	2	3	1	4	1	1	4	1	2	2	1	2	2	1	3	2	Burkina Faso
4 WL	119a	1	3	2	3	1	2	1	1	4	1	2	2	1	2	2	1	3	2	Mauritius
5 WL	104	1	3	2	3	1	5	1	1	4	1	2	2	1	2	2	1	3	2	Burkina Faso
6 WL	76	1	3	2	3	1	3	1	1	4	1	2	1	1	2	2	1	3	2	Reunion
7 WL	27	1	3	2	3	1	4	2	1	4	1	2	2	1	2	2	1	3	2	Thailand
8	17	5	3	1	2	2	4	1	2	4	2	2	4	1	3	1	2	3	2	Japan
9	93	3	1	1	2	2	9	2	2	3	2	3	4	1	3	1	2	2	1	Philippines
10	042a	3	3	1	3	2	9	1	1	3	2	3	4	1	2	1	2	3	2	Philippines
11	41	1	3	2	2	1	8	2	4	4	1	2	2	1	2	2	1	3	2	India
12	28	1	2	2	3	1	4	1	1	4	1	1	2	1	2	2	1	3	2	Thailand
13	90	1	1	2	2	2	9	2	2	3	2	2	2	1	3	2	2	3	1	Philippines
14	45	1	1	2	3	2	4	2	2	3	1	3	2	1	3	1	1	1	1	Philippines
15	042b	1	2	2	1	1 + 2	7	2	1	4	1	1	3	2	1	2	1	3	2	Philippines
16	88	1	2	2	1	2	6	2	4	4	1	1	2	1	2	2	1	3	2	Philippines
17	94	1	2	2	3	1	1	1	3	4	1	2	2	1	1	2	1	3	2	Philippines
18	97	1	2	2	3	1	1	2	3	1	2	2	3	2	1	2	2	3	2	Philippines
18	92	1	2	2	3	1	4	1	4	2	1	2	2	1	2	2	1	3	2	Philippines
20	99	1	2	2	3	2	4	2	1	1	1	1	3	2	2	2	1	1	2	Philippines
21	98	1	3	1	3	2	9	1	1	3	2	3	2	1	3	1	2	3	2	Philippines
22	87	1	3	2	1	1	7	1	4	2	2	2	2	2	2	2	2	1	2	Philippines

(1)Reunion, Mauritius, Senegal, Burkina Faso, South Africa, Thailand, Indonesia, Brazil, Hawaii, Colombia, Texas, Guadeloupe.
(2)Reunion, Colombia, Burkina Faso, Mauritius.
n, number of strains displaying this genotype.

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

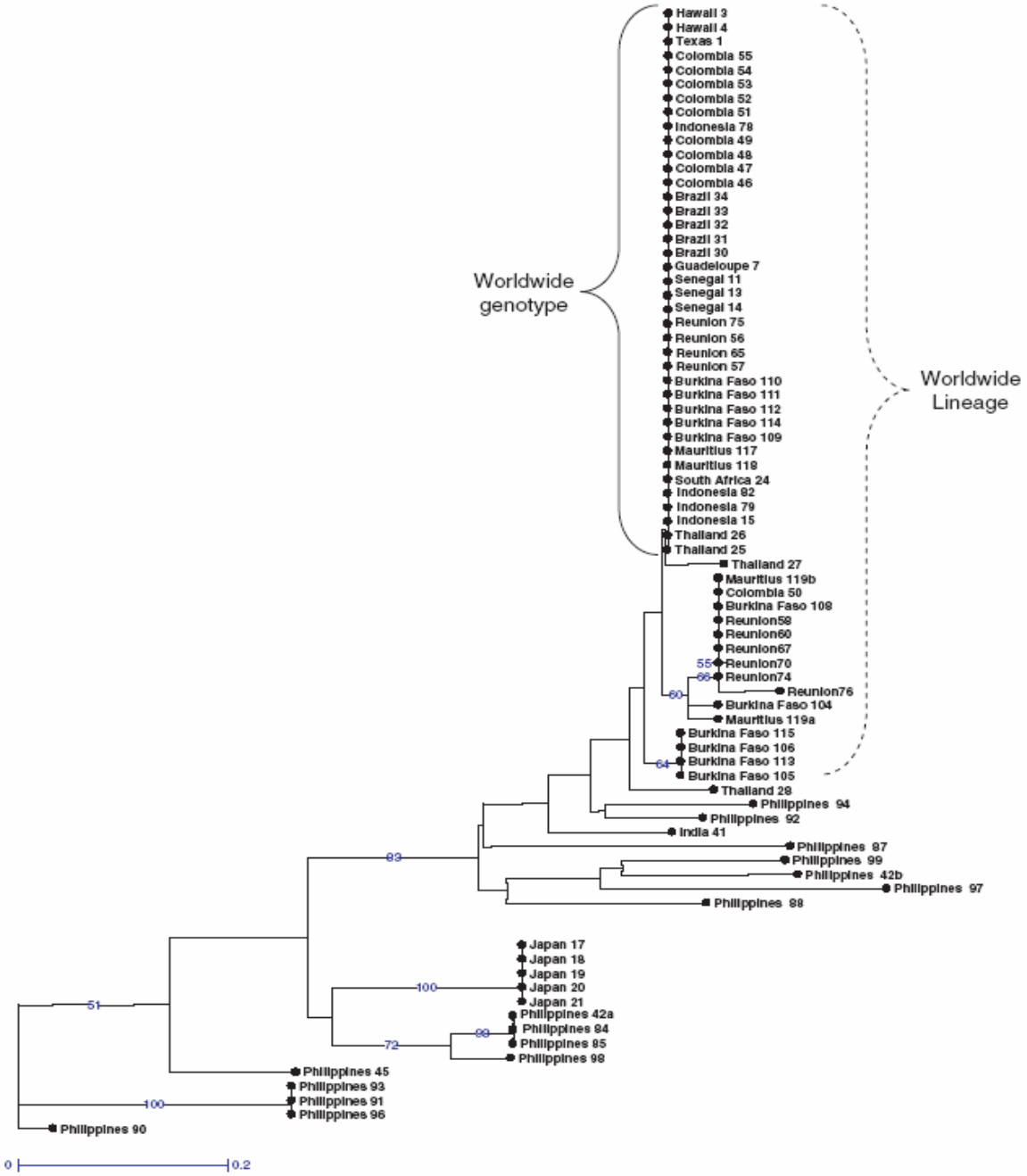


Fig. 2. Neighbour joining tree of the 79 *U. scitaminea* genotypes isolated from 77 distinct sori collected in 15 different countries. Single spore derived strains from a same sorus displaying distinct genotypes are identified by their sorus number followed by a letter.

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 4
Diversity statistics of *U. scitaminea* geographical populations

	Asia		America		Africa		
	All	Philippines	All	Colombia	All	Burkina	Reunion
N	31	16	19	10	29	11	10
Hn.b.	0.5055 (0.1178)	0.573 (0.1045)	0.0060 (0.0248)	0.0118 (0.0474)	0.0477 (0.1312)	0.0509 (0.1419)	0.0434 (0.1319)
Hobs	0.0020 (0.0081)	0.0039 (0.0156)	0.00 (0.00)	0.00 (0.00)	0.00 0.00	0.00 (0.00)	0.00 (0.00)
P(0.95)	1	1	0.0588	0.0625	0.1176	0.1250	0.1250
A	2.8235	2.8125	1.0588	1.0625	1.2941	1.1875	1.1250

Hn.b., unbiased gene diversity (Nei, 1978).
Hobs, observed heterozygosity.
P(0.95), proportion of polymorphic loci when most frequent allele does not exceed 95%.
A, mean number of alleles per locus.
Standard deviations are indicated in parenthesis.

4. Discussion

4.1. Selfing is the predominant reproduction mode in U. scitaminea

Since *U. scitaminea* strains were dikaryotic mycelium cultures originating from single diploid teliospores it was possible—using the co-dominant microsatellites markers developed in this study—to infer the reproduction mode of this fungus. All strains but one analysed in this study were homozygous at all 17 polymorphic microsatellite loci (Table 3). Moreover, almost no polymorphism was detected between single-teliospore strains from a same sorus, even in areas where polymorphism does exist such as the Philippines. Heterozygous teliospores from the same sorus would generate polymorphic and heterozygous dikaryotic genotypes (single-spore strains) as a result of recombination and association between haploid nuclei with different genotypes. Conversely (our situation), homozygous teliospores from a same whip should result in monomorphic and homozygous dikaryotic genotypes. It is also noteworthy that very different homozygous genotypes could be present at the same location, even in the same infected plant, as observed in the Philippines. Since we analysed DNA extracted from mycelium cultures derived from the selfing of single spore isolates, some heterozygosity may have been lost as a consequence of distorted segregations and drift during cultures. However, the original heterozygosity in spores would still be extremely low since the loss of heterozygosity at all loci and from all strains sampled from the same sori is very unlikely. Therefore, our results suggest that *Ustilago scitaminea* reproduces mainly by selfing and that populations of this fungus are mainly composed of inbred lineages.

This situation has already been observed in other Ustilaginales such as *Ustilago hordei* in which outcrossing is uncommon (Abdennadher and Mills, 2000), and *Microbotryum violaceum* (formerly *Ustilago violaceum*) that displayed unexpectedly high frequencies of homozygotes indicating that selfing is the predominant mode of reproduction (Bucheli et al., 2001; Delmotte et al., 1999; Giraud, 2004; Kaltz and Shykoff, 1999). *M. violaceum* predominent selfing results from its propensity to intra-tetrad mating (Baird and Garber, 1979; Giraud et al., 2005; Hood and Antonovics, 2000). The lack of available mating partners may also play a role. Indeed, this fungus is transmitted by insect pollinators which prefer healthy over diseased plants, thus reducing the probability of spore mixtures being deposited on the same flower (Shykoff and Bucheli, 1995) and spore dispersal is generally limited to nearby plants (Giraud, 2004). Therefore, only one genotype may arrive at a time on a given flower making selfing the only option (Giraud et al., 2005).

Ustilago scitaminea may have the potential for intratetrad mating since Leu (1972) and Bock (1964) observed the development of dikaryotic infection hyphae directly from the promycelial cells of a single teliospore. However, we still have to determine whether the high level of selfing observed in this study results from a selfing preference or from a lack of mating opportunities in outcrossing combinations. Indeed, it may be very unlikely that two teliospores from distinct genetic sources germinate simultaneously and close enough on the same cane bud to allow outcrossing. The possibility and frequency of outcrossing should be clarified through the genetic analysis of population structures at the field and plant level in areas where *U. scitaminea* genetic diversity is high or through experimental inoculations of sugarcane with mixtures of spores from distinct genotypes.

4.2. Global genetic structure of Ustilago scitaminea populations

In spite of our weak and unbalanced sampling, our results very clearly reveal the main characteristics of the genetic structures of *U. Scitaminea* global population. The highest genetic diversity of *U. scitaminea* populations was observed in Asia, mainly in the Philippines (Fig. 2) whereas outside Asia, the genetic diversity of *U. scitaminea* populations was extremely low. Indeed, *U. scitaminea* strains from America and Africa are closely related genetically and are likely belonging to a single worldwide lineage. This situation is reminiscent of the preliminary results obtained by Braithwaite et al. (2004), in a worldwide AFLP-based study of *U. scitaminea* populations. These authors have observed, although with a very low number of strains, that *U. scitaminea* populations from South Africa, Zimbabwe, Mauritius, United States of America (Louisiana, Florida, Texas, Hawaii), Argentina, Brazil, Columbia and Venezuela displayed a very low level of genetic diversity. In contrast, they have observed a higher level of polymorphisms among *U. scitaminea* Asian strains (Taiwan, Thailand, Philippines). RAPD markers also revealed the lack of polymorphism between strains from South Africa, Louisiana, Hawaii and Reunion Island (Singh et al., 2005). In our study, the single worldwide lineage was also detected in some Asian countries such as Indonesia and Thailand. Furthermore, the

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Table 5
Allele frequencies for 17 polymorphic microsatellite loci across *U. scitaminea* populations

Microsatellite locus	Estimated allele size	Allele code	WG[a]	Asia		America		Africa		
				All	Philippines	All	Colombia	All	Burkina	Reunion
Usc1	187	1		0.17	0.31	0	0	0	0	0
	181	2		0.23	0.38	0	0	0	0	0
	177	3	WG	0.6	0.31	1	1	0.86	0.64	1
	175	4	WL	0	0	0	0	0.14	0.36	0
Usc3	310	1		0.4	0.44	0	0	0	0	0
	308	2	WG	0.6	0.56	1	1	1	1	1
Usc7	280	1		0.1	0.19	0	0	0	0	0
	282	2		0.33	0.25	0	0	0	0	0
	286	3	WG	0.57	0.56	1	1	1	1	1
Usc8	200	1	WG	0.45	0.29	1	1	1	1	1
	158	2		0.55	0.71	0	0	0	0	0
Usc9	288	1		0.07	0.13	0	0	0	0	0
	280	2	WL	0	0	0	0	0.03	0	0
	278	3	WL	0	0	0.05	0.1	0.28	0.09	0.6
	276	4	WG	0.53	0.19	0.95	0.9	0.66	0.82	0.4
	274	5	WL	0	0	0	0	0.03	0.09	0
	272	6		0.03	0.06	0	0	0	0	0
	270	7		0.07	0.13	0	0	0	0	0
	252	8		0.03	0	0	0	0	0	0
	242	9		0.27	0.5	0	0	0	0	0
Usc10	163	1	WG	0.63	0.44	1	1	1	1	1
	165	2	WL	0.37	0.56	0	0	0	0	0
Usc12	205	1	WG	0.47	0.38	1	1	1	1	1
	203	2		0.33	0.31	0	0	0	0	0
	193	3		0.07	0.13	0	0	0	0	0
	189	4		0.13	0.19	0	0	0	0	0
Usc13	225	1		0.07	0.13	0	0	0	0	0
	175	2		0.07	0.13	0	0	0	0	0
	171	3		0.3	0.56	0	0	0	0	0
	169	4	WG	0.57	0.19	1	1	1	1	1
Usc15	122	1	WG	0.5	0.38	1	1	1	1	1
	118	2		0.5	0.63	0	0	0	0	0
Usc18	278	1		0.13	0.19	0	0	0	0	0
	276	2	WG	0.6	0.31	1	1	1	1	1
	272	3		0.27	0.5	0	0	0	0	0
Usc21	150	1	WL	0	0	0	0	0.03	0	0.1
	148	2	WG	0.53	0.44	1	1	0.97	1	0.9
	144	3		0.1	0.19	0	0	0	0	0
	142	4		0.37	0.38	0	0	0	0	0
Usc29	202	1	WG	0.87	0.75	1	1	1	1	1
	192	2		0.13	0.25	0	0	0	0	0
Usc30	128	1		0.1	0.19	0	0	0	0	0
	112	2	WG	0.53	0.44	1	1	1	1	1
	110	3		0.37	0.38	0	0	0	0	0
Usc32	269	1		0.43	0.5	0	0	0	0	0
	267	2	WG	0.57	0.5	1	1	1	1	1
Usc33	122	1	WG	0.5	0.38	1	1	1	1	1
	118	2		0.5	0.63	0	0	0	0	0
Usc34	292	1		0.1	0.19	0	0	0	0	0
	284	2		0.1	0.19	0	0	0	0	0
	278	3	WG	0.8	0.63	1	1	1	1	1
Usc38	277	1		0.17	0.31	0	0	0	0	0
	273	2	WG	0.83	0.69	1	1	1	1	1

[a] WG indicates alleles of the worldwide genotype and WL alleles of the worldwide lineage.

microsatellites markers used in the present study showed that this worldwide lineage is a combination of alleles mostly detected in strains from Asia. These results strongly suggest that *U. scitaminea* WL isolates have originated from a single isolate/genotype (WG) from Asia that has spread the disease around the world.

Migration of *U. scitaminea* WG isolate from Asia to other continents has likely occurred through exchange of infected stem cuttings of sugarcane cultivars. Indeed, historical records of sugarcane cultivation show the occurrence of successive cultivar replacements on a worldwide scale. For example, sugarcane hybrids were first produced

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

in Java (Indonesia) and Coimbatore (India) and distributed worldwide during the 1920s and 1930s. Thus, the movement of infected material from a unique origin is the likely explanation for the strong founder effect observed in *U. scitaminea* African and American populations. Since significant genetic diversity of *U. scitaminea* populations was only detected in Asia, a thorough investigation on its population genetic structure should be conducted in this continent using hierarchical sampling methods, in order to determine precisely the centre of diversity of *U. scitaminea*, to estimate the degree of gene flow between Asian populations and to evaluate the relative importance of movement of infected material and wind dispersal of teliospores in the spread of the disease at different geographical scales.

The global genetic structure of *U. Scitaminea* populations observed in this study could be compared to the case of *Phytophthora infestans* where a single genotype was distributed panglobally resulting probably from very low level of gene flow (through a recent spread from a presumed ancestral population in Mexico) and recombination (Goodwin et al., 1994). These cases appear extreme in comparison to the few other reports available on other plant pathogenic fungus (Carlier et al., 1996; Hayden et al., 2003; Zhan et al., 2002). For example, the wheat pathogen *Mycosphaerella graminicola*, display a high level of genetic diversity within populations (at field and plant scales) and a low level of genetic differentiation between populations from different continents. The populations of this fungal pathogen are panmictic and its spread around the world is ancient. Thus, the population structure observed might result from gene flow with regular exchange of infected plant material during time and recombination. The populations of the banana pathogens *M. fijiensis* and *M. musicola* from different continents are also panmictic and maintained high level of genetic diversity within populations at field and plants scales. However, a higher level of genetic diversity was observed in the Asia and a high level of genetic differentiation between populations from different or the same continents was detected. The recent spread of these pathogens around the world through occasional movement of infected plant material and recombination might be the main factors that shaped such a structure. The extreme global genetic structure of *U. scitaminea* populations might result from the combination of several factors: a very low level of genetic diversity at the plant level, a low level of recombination and a very limited gene flow with a recent spread around the world of the pathogen through occasional movements of infected plant material.

4.3. Implications for studies on pathogenic variation and resistance

Race-cultivar interactions controlled by a gene-for-gene interaction were demonstrated in *Ustilago hordei* (Linning et al., 2004). In *Ustilago scitaminea*, race-cultivar interactions were observed through controlled artificial inoculations in Taiwan (Leu and Teng, 1972; Leu, 1978; Lee et al., 1999). Interestingly, the pathogenicity variations observed in Taiwan coincide with the genetic variation revealed by AFLP markers in these populations (Braithwaite et al., 2004). Other reports suggested race-cultivar interactions in Pakistan, Brazil and Hawaii (Muhammad and Kausar, 1962; Toffano, 1976; Comstock and Heinz, 1977). Nonetheless, an international project involving Argentina, Barbados, Colombia, India, South Africa, Sudan, Taiwan, USA (including Hawaii) and Zimbabwe could not reveal such race-cultivar interactions except in Taiwan (Grisham, 2001). Similarly, Péros and Baudin (1983) did not observe differences in pathogenicity nor race-cultivar interactions among isolates from Reunion Island, Ivory Coast, Mali, West Indies, Kenya and Burkina Faso. These observations are likely the result of the absence of genetic diversity in these populations, as suggested by our study. Thus, we hypothesize that pathogenicity variability is greater in Asia where a high level of genetic variation in *U. scitaminea* does exist. To further test this hypothesis, pathogenicity should be analysed using appropriate sampling and tested under controlled conditions on a reference set of sugarcane cultivars.

The results obtained in this study suggest that the use of resistant cultivars to *U. scitaminea* might be an efficient and durable strategy to control sugarcane smut outside Asia. Gene flow between continents appeared accidental through movement of infected materials. Quarantine measures applied today should greatly reduce the risk of migration of new pathotypes from Asia to other sugarcane-producing areas (Frison and Putter, 1993). Furthermore, *U. scitaminea* appeared to reproduce mainly by selfing limiting evolution through recombination. However, since variability was detected in Asia, disease control with resistant cultivars might be more difficult in this region. To develop further disease control using resistant cultivars, a better characterisation of resistant germplasm is needed to identify resistance genes within modern sugarcane cultivar germplasm. Resistance should also be searched for in wild *Saccharum spontaneum* since modern sugarcane cultivars are derived from interspecific hybrids between the domesticated sugar producing species *S. officinarum* and *S. spontaneum* (Daniels and Roach, 1987) and numerous resistant clones have been found in *S. spontaneum* (Srinivasan and Alexander, 1971). This situation could be the result of co-evolution between smut and *Saccharum spontaneum* that has a large distribution from West Africa to Japan, with Asia being the most probable center of diversity of this species (Panje and Babu, 1960). By contrast, smut disease has never been reported in Papua New Guinea, the center of origin of *S. officinarum* (Antoine, 1961).

Acknowledgments

We gratefully acknowledge Genoscope (Centre National de Séquençage, Evry, France) for sequencing the sugarcane smut SSR clones and all the institutions listed

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

in Table 2 that kindly provided us with smut samples. We also thank Georges Piperidis, Laurent Costet, Fabien Halkett, Didier Tharreau and Jean Pierre Péros for critical reading of the manuscript. We also gratefully thank Marc-Henri Lebrun for his helpful revision of the manuscript.

References

Abdennadher, M., Mills, D., 2000. Telomere-associated RFLPs and electrophoretic karyotyping reveal lineage relationships among race-specific strains of *Ustilago hordei*. Curr. Genet. 38, 141–147.

Alexander, K.C., Srinivasan, K.V., 1966. Sexuality in *Ustilago scitaminea* Syd. Curr. Sci. 23, 603–604.

Amire, O.A., Trione, E.J., Schmitt, R.A., 1982. Characterization of pathogenic races of the sugarcane smut fungus by neutron activation analysis. J. Radioanal. Chem. 75, 195–203.

Antoine, R., 1961. Smut. In: Martin, J.P., Abbott, E.V., Hughes, C.G. (Eds.), Sugar Cane Disease of the World. Elsevier Publishing, Amsterdam, pp. 327–345.

Baird, M.L., Garber, E.D., 1979. Genetics of *Ustilago violacea* V. Outcrossing and selfing in teliospore inocula. Bot. Gaz. 140, 89–93.

Barat, H., 1971. Etat Phytosanitaire des cannes à sucre de Côte d'Ivoire. Dangers résultant de la présence du charbon en Haute-Volta. Rapport IRAT.

Bates, J.F., 1975. Smut disease in Guiana. Sugarcane Pathologists' Newslett. 13/14, 8.

Bakkeren, G., Kronstad, J.W., 1994. Linkage of mating type loci distinguishes bipolar from tetrapolar mating in Basidiomycetous smut fungi. Proc. Natl. Acad. Sci. USA 91, 7085–7089.

Billotte, N., Lagoda, P.J.L., Risterucci, A.M., Baurens, F.C., 1999. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54 (4), 277–288.

Belkhir, K., Borsa, P., Chiki, L., Raufaste, N., Bonhomme, F., 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique despopulations. Laboratoire Génome, Populations, Interactions. CNRS UMR5171, Université de Montpellier II, Montpellier (France). Available from: <http://www.genetix.univ-montp2.fr/genetix/intro.htm>.

Bock, K.R., 1964. Studies on sugar-cane smut (*Ustilago scitaminea*) in Kenya. Trans. Brit. Mycol. Soc. 47 (3), 403–417.

Braithwaite, K.S., Bakkeren, G., Croft, B.J., Brumbley, S.M., 2004. Genetic variation in a worldwide collection of the sugarcane smut fungus *Ustilago scitaminea*. Proc. Aust. Soc. Sugarcane Technol. 26, abstract p 48.

Bucheli, E., Gautschi, B., Shykoff, J.A., 2001. Differences in population structure of the anther smut fungus *Microbotryum violaceum* on two closely related host species. Mol. Ecol. 10, 285–294.

Carlier, J., Lebrun, M.H., Zapater, M.F., Dubois, C., Mourichon, X., 1996. Genetic structure of the global population of banana black leaf streak fungus, *Mycosphaerella fijiensis*. Mol. Ecol. 5, 499–510.

Chakraborty, R., Jin, L., 1993. A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and populations distances. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T., Jeffreys, A.J. (Eds.), DNA Fingerprinting: State of the Science. Birkhaüser Verlag, Basel, pp. 153–175.

Chao, C.P., Hoy, J.W., Saxton, A.M., Martin, F.A., 1990. Heritability of resistance and repeatability of clone reactions to sugarcane smut in Louisiana. Phytopathology 80, 622–626.

Comstock, J.C., 2000. Smut. In: Rott, P., Bailey, R.A., Comstock, J.C., Croft, B.J., Salem Saumtally, A. (Eds.), A Guide to Sugarcane Diseases. CIRAD and ISSCT, pp. 181–185.

Comstock, J.C., Heinz, D.J., 1977. A new race of culmiculous smut of sugarcane in Hawaii. Sugarcane Pathologists' Newslett. 19, 24–25.

Daniels, J., Roach, B.T., 1987. Taxonomy and evolution. In: Heinz, D.J. (Ed.), Sugarcane Improvement through Breeding. Elsevier, Amsterdam, pp. 7–84.

Delassus, M., 1975. Rapport sur le charbon de la canne à sucre. Situation au Niger et méthodes de lutte à envisager. Rapport de l'IRAT.

Delmotte, F., Bucheli, E., Shykoff, J.A., 1999. Host and parasite population structure in a natural plant-pathogen system. Heredity 82, 300–308.

Ferreira, S.A., Comstock, J.C., 1989. Smut. In: Ricaud, C., Egan, B.T., Gillaspie, A.G., Hughes, C.G. (Eds.), Diseases of Sugarcane. Elsevier, Amsterdam, pp. 211–229.

Frison, E.A., Putter, C.A.J., 1993. FAO/IBPGR Technical guidelines for the safe movement of sugarcane germplasm. Rome, Italy, Food and Agriculture Organization of the United Nations and International Board for Plant Genetic Resources, 44p.

Goodwin, S.B., Barak, A.C., William, E.F., 1994. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. Proc. Natl. Acad. Sci. USA 91, 11991–11995.

Giraud, T., 2004. Patterns of within population dispersal and mating of the fungus *Microbotryum violaceum* parasiting the plant *Silene latifolia*. Heredity 93, 559–565.

Giraud, T., Jonot, O., Shykoff, J.A., 2005. Selfing propensity under choice conditions in a parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection success in artificial conditions. Int. J. Plant Sci. 166, 649–657.

Grisham, M.P., 2001. An international project on sugarcane variability within sugarcane smut. Proc. Int. Soc. Sugarcane Technol. 24 (2), 459–461.

Hayden, H.L., Carlier, J., Aitken, E.A.B., 2003. Population differentiation in the banana leaf spot pathogen *Mycosphaerella musicola*, examined at a global scale. Plant Pathol., 713–719.

Hirshhorn, E., 1949. Un nuevo metodo de infeccion artificial con el carbon de la cana de azucar. Revista de Investigaciones Agricolas III (4), 335–344.

Hood, M.E., Antonovics, J., 2000. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (= *Ustilago violacea*). Heredity 85, 231–241.

James, N.I., 1978. Sugarcane smut–an eminent threat to the U.S. mainland. Proc. Amer. Soc. Sugarcane Technol. 7, 63–66.

Kaltz, O., Shykoff, J.A., 1999. Selfing versus outcrossing propensity of the fungal pathogen *Microbotryum violaceum* across *Silene latifolia* host plants. J. Evol. Biol. 12, 340–349.

Lambat, A.K., Chenulu, V.V., Chona, B.L., 1968. Morphological variation in the sugarcane smut *Ustilago scitaminea* Syd. Mycopathologia Mycologia Applicata 36, 301–304.

Lee, C.S., Yuan, C.H., Liang, Y.G., 1999. Occurrence of a new pathogenic race of culmiculous smut of sugarcane in Taiwan. Proc. Int. Soc. Sugarcane Technol. 23 (2), 406–407.

Lee-Lovick, G., 1978. Smut of sugarcane—*Ustilago scitaminea*. Rev. Plant Pathol. Commonwealth Mycol. Inst. 57 (5), 181–187.

Leu, L.S., 1972. Culmiculous smut of sugar cane in Taiwan III. Germination and storage of teliospores, and compatibility of *Ustilago scitaminea* Sydow. Report Taïwan Sugarcane Experimental Station 56, 37–45.

Leu, L.S., 1978. Culmicolous smut of sugar cane in Taiwan VI. New pathogenic strain obtained by artificial hybridization and further studies on compatibility of *Ustilago scitaminea* Sydow. Ann. Phytopath. Soc. Japan 44, 321–324.

Leu, L.S., Teng, W.S., 1972. Pathogenic Strains of *Ustilago scitaminea* Sydow. Sugarcane Pathologists' Newslett. 8, 12–13.

Leung, H., Nelson, R.J., Leach, J.E., 1993. Population structure of plant pathogenic fungi and bacteria. Adv. Plant Pathol. 10, 157–205.

Linning, R., Lin, D., Lee, N., Abdennadher, M., Gaudet, D., Thomas, P., Mills, D., Kronstad, J.W., Bakkeren, G., 2004. Marker-based cloning of the region containing the UhAvr1 avirulence gene from the basidiomycete barley pathogen *Ustilago hordei*. Genetics 166, 99–111.

McDonald, B.A., Linde, C., 2002. Pathogen population genetics, evolutionary potential , and durable resistance. Annu. Rev. Phytopathol. 40, 349–379.

Martinez-Espinoza, A.D., Garcia-Pedrajas, M.D., Gold, S.E., 2002. The Ustilaginales as plant pests and model system. Fung. Genet. Biol. 35, 1–20.

Muhammad, S., Kausar, A.G., 1962. Preliminary studies on the genetics of sugarcane smut, *Ustilago scitaminea* Sydow. Biologia 8, 65–74.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.

Panje, R.R., Babu, C.N., 1960. Studies in *Saccharum spontaneum*. Distribution and geographical association of chromosome numbers. Cytologia 25, 152–172.

Péros, J.P., Baudin, P., 1983. Etude de la variabilité d'*Ustilago scitaminea* Syd., agent du charbon de la canne à sucre. L'Agronomie Tropicale 38 (3), 234–239.

Perrier, X., Flori, A., Bonnot, F., 2003. Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X., Glaszmann, J.C. (Eds.), Genetic Diversity of Cultivated Tropical Plants. Enfield, Science Publishers, Montpellier, pp. 43–76.

Riley, I.T., Jubb, T.F., Egan, B.T., Croft, B.J., 1999. First outbreak of sugarcane smut in Australia. Proc. Int. Soc. Sugarcane Technol. 23 (2), 333–336.

Rogers, S., Bendich, A., 1988. Extraction of DNA from plant tissues. In: Gelvin, S.B., Schilperoort, R.A. (Eds.), Plant Molecular Biology Manual A6. Kluwer Academic Publishers, Dordrecht, pp. 1–10.

Singh, N., Somai, B.M., Pillay, D., 2005. Molecular profiling demonstrates limited diversity amongst geographically separate strains of *Ustilago scitaminea*. FEMS Microbiol. Lett. 247 (1), 7–15.

Shykoff, J., Bucheli, E., 1995. Pollinator visitation patterns, floral rewards and the probability of transmission of *Microbotryum violaceum*, a venereal disease of plants. J. Ecol. 83, 189–198.

Srinivasan, K.V., Alexander, K.C., 1971. Sources of resistance to red rot and smut in the species of *Saccharum*. Sugarcane Pathologist's Newslett. 6, 6–7.

Toffano, W.B., 1976. Estudos sobre polaridade sexual e possiveis raças fisiologicas de *Ustilago scitaminea* Syd., no estado de Sao Paulo. Arq. Inst. Biol., Sao Paulo 43, 65–79.

Trione, E.J., 1990. Growth and sporulation of *Ustilago scitaminea*, *in vivo* and *in vitro*. Mycol. Res. 94 (4), 489–493.

Veiga, F.M., 1972. Smut in Brazil. Sugarcane Pathologists' Newslett. 9 (25), 17.

Wu, K.K., Heinz, D.J., Hogarth, D.M., 1988. Association and heritability of sugarcane smut resistance to races A and B in Hawaii. Theor. Appl. Genet. 75, 754–760.

Zhan, J., Pettway, R.E., McDonald, B.A., 2002. The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. Fungal Genet. Biol. 38, 286–297.

Please cite this article as: Louis-Marie Raboin et al., Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*, Fungal Genetics and Biology (2006), doi:10.1016/j.fgb.2006.07.004.