Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Mapeamento genético de cana-de-açúcar (*Saccharum* spp.) por associação empregando marcadores SSR e AFLP

Francisco Claudio da Conceição Lopes

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2011

Francisco Claudio da Conceição Lopes
Engenheiro Agrônomo

Mapeamento genético de cana-de-açúcar (*Saccharum* spp.) por associação empregando marcadores SSR e AFLP

Orientador:
Prof. Dr. **ANTONIO VARGAS DE OLIVEIRA FIGUEIRA**

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2011

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP

Lopes, Francisco Claudio da Conceição
Mapeamento genético de cana-de-açúcar (*Saccharum* spp.) por associação empregando marcadores SSR e AFLP / Francisco Claudio da Conceição Lopes. - - Piracicaba, 2011.
140 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2011.

1.Cana-de-açúcar 2. Cromossomos vegetais 3. Genética estatística 4. Marcador molecular 5. Mapeamento genético 6. Melhoramento genético vegetal I.Título

CDD 633.61
L864m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Aos amados de minha vida

*À minha mãe **Zilda**, pelo Amor puro, inocente e incondicional dedicado a mim,*

*Ao meu pai **Manoel**, por sempre ter me apoiado e me dado o fundamental: liberdade,*

*Às minhas irmãs, **Claudete** e **Claudiane,** e sobrinhos, **Vinícius**, **Rian**, **Alejandro** e à recém-chegada **Júlia**, os quais tenho aprendido a amar cada vez mais,*

*À **Dudinha**, pequena jóia, que desperta o que há de mais paternal em mim,*

DEDICO

*À **Linéia**, minha esposa, pelo companheirismo e pela busca que só faz sentido a dois...*

OFEREÇO

AGRADECIMENTOS

Gostaria de agradecer profundamente a todos aqueles que contribuíram para que este trabalho fosse realizado.

À Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/USP), ao Centro de Energia Nuclear na Agricultura (CENA/USP) e às instituições CAPES e CNPq pelas bolsas de estudo e apoio financeiro concedidos. De forma mais específica, agradeço ao Departamento de Genética da ESALQ, por ter proporcionado pela segunda vez o ambiente de aperfeiçoamento profissional.

Meus agradecimentos também ao Centro de Cana do Instituto Agronômico de Campinas (IAC) pela cooperação e apoio dados ao projeto através dos pesquisadores Dr. Marcos Guimarães de Andrade Landell, Dra. Silvana Aparecida Creste Dias de Souza, Dra. Luciana Rossini Pinto e Dr. Márcio Bidóia. Agradeço também a todos os colegas do Laboratório de Biologia Molecular da mesma instituição, em especial aos técnicos de laboratório Luiz e Thaís.

Ao prof. Dr. Antonio Vargas de Oliveira Figueira por ter aceitado de imediato meu pedido para trabalhar em seu laboratório.

Ao prof. Dr. Alexandre Coelho Guedes Siqueira da Universidade Federal de Goiás pela inestimável ajuda com as análises de associação.

A todos os amigos do Laboratório de Melhoramento de Plantas do CENA/USP: Aline Harissis, Amanda, Ângela, André, Bressan, Dani, Déborah, Felipe Jóia, Gabi, Ilse, Isabela, Joni, Juliana Deganello, Juliana Leles, Letícia, Lorena, Mariana Belloti, Marielle, Nebó, Renatão, Roberto, Thaís, Thaísa e Verusca. Aos amigos que já deixaram o laboratório, Celsão (pode crer...) e Onildo (ó pai ó), com quem tive bons momentos de descontração e ao Jupará pelos cafés filosóficos. Obrigado a todos pela convivência amigável e interessante.

Aos técnicos Eduardo Fonseca, Raquel Orsi e Wlamir Godoy por todo o apoio às atividades de laboratório nestes quatro anos.

Aos meus queridos irmãos de coração Thiagão e Edi pelos almoços e jantares (valeu Edi!) e pelas boas horas de lazer compartilhadas.

"Emancipate yourselves from mental slavery
None but ourselves can free our minds…"
(Redemption Song)
Bob Marley

"Don't let them fool you
Or even try to school you, Oh! No
We've got a mind of our own
So go to hell if what you're thinkin' isn't right
Love would never leave us alone
In the darkness there must come out to light
Could you be loved and be loved?
Could you be loved and be loved…"
(*Could you be loved)*
Bob Marley

"Ouse, ouse... ouse tudo!!
Não tenha necessidade de nada!
Não tente adequar sua vida a modelos,
nem queira você mesmo ser um modelo para ninguém.
Acredite: a vida lhe dará poucos presentes.
Se você quer uma vida, aprenda... a roubá-la!
Ouse, ouse tudo! Seja na vida o que você é, aconteça o que acontecer.
Não defenda nenhum princípio, mas algo de bem mais maravilhoso:
algo que está em nós e que queima como o fogo da vida!"

Lou Andreas-Salomé

SUMÁRIO

RESUMO 11
ABSTRACT 13
LISTA DE FIGURAS 15
LISTA DE TABELAS 17
1 INTRODUÇÃO 19
1.1 Aspectos econômicos do agronegócio sucroalcooleiro no Brasil 19
1.2 Taxonomia e origem da cana-de-açúcar 20
1.3 Melhoramento e uso de germoplasma no desenvolvimento das cultivares modernas 22
1.4 Marcadores moleculares em cana-de-açúcar 26
1.5 Mapeamento de QTLs 29
1.5.1 Mapeamento via análise de ligação 29
1.5.2 Mapeamento por associação 31
1.5.3 Estudos de desequilíbrio de ligação e Mapeamento por associação em cana..... 35
1.6 Modelos mistos em estudos de associação 38
2 MATERIAL E MÉTODOS 41
2.1 Material vegetal 41
2.2 Caracteres fenotípicos 41
2.3 Extração de DNA 47
2.4 Dados moleculares 47
2.4.1 Marcadores SSR 47
2.4.2 Marcadores AFLP 53
2.4.2.1 Restrição do DNA 53
2.4.2.2 Ligação dos adaptadores 53
2.4.2.3 Reação de pré-amplificação 53
2.4.2.4 Reação de amplificação seletiva 54
2.4.2.5 Análise dos fragmentos AFLP 55
2.5 Análise dos dados 56
2.5.1 Análises de diversidade 56
2.5.1.1 Polimorfismo e Diversidade alélica 56
2.5.1.2 Análise de similaridade genética 57
2.5.2 Desequilíbrio de ligação 58
2.5.3 Análise de estruturação da população 59
2.5.4 Coeficientes de coancestralidade 60
2.5.5 Análise dos dados fenotípicos 61
2.5.6 Análise de associação 62
3 RESULTADOS E DISCUSSÃO 65
3.1 Seleção dos genótipos para compor a população de mapeamento 65
3.2 Análises de diversidade 65
3.3 Análise de similaridade genética 70

3.4 Níveis e estrutura do desequilíbrio de ligação no genoma de cana 78
3.5 Análise de estruturação pelo programa STRUCTURE 85
3.6 Medidas dos coeficientes de coancestralidade 89
3.7 Análises de associação 90
4 CONCLUSÕES 97
REFERÊNCIAS 99
ANEXOS 117

RESUMO

A cultura da cana-de-açúcar (*Saccharum* spp.) possui uma importância histórica e econômica para o Brasil. O agronegócio sucroalcooleiro vem experimentando forte expansão na última década não só no Brasil como também em todo o mundo em função, principalmente, da demanda por fontes de energia menos agressivas ao ambiente. Para atender a uma maior demanda por seus subprodutos, principalmente de etanol, a área cultivada com cana-de-açúcar vem aumentando a cada ano no Brasil, ocupando áreas novas de cultivo nas regiões centrais do país. Nesse contexto, o melhoramento genético tem um papel fundamental no desenvolvimento de novas cultivares adaptadas a essas condições de cultivo. A maioria dos caracteres de importância econômica possui uma natureza genética complexa fazendo com que o desenvolvimento de uma nova cultivar de cana-de-açúcar leve mais de 10 anos. Desta forma, o uso de abordagens que permitam a identificação de genes ou de QTLs associados a caracteres quantitativos de forma precisa e rápida tem grande utilidade no melhoramento dessa espécie. O mapeamento por associação baseado no fenômeno do desequilíbrio de ligação é uma metodologia que visa detectar associações entre genes e caracteres agronômicos, podendo contribuir desta forma para o melhoramento da cana-de-açúcar. Assim, este trabalho teve como principal objetivo avaliar o uso da abordagem de mapeamento por associação na detecção de associações importantes entre marcadores moleculares do tipo SSR e AFLP e os caracteres Altura, Diâmetro e Número de Colmos; Percentual de Fibra na Cana (Fibra % Cana); Porcentagem em massa de sacarose (Pol % cana) e Tonelada de Cana por Hectare (TCH) em cana-de-açúcar. Os dados fenotípicos dos genótipos avaliados são oriundos de experimentos conduzidos em quatro regiões: Ribeirão Preto, Jaú e Piracicaba em São Paulo e Goianésia em Goiás no período entre 1990 e 2009. Associações entre os marcadores e os caracteres fenotípicos foram avaliadas em cada região e em todas simultaneamente. A análise de associação realizada através de modelos mistos sugeriu a existência de doze associações envolvendo os caracteres Número de Colmos, Porcentagem em massa de sacarose (Pol % cana) e Tonelada de Cana por Hectare (TCH). Quatro associações envolveram a característica Número de Colmos sendo três (*CIR56*, *ACG_CGT* e *AGG_CAG*) de caráter geral, ou seja, relacionada à média das quatro regiões e uma associação na região de Goiás (*ACG_CAT*). Sete associações entre a característica Pol % Cana e as marcas *CV60*, *CV106*, *AAG_CAC*, *AAG_CAG* e *ACG_CTT* foram específicas para a região de cultivo de

Ribeirão Preto. Uma associação entre Tonelada de Cana por Hectare (TCH) e a marca *ACG_CGC* foi detectada na região de Piracicaba.

Palavras-chave: Poliplóide; Cana-de-açúcar; Desequilíbrio de ligação, Mapeamento por Associação; Modelos mistos; QTL

ABSTRACT

Sugarcane (*Saccharum* spp.) has an historical and economic relevance in Brazil. We are the largest producer and exporter of sugar and ethanol in the world. Sugar agribusiness has experienced a xx in the last decade not only in Brazil but also around the world as a consequence of increasing demand on renewable and clean sources of energy. As a consequence, the growing area with sugarcane in Brazil is expanding, reaching the central regions of the country. Sugarcane breeding has an important role in developing new cultivars adapted to these new conditions. However, most traits of economic importance in sugarcane have complex genetic architecture making the improvement of new sugarcane cultivars a challenging process. Thus, adoption of strategies that allow for rapid and precise detection of genes associated with quantitative traits is of great interest, representing a valuable tool for sugarcane breeding. Association mapping based on linkage disequilibrium represents a strategy useful for detection of marker-trait associations and may contribute for identifying genes useful for sugarcane breeding. In the present study, association mapping approach was applied to sugarcane in order to evaluate its potential contribution in detecting important associations between SSR and AFLP molecular markers and the characters Height, Diameter and Number of Stalks; % Fiber; % Pol and TCH. Phenotypic data for genotypes were collected from field trials in four locations: Ribeirão Preto, Jaú and Piracicaba in São Paulo and Goianésia in Goiás between 1993 and 2009. Marker-trait associations were tested for each location individually and for all locations simultaneously. A mixed model approach was adopted to test for marker-trait associations. The results suggested the existence of twelve associations involving the characters Stalk Number, % Pol and TCH. Four associations involved stalk number from which three (markers *CIR56*, *ACG_CGT* e *AGG_CAG*) were for all locations and one specific to Goiás (*ACG_CAT*). Seven associations between % Pol and markers *CV60*, *CV106*, *AAG_CAC*, *AAG_CAG* e *ACG_CTT* were detected in Ribeirão Preto. One association between TCH and *ACG_CGC* was detected in Piracicaba

Keywords: Polyploid; Sugarcane; Linkage disequilibrium; Association mapping; Mixed models; QTL

LISTA DE FIGURAS

Figura 1 - Histogramas dos valores fenotípicos para os caracteres analisados 65

Figura 2 - Distribuição do número de alelos para os marcadores SSR, AFLP e para o Total (soma de ambos) avaliada nos 103 genótipos cultivados 69

Figura 3 - Dendrograma mostrando as relações genéticas entre os 103 genótipos cultivados de cana e 3 genótipos selvagens. O método de agrupamento foi *Neighbor-Joining* modificado segundo Gascuel (1997) utilizando a matriz de dissimilaridade genética de Dice. São mostrados apenas os bootstraps dos principais grupos. A barra indica a distância entre os genótipos 75

Figura 4 - Agrupamento dos 103 genótipos cultivados de cana revelados pelo método PCoA. A numeração dos grupos corresponde à mesma numeração adotada no dendrograma. Os dois primeiros eixos ou componentes principais estão indicados na figura com os respectivos valores de variação distribuídos em cada um 77

Figura 5 - Estrutura do desequilíbrio de ligação em função da distância genética para 243 marcas AFLP, RFLP, SSR e TRAP do mapa genético construído a partir de uma população de 92 indivíduos do cruzamento entre os clones 'SP80-180' x 'SP80-4966'. A linha horizontal tracejada representa o percentil 97,5% da distribuição de r^2 para marcas localizadas em grupos de co-segregação diferentes (marcas não ligadas) no mapa genético. A linha sólida representa a curva de regressão para os dados ajustada com a função de LOESS 79

Figura 6 - Genealogia das cultivares SP80-180 e SP80-4966 usadas como parentais para obtenção do mapa genético de ligação brasileiro. Em negrito está destacado também o genótipo POJ100. O símbolo "?" indica que o parental desconhecido (policruzamento). O termo "Self" significa parental autofecundado 81

Figura 7 - Determinação do número de subgrupos na população de mapeamento por associação de acordo com o método proposto por Evanno, Regnault e Goudet (2005). No eixo x é representado o número provável de subgrupos K e no eixo y a estatística ΔK.....87

Figura 8 - Composição dos grupos identificados através do programa STRUCTURE. As cultivares estão dispostas ao logo do eixo horizontal e cada uma é representada por uma única linha vertical dividida em segmentos coloridos proporcionais à fração de seus genoma (valor Q) oriundo em um dos K grupos 87

Figura 9 - Distribuição das estimativas dos coeficientes de coancestralidade (F_{ij}) calculados para os 103 genótipos de cana-de-açúcar pelo segundo o método proposto por Hardy (2003) 89

LISTA DE TABELAS

Tabela 1 - Genótipos de cana-de-açúcar selecionados para compor a população de mapeamento por associação, com seus respectivos parentais 42

Tabela 2 - Valores médios dos caracteres fenotípicos nas quatro regiões onde as variedades foram avaliadas. Cada valor representa a média de todos os experimentos conduzidos em cada região 44

Tabela 3 - Relação dos locos microssatélite selecionados, indicando o critério de seleção e os caracteres aos quais se mostraram associados, seus grupos de homologia e/ou co-segregação e os respectivos trabalhos de referências usados para a seleção 49

Tabela 4 - Sequências dos iniciadores dos locos microssatélites com as respectivas caudas fluorescentes, temperaturas de anelamento e tamanhos observados dos alelos neste trabalho 51

Tabela 5 - Combinações de iniciadores AFLP usados na genotipagem da população. Estão indicados apenas os três nucleotídeos seletivos posicionados na extremidade 3´ dos iniciadores *Eco*RI e *Mse*I. Os iniciadores marcados estão indicados com seus respectivos fluoróforos 55

Tabela 6 - Total de alelos polimórficos e Conteúdo de Informação Polimórfica médio (PIC*m*) para os locos SSR. Os alelos correspondem aos picos observados nos eletroferogramas produzidos no analisador genético ABI PRISM 310, e analisados através dos programas GeneScan 3.7 e Genotyper 3.7 (Applied Biosystems) 67

Tabela 7 - Total de bandas polimórficas e Conteúdo de Informação Polimórfica médio (PIC*m*) para as combinações de iniciadores AFLP. As bandas foram reveladas no analisador automático de DNA Li-cor modelo 4300 e analisadas através do programa SAGA^{MX} (LI-COR Biosciences) 68

Tabela 8 - Marcadores SSR ou AFLP associados com caracteres fenotípicos em cana-de-açúcar de acordo com diferentes modelos de associação.....92

1 INTRODUÇÃO

1.1 Aspectos econômicos do agronegócio sucroalcooleiro no Brasil

Historicamente a cana-de-açúcar sempre possuiu um papel relevante para a economia brasileira e, desde que foi introduzida pelos portugueses no século XVI, figura entre as principais culturas do Brasil. Em 1975 o governo federal lançou o programa Nacional do Álcool (Proálcool) com o objetivo de reduzir a dependência do país em relação ao petróleo. A partir desse momento, iniciou-se um processo de intensa transformação do agronegócio da cana, resultante de grandes investimentos em melhorias tecnológicas. Nos últimos anos, em função da demanda por fontes de energia menos agressivas ao ambiente, o setor passa por uma nova fase de intensos investimentos diante da perspectiva de aumento no consumo mundial de etanol como fonte de combustível alternativo ao petróleo. Além disso, novos processos na área biotecnológica abriram possibilidades de uso antes não exploradas para a cultura e seus derivados como, por exemplo, a produção de plásticos biodegradáveis. Como conseqüência, o agronegócio da cana vem experimentando forte expansão na última década o que pode ser evidenciado pelos números crescentes relacionados ao setor.

Segundo a Companhia Nacional de Abastecimento (CONAB, 2011), a área cultivada com cana-de-açúcar está em expansão, sendo estimada em aproximadamente 8 milhões de hectares na safra 2010/11. Nesta mesma safra foram processadas cerca de 620 milhões de toneladas, sendo 10,3% nas regiões Norte e Nordeste e 89,7% na região Centro-Sul, com destaque para o estado de São Paulo como o maior produtor brasileiro, respondendo com quase 60% da produção nacional. A produtividade média desta safra foi de 77,8 toneladas por hectare. No mesmo período o país produziu cerca de 38 milhões de toneladas de açúcar e cerca de 27 bilhões de litros de etanol. A expressividade desses números coloca o Brasil numa posição de destaque no cenário internacional do agronegócio da cana-de-açúcar e seus principais derivados.

O país é o maior produtor e exportador de açúcar, respondendo por cerca de 20% de toda a produção mundial, sendo que aproximadamente dois terços dessa produção são destinados à exportação, fazendo com que sozinho, o Brasil responda por 37% do mercado internacional deste produto (BRASIL, 2011a).

Em relação ao etanol, o país disputa com os Estados Unidos a liderança na produção mundial, embora a fonte primária para a produção americana deste combustível seja o milho. As perspectivas para o etanol são de demanda crescente para este produto no mercado internacional, de modo que as exportações brasileiras vêm mostrando um expressivo crescimento na última década, atingindo níveis suficientes para colocar o país também na liderança das exportações desse produto. De acordo com a Secretaria de Comércio Exterior do Ministério do Desenvolvimento, Indústria e Comércio (BRASIL, 2011c), em 2008 o Brasil atingiu o ápice da série histórica, exportando 5,1 bilhões de litros de etanol ou o equivalente a 20% do total produzido naquela safra. Em 2010 o desempenho das exportações foi tímido, cerca de 1,9 bilhões de litros, em função da prioridade de abastecimento do mercado interno de carros bicombustíveis e devido à redução das importações americanas que aumentaram sua produção nos últimos anos (CONAB, 2011).

Outra constatação importante ao se analisar as estatísticas referentes à cana-de-açúcar no Brasil é que a forte expansão da produção da cultura nos últimos anos é função muito mais de um aumento na área plantada do que de aumento no rendimento médio da cultura, que parece ter atingido um platô. De 1975, ano de lançamento do Proálcool, até 2010 a área plantada aumentou aproximadamente 400%, enquanto o incremento no rendimento médio no mesmo período foi de pouco mais de 63% (BRASIL, 2011b). Portanto, além do aumento de produção devido à expansão da cultura para áreas ditas marginais ou não tradicionais que apresentam, por exemplo, maior déficit hídrico, a garantia de competitividade e sustentabilidade do agronegócio da cana exigirá investimentos no desenvolvimento contínuo de variedades geneticamente melhoradas visando aumentar o rendimento médio da cultura, além de considerar novos fatores bióticos e abióticos de cada região e sistema de cultivo.

1.2 Taxonomia e origem da cana-de-açúcar

A cana-de-açúcar (*Saccharum* spp.) é uma espécie pertencente à família *Poaceae*, sub-família *Panicoideae*, tribo *Andropogoneae*, subtribo *Saccharinae* e ao gênero *Saccharum* (TZVELEV, 1989). A maioria dos autores considera que o gênero *Saccharum* é composto por seis espécies, sendo duas selvagens, *Saccharum spontaneum* Linnaeus (2*n*=40-128) e *Saccharum robustum* Brandes and Jeswiet ex Grassl. (2*n*=60-205) e quatro formas cultivadas, *Saccharum*

officinarum Linnaeus (2*n*=80), *Saccharum barberi* Jeswiet (2*n*=81-124), *Saccharum sinense* Roxb. (2*n*=111-120) e *Saccharum edule* Hassk. (2*n*=60-80).

Estudos moleculares e citogenéticos realizados principalmente nas duas últimas décadas ajudaram a definir melhor o processo evolutivo que originou as espécies de cana-de-açúcar, a estabelecer uma clara distinção entre os gêneros *Saccharum*, *Erianthus* e *Miscanthus* e reduzir a possibilidade de uma provável contribuição destes últimos dois gêneros para a origem das espécies de cana (GLASZMANN et al., 1989; GLASZMANN; LU; LANAUD, 1990; LU et al., 1994a; D'HONT et al., 1993; D'HONT et al., 1995; BESSE; McINTYRE; BERDING, 1997; ALIX et al., 1998; ALIX et al., 1999; D'HONT; PAULET; GLASZMANN, 2002). Em linhas gerais, os resultados destes estudos suportam o cenário descrito por Brandes (1958 apud GRIVET et al., 2004), no qual se sugere que a cana evoluiu a partir de uma linhagem ancestral pertencente exclusivamente ao gênero *Saccharum* (GRIVET et al., 2004). Inicialmente, este ancestral comum encontrava-se amplamente disperso em duas áreas principais: uma na região continental do sudeste asiático (Java e Sumatra) e outra correspondendo à Austrália, Nova Guiné e ilhas próximas. Acredita-se que no período quaternário o isolamento de populações deste ancestral nas duas regiões permitiu sua diferenciação, dando origem às espécies selvagens *S. spontaneum*, na primeira área, e *S. robustum*, na segunda. Na Nova Guiné, a domesticação de plantas de *S. robustum* ricas em açúcar originou a espécie das canas nobres (*S. officinarum*). Em conseqüência, principalmente da ação humana, clones de *S. officinarum* alcançaram as regiões continentais próximas e ilhas do Pacífico, onde hibridações com populações naturais de *S. spontaneum* deram origem a híbridos interespecíficos robustos, melhor adaptados às condições subtropicais do que os clones nobres. A introdução de clones nobres no norte da Índia e da China permitiu a ocorrência de hibridações interespecíficas entre *S. officinarum*, *S. spontaneum* e *S. robustum*, dando origem às espécies cultivadas *S. barberi* e *S. sinense* (D'HONT; PAULET; GLASZMANN, 2002; GRIVET et al., 2004). Para a espécie *S. edule* os estudos realizados sugerem que sua origem pode ter sido por um processo semelhante ao de *S. officinarum*, ou seja, por seleção direta dentro de *S. robustum*. No entanto, tem sido sugerida também uma origem intergenérica entre *S. officinarum* ou *S. robustum* com uma espécie pertencente a algum gênero relacionado (DANIELS; ROACH, 1987).

1.3 Melhoramento e uso de germoplasma no desenvolvimento das variedades modernas

O melhoramento de cana começou em Barbados e em Java quase que simultaneamente no fim do século XIX. Ming et al. (2006) descrevem cinco fases principais na história de melhoramento da cana. A primeira relacionada ao cruzamento e seleção de clones da cana nobre para obter as cultivares nobres. Apesar de possuírem boa qualidade de moagem, alto conteúdo de açúcar e baixo teor de fibra, faltava-lhes vigor, capacidade de rebrota e resistência a doenças. Na segunda fase ocorreram hibridações interespecíficas entre as canas *S. officinarum* e a espécie selvagem *S. spontaneum*, seguido de sucessivos retrocruzamentos com *S. officinarum*, procedimento que ficou conhecido como nobilização. O objetivo principal era obter novas cultivares que possuíssem o potencial de produção de açúcar das canas nobres e a robustez, vigor e resistência a doenças de *S. spontaneum*. Um fato de grande importância nessa fase foi o desenvolvimento em 1921 da cultivar POJ2878 por melhoristas em Java.

Durante o processo de nobilização ocorreu a transmissão assimétrica de cromossomos, de modo que na geração F_1 e na primeira geração de retrocruzamento (RC_1), os indivíduos da progênie receberam o número somático de cromossomos do parental feminino *S. officinarum* ($2n$=80, x=10) somado ao número gamético do parental masculino *S. spontaneum* ($2n$ = 40-128, x=8), resultando em cultivares altamente poliplóides e com frequente aneuploidia (BUTTERFIELD; D'HONT; BERDING, 2001; D'HONT; GLASZMANN, 2001; GRIVET; ARRUDA, 2001).

A terceira fase consistiu no cruzamento das canas nobilizadas para produzir as canas híbridas melhoradas nos anos de 1930. Estas cultivares tiveram grande importância para a produção de açúcar durante as décadas seguintes em diversas regiões do mundo. Na quarta fase foram desenvolvidas as atuais cultivares de cana-de-açúcar a partir de cruzamentos entre os híbridos obtidos na fase anterior e na de nobilização. Como conseqüência, houve um forte efeito de gargalo no desenvolvimento das canas modernas, pois elas são derivadas essencialmente de poucos genótipos parentais (ARCENEAUX, 1965; ROACH, 1989; JACKSON, 2005). Conscientes desse problema, os melhoristas deram início por volta de 1960 à quinta fase do melhoramento, caracterizada pelo uso de germoplasma selvagem nos programas de melhoramento, na tentativa de ampliar a base genética estreita das cultivares até então desenvolvidas.

Portanto, as cultivares modernas de cana-de-açúcar, denominadas coletivamente *Saccharum* spp., são essencialmente gerações avançadas de híbridos entre *S. officinarum* e *S. spontaneum*, que possuem tipicamente mais de 100 cromossomos (2*n*=100-130) sendo que aproximadamente 80% destes são derivados de *S. officinarum*, 10-15% de *S. spontaneum* e o restante 5-10% são cromossomos recombinantes entre as duas espécies ancestrais (D'HONT et al., 1996).

No Brasil, a cana-de-açúcar foi introduzida por Martin Afonso de Souza em 1502, a partir de mudas trazidas provavelmente da Ilha da Madeira (CESNIK; MIOCQUE, 2004). Nesta fase inicial, a introdução de germoplasma originário de outros países representava a maneira mais rápida e econômica de obtenção de cultivares (LANDELL; BRESSIANI, 2008). No final do século XIX e início do século XX houve intensa preocupação em todo o mundo em relação aos níveis de produtividade, aos problemas fitossanitários, bem como com a melhoria da qualidade das cultivares desenvolvidas (AGUIRRE Jr., 1936; ANDRARDE, 1985; DANTAS, 1960; DEERR, 1921; EDGERTON, 1955; STEVENSON, 1965 apud MATSUOKA; GARCIA; ARIZONO, 2005). Assim como em outras partes do mundo, iniciaram-se no Brasil as primeiras atividades envolvendo o melhoramento de cana, de forma que nas três primeiras décadas do século XX foi organizado um sistema de experimentação voltado para a avaliação de cultivares introduzidas e desenvolvimento de genótipos adaptados às condições brasileiras (LANDELL; BRESSIANI, 2008; MATSUOKA; GARCIA; ARIZONO, 2005).

As duas primeiras estações brasileiras de pesquisa com cana-de-açúcar foram criadas em 1910. A Estação Experimental de Escada, instalada efetivamente em 1913 no estado de Pernambuco, iniciou estudos para a seleção de cultivares resistentes à broca *Diatraea saccharalis* e ao piolho *Trionymus sacchari*, bem como projetos de irrigação com a assessoria de técnicos holandeses. No entanto, as atividades nesta estação foram interrompidas na década de 1920 e o acervo transferido para a Estação Experimental de Barreiros também em Pernambuco (CESNIK; MIOCQUE, 2004). As informações a respeito das cultivares desenvolvidas nesta estação, e nas que se seguiram a ela em Pernambuco são escassas, mas há registro de cultivares como EB4 e EB10 (LANDELL; BRESSIANI, 2008). A Estação de Campos, no estado do Rio de Janeiro, apesar de criada em 1910 passou a atuar efetivamente nas atividades de melhoramento em 1916. Além da introdução e avaliação de cultivares estrangeiras, foi instalado nesta estação um

programa de melhoramento próprio, resultando no desenvolvimento das primeiras cultivares híbridas (‘CB100’ e ‘CB3600’) (LANDELL; BRESSIANI, 2008). Durante várias décadas, as cultivares CB (Campos Brasil), tiveram significativa participação na canavicultura brasileira chegando a ter o predomínio da área cultivada no país (MATSUOKA; GARCIA; ARIZONO, 2005).

Outro programa de grande relevância para o melhoramento da cana-de-açúcar é o do Instituto Agronômico de Campinas (IAC). Já em 1892, os primeiros estudos foram conduzidos através de um ensaio de competição com 42 cultivares nobres introduzidas ou selecionadas por produtores (LANDELL; ALVAREZ, 1993 apud LANDELL; BRESSINAI, 2008). Efetivamente, as atividades de melhoramento neste instituto só se iniciaram em 1934 e as primeiras cultivares desse programa (‘IAC36-25’ e ‘IAC47-31’) foram lançadas na década de 1950. Em seguida, foram desenvolvidas as cultivares (‘IAC48-65’, ‘IAC51-205’, ‘IAC52-250’, ‘IAC58-480’ e ‘IAC64-257’) que atingiram efetivo sucesso comercial em função da expansão da cultura para áreas com solos de fertilidade mais baixa (LANDELL; BRESSINAI, 2008). Além de ser um dos pioneiros na criação de um programa de melhoramento realmente estruturado, o IAC é o programa com maior tradição nesta área no Brasil. Além disso, na década de 1960, dois pesquisadores deste instituto, Carlos Alberto Krug e Hermindo Antunes Filho, contribuíram significativamente para a formação de outros dois importantes programas de melhoramento genético, o do Planalsucar e o da Copersucar (IAC, 2011; disponível em: <http://www.iac.br/>). Atualmente o instituto funciona como Centro Avançado da Pesquisa Tecnológica do Agronegócio de Cana (APTA Cana), com sede em Ribeirão Preto/SP, desenvolvendo pesquisa não apenas na área de melhoramento, mas também em outras áreas relevantes para a cultura. Além disso, em sintonia com os avanços tecnológicos, o centro tem investido também em pesquisas na área de biotecnologia de cana-de-açúcar.

Nos anos de 1969 e 1970 surgiram dois dos maiores programas de melhoramento de cana do Brasil. O primeiro foi o do Centro de Tecnologia Copersucar, criado em 1969 pela Cooperativa Central dos Produtores de Açúcar e Álcool do Estado de São Paulo na cidade de Piracicaba/SP. O ciclo das cultivares da Copersucar (que receberam a sigla SP) teve início na segunda metade da década de 80, com cultivares como a ‘SP70-1143’, ‘SP70-1284’, ‘SP71-1406’ e ‘SP71-6163’, as quais figuraram entre as mais cultivadas na região Centro-Sul nas décadas de

1980 e parte da década de 1990 (COPERSUCAR, 1995 apud MATSUOKA, GARCIA. ARIZONO, 2005). Além das atividades de melhoramento tradicional, este programa se destaca também pela forte atuação na área de biotecnologia. Em Agosto de 2004, o centro passou a se chamar Centro de Tecnologia Canavieira (CTC) e as cultivares liberadas desde então recebem a sigla CTC. Atualmente, o CTC é o maior centro privado de pesquisas com cana-de-açúcar do Brasil, atuando dinamicamente em toda a cadeia produtiva do setor canavieiro, desde o melhoramento da cana até a produção final de açúcar, etanol e de energia (CTC, 2011; disponível em: <http://www.ctcanavieira.com.br/>).

O segundo programa foi o do Planalsucar (Programa Nacional de Melhoramento de Cana-de-açúcar), responsável pelo desenvolvimento das cultivares RB (República do Brasil). Este programa foi criado pelo Instituto do Açúcar e do Álcool (IAA), do então Ministério da Indústria e do Comércio. O objetivo do governo federal era desenvolver um programa de abrangência nacional com uma estrutura de pesquisa capaz de lidar com os problemas tecnológicos que a cultura estava sujeita na segunda metade da década de 1960. O programa liberou 18 cultivares até sua extinção em 1990, sendo a mais importante delas a 'RB72-454' (MATSUOKA; GARCIA; ARIZONO, 2005). Em 1992, toda a estrutura do Planalsucar foi transferida para um conjunto de universidades federais, compondo assim a Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RIDESA), com o objetivo de dar continuidade ao desenvolvimento de pesquisas e melhoramento genético da cana-de-açúcar. Para se ter uma idéia da importância deste programa, atualmente as cultivares RB ocupam aproximadamente 58% da área cultivada com cana no Brasil (RIDESA, 2011; disponível em: < http://www.ridesa.com.br/>).

Nos diferentes programas de melhoramento no Brasil e no mundo, a exploração da hibridação pelos melhoristas proporcionou uma diversificação do perfil varietal da cultura, refletindo-se em cultivares produtivas, com boa capacidade de perfilhamento, diâmetro dos colmos adequado, maior resistência a pragas e doenças e maior capacidade de rebrota após vários cortes. Apesar destes avanços, é necessário garantir que os incrementos de produtividade a cada ciclo de melhoramento sejam suficientes para atender à demanda projetada de cana e seus derivados para as próximas décadas.

Os programas atuais de melhoramento de cana baseiam-se essencialmente em esquemas de seleção recorrente. Para maximizar a eficiência da seleção, a maioria dos programas utiliza grandes populações de indivíduos nas etapas iniciais de melhoramento, mas a eficiência pode não ser suficiente para permitir os níveis desejados de ganhos. Devido ao pouco conhecimento do controle genético de caracteres quantitativos, responsáveis pelas características de maior importância comercial, a identificação dos genótipos promissores baseia-se na seleção para caracteres secundários, correlacionados com o caráter principal (BRESSIANI, 1993). Uma grande desvantagem desta estratégia é que as correlações obtidas dependem em certo grau das condições ambientais de avaliação, o que prejudica bastante na identificação dos genótipos superiores.

Neste contexto, os marcadores moleculares representam uma abordagem importante para a análise do genoma de cana, pois oferecem a possibilidade de analisar diretamente ao nível molecular a variação existente e identificar locos relacionados com características de importância econômica, aumentando a eficiência da seleção.

1.4 Marcadores moleculares em cana-de-açúcar

O uso de abordagens baseadas em marcadores moleculares de DNA oferece a oportunidade de explorar o genoma das espécies fornecendo informações importantes a respeito de sua estrutura e função. Desde o seu desenvolvimento, uma de suas principais aplicações tem sido na construção de mapas genéticos de ligação. Numa abordagem tradicional, os mapas genéticos são usados como ponto de partida para a análise dos genomas permitindo, por exemplo, dissecar geneticamente os caracteres de herança complexa em caracteres de herança simples, fornecendo informações importantes para o melhoramento das espécies vegetais cultivadas.

No entanto, o mapeamento, tal como é feito em plantas diplóides, não pode ser diretamente aplicado em cana-de-açúcar devido à complexidade genômica das cultivares modernas, representada pela elevada poliploidia com ocorrência concomitante de aneuploidia; alta heterozigosidade; dosagens alélicas não conhecidas e difíceis de determinar; e a contribuição exata de cada genoma que originou as cultivares modernas de cana (ALWALA; KIMBENG, 2010). Além disso, de modo geral os métodos estatísticos para poliplóides são mais complexos e,

especialmente no caso de cana, há carência de modelos adequados às especificidades de sua natureza genética.

Um método de mapeamento genético para poliplóides com pareamento bivalente na meiose foi proposto por Wu et al. (1992). Esse método é baseado no uso de marcadores em dose única (MDU), ou seja, em marcadores presentes em uma única cópia ou dose em um ou nos dois parentais. Dessa forma, os indivíduos da progênie são genotipados com base no padrão de presença e ausência das marcas MDU, as quais segregam nas proporções 1:1 ou 3:1, independentemente da natureza do marcador utilizado. Essa estratégia permitiu a construção de mapas genéticos de ligação em cana-de-açúcar utilizando de diferentes marcadores moleculares como RFLP (*Restriction Fragment Length Polymorphism*), microssatélites ou SSR (*Simple Sequence Repeats*), RAPD (*Random Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*) e TRAP (*Target Region Amplification Polymorphism*). Estes mapas foram obtidos a partir de populações F_1 derivadas de cruzamentos biparentais, envolvendo diferentes tipos de parentais como, por exemplo, *S. spontaneum* 'SES 208' (2*n* = 64) e seu duplo-haplóide 'ADP068' (AL-JANABI et al., 1993; Da SILVA et al., 1993; 1995); cruzamentos interespecíficos entre *S. officinarum* 'Louisiana Purple' (2*n* = 80) x *S. robustum* 'Mol 5829' (2*n* = 80) (MUDJE et al., 1996; GUIMARÃES et al., 1999) ou entre *S. officinarum* 'Green German' (2*n* = 97–117) x *S. spontaneum* 'IND 81-146' (2*n* = 52-56), *S. officinarum* 'Muntok Java' (2*n* = 140) X *S. spontaneum* 'PIN 84-1' (2*n* = 96) (MING et al., 1998; 2002a) e *S. officinarum* 'La Striped' (2*n* = 80) x *S. spontaneum* 'SES 147B' (2*n* = 64) (ALWALA et al., 2008). Há também mapas desenvolvidos a partir de populações derivadas ou envolvendo variedades comerciais ou clones elite como, por exemplo, *S. officinarum* 'IJ76-514' (2*n* = 80) x 'Q165' (2n ≈ 115) (AITKEN; JACKSON, McINTYRE, 2005, 2007); 'Q117' x 'MQ77-340' (REFFAY et al., 2005); 'Q117' x Linhagem '74C42' (McINTYRE et al., 2005a); 'SP80-180' x 'SP80-4966' (GARCIA et al., 2006; OLIVEIRA et al., 2007); auto-fecundação da cultivar 'R570' (2*n* = 107–115) (GRIVET et al., 1996; ASNAGHI et al., 2000; HOARAU et al., 2001; ROSSI et al., 2003); 'R570' x 'MQ76-53' (RABOIN et al., 2006). Devido ao tamanho do genoma de cana um grande número de marcadores é exigido e estima-se que mesmo o mais saturado dos mapas cobre somente cerca de dois terços do seu genoma (AITKEN; JACKSON; McINTYRE, 2005). Além disso, a cobertura genômica obtida nos mapas desenvolvidos até o momento é irregular, ou seja, parece que a fração do genoma originária de *S. spontaneum* apresenta uma densidade maior de

marcadores quando comparada com os cromossomos originários de *S. officinarum* (GRIVET; ARRUDA, 2001).

Outra aplicação de grande importância dos marcadores tem sido na investigação da diversidade envolvendo o germoplasma de *Saccharum*. De especial interesse tem sido a caracterização da origem e dos níveis de diversidade dentro das cultivares modernas de cana e das duas principais espécies que contribuíram para o seu genoma *S. officinarum* e *S. spontaneum*. Os primeiros estudos neste sentido sugeriram haver pouca diversidade no germoplasma de *S. officinarum* (GLASZMANN et al., 1989; GLASZMANN; LU; LANAUD, 1990; BURNQUIST; SORRELLS; TANKSLEY, 1992; D'HONT et al., 1993; LU et al., 1994a, 1994b). No entanto, estudos recentes têm demonstrado haver uma quantidade considerável de diversidade nesta espécie (JANNOO et al., 1999b; SELVI et al., 2003; BROWN et al., 2007; AITKEN; JACKSON; McINTYRE, 2006, ALWALLA et al., 2006) e que as observações iniciais parecem ter sido muito mais um problema relacionado ao baixo número de acessos avaliados do que a uma baixa diversidade de fato. Análises moleculares feitas com diferentes marcadores no germoplasma de *S. spontaneum* têm mostrado que esta espécie possui os níveis mais elevados de diversidade intra-específica quando comparada com as demais espécies do gênero (GLASZMANN; LU; LANAUD, 1990; BURNQUIST; SORRELLS; TANKSLEY, 1992; SOBRAL et al., 1994; NAIR et al., 1999; PAN et al., 2004; SHEJI et al., 2006). Evidências recentes sugerem inclusive que genes com contribuição positiva para o acúmulo de açúcar poderiam estar presentes em seu genoma (REFFAY et al., 2005; ALWALA et al., 2008).

Por causa de sua origem híbrida, da poliploidia e da propagação vegetativa, as cultivares modernas de cana são altamente heterozigotas, ou seja, possuem elevada diversidade alélica nos locos (JANNOO et at., 1999b; SELVI et al., 2003; ALWALA et al., 2006; BROWN et al., 2007). Os principais estudos conduzidos para caracterização da diversidade dentro do germoplasma cultivado têm demonstrado que uma parte significativa da variação observada é derivada da fração do genoma originária de *S. spontaneum* (D'HONT et al., 1996; JANNOO et al., 1999a; SELVI et al., 2005; AITKEN; JACKSON; McINTYRE, 2006).

Apesar de estes estudos mostrarem que há considerável variabilidade disponível nas espécies selvagens *S. officinarum* e *S. spontaneum*, somente uma fração pequena desta diversidade foi efetivamente utilizada. Isso se deve ao fato de que um número restrito de

genótipos destas espécies parentais foi usado no início do desenvolvimento das cultivares modernas. Consequentemente, parece que a maior parte da variabilidade presente no germoplasma destas cultivares já foi explorada pelo melhoramento e que esta seria a causa principal das baixas taxas de progresso genético observadas nos programas de melhoramento (AITKEN; McNEIL, 2010). Portanto, a ampliação da base genética de modo a permitir a continuidade dos ganhos por seleção, depende da caracterização de novas fontes de germoplasma dentro do gênero *Saccharum* ou em gêneros relacionados como *Erianthus* e *Miscanthus*, principalmente.

1.5 Mapeamento de QTLs

1.5.1 Mapeamento via análise de ligação

A maioria das características com as quais lidamos no melhoramento de plantas são características complexas, controladas por muitos genes e cuja expressão é influenciada pelo ambiente e por interações entre os genes. As regiões no genoma onde se localizam os genes com efeitos significativos sobre tais características são denominadas QTLs (*Quantitative Trait Locus*) (FALCONER; MACKAY, 1996).

Nas últimas décadas grande esforço tem sido feito para a identificação, localização e mensuração dos efeitos dos QTLs em diversas espécies de interesse econômico. A maioria destes estudos utilizou estratégias de mapeamento de QTLs que incluem primeiramente a geração de uma população de mapeamento (F_1, F_2, retrocruzamento, RILs e NILs) a partir de cruzamentos biparentais, genotipagem dos indivíduos com marcadores moleculares, fenotipagem dos indivíduos para a(s) característica(s) de interesse e, por último, análise dos resultados via mapeamento de ligação. Neste processo, a construção criteriosa dos mapas genéticos de ligação é uma etapa de fundamental importância para o mapeamento e utilização dos QTLs no melhoramento de espécies agrícolas (STAUB; SERQUEN; GUPTA, 1996).

Utilizando esta abordagem em populações F_1, diversos caracteres em cana-de-açúcar foram mapeados, incluindo caracteres relacionados a doenças, produção de massa e de açúcar (DAUGROIS et al., 1996; MUDGE et al., 1996; MING et al., 2001, 2002b, 2002c; HOARAU et al., 2002; McINTYRE et al., 2005a, 2005b; AITKEN; JACKSON; McINTYRE, 2006; RABOIN et al., 2006; AL-JANABI et. al., 2007; AITKEN et al., 2008; ALWALA et al., 2008; PIPERIDIS

et al., 2008; PINTO et al., 2010). Por causa da natureza poliplóide e altamente heterozigótica de seu genoma, a maioria dos caracteres de interesse em cana possui padrão quantitativo de herança. Esse fato tem sido confirmado em diversos estudos que identificaram grande número de alelos com pequenos efeitos sobre a variação fenotípica dos caracteres (SILLS et al., 1995; MING et al., 2001, 2002b, 2002c; HOARAU et al., 2002; McINTYRE et al., 2005a; AL-JANABI et. al., 2007; PIPERIDIS et al., 2008; PINTO et al., 2010).

Apesar da contribuição indiscutível desses estudos, o mapeamento de QTLs utilizando mapas de ligação possui importantes limitações como: 1) necessidade de obtenção de populações segregantes a partir de cruzamentos planejados, o que demanda tempo e recursos; 2) baixa precisão na localização dos QTLs devido ao limitado número de recombinações desde a geração de parental, fazendo com que marcadores localizados a distâncias relativamente grandes do QTL mostrem associações significativas; 3) amostragem parcial dos alelos envolvidos no controle do caráter, ou seja, a quantidade de variação genética segregando na população de mapeamento é limitada. Além disso, geralmente são usados apenas dois parentais para a obtenção da população de mapeamento e, nem sempre, estes genótipos são representativos do germoplasma efetivamente usado no melhoramento; 4) baixa sensibilidade na detecção de QTLs com efeitos pequenos sobre a variação quantitativa.

No contexto específico do mapeamento de QTLs em cana via análise de ligação, a identificação de associações significativas tem sido limitada também por razões como a baixa saturação dos mapas; dificuldade na obtenção de linhagens endogâmicas, o que restringe ou impede a aplicação de abordagens estatísticas mais robustas e precisas; elevados níveis de poliploidia e heterozigosidade, que juntos tornam complexos os padrões de segregação, não permitindo que se conheça o número de alelos segregando em cada loco e impossibilitando a determinação das fases de ligação entre o marcador e o QTL. Além disso, há necessidade de métodos estatísticos mais específicos capazes de extrair o máximo de informação dos estudos de QTL em poliplóides complexos como a cana (LAKSHMANAN et al., 2005).

Estas limitações e dificuldades motivaram o interesse no uso de uma abordagem alternativa para o mapeamento de QTLs em plantas, conhecida como mapeamento por associação ou mapeamento por desequilíbrio de ligação.

1.5.2 Mapeamento por associação

Os termos Mapeamento por Associação e Mapeamento por Desequilíbrio de Ligação são usados na literatura como sinônimos. Além destes, outros termos podem ser encontrados como Genética Associativa e Estudos de Associação Genética.

O mapeamento por associação emergiu recentemente como um método promissor para a detecção de QTLs em plantas (REMINGTON et al., 2001; GUPTA et al., 2005; MACKAY; POWELL, 2006). Através desta metodologia, genes envolvidos na variação de características fenotípicas podem ser identificados se associações estatísticas significativas entre elas e os polimorfismos nos marcadores moleculares puderem ser detectadas. O fenômeno responsável por essas associações é denominado desequilíbrio de ligação ou desequilíbrio de fase gamética, o qual se refere à associação não casual entre alelos em diferentes locos, ou seja, quando determinado alelo num loco está associado a outro alelo em um segundo loco numa freqüência maior do que a esperada caso estes alelos se combinassem de forma independente. O desequilíbrio de ligação está por trás de todas as formas de mapeamento de QTLs. Contudo, no mapeamento via análise de ligação, o desequilíbrio decorre exclusivamente de associações físicas, detectadas através de populações estruturadas, resultantes de cruzamentos planejados. No contexto do mapeamento por associação o desequilíbrio de ligação pode ser derivado de associações físicas ou de outros fatores presentes na população onde os indivíduos são, a princípio, não relacionados (NORDBORG; TAVARÉ, 2002).

Estatisticamente, o desequilíbrio de ligação é medido por meio do desvio entre as freqüências observadas para determinadas combinações de alelos e suas freqüências esperadas, assumindo que há independência total na segregação destes alelos. Desta forma, desvios diferentes do valor zero sugerem uma situação de desequilíbrio, ou seja, não há independência entre os alelos. As duas medidas estatísticas mais comuns do desequilíbrio e que refletem diferentes aspectos relacionados a ele são D' e r^2. A medida D' reflete o histórico de recombinação da espécie e é, portanto, muito útil para estimar diferenças de recombinação entre locos. No entanto, estimativas de D' podem ser fortemente superestimadas em amostras pequenas. Já a medida r^2 resume os eventos de recombinação e mutação e também dá um indicativo de como marcadores podem estar correlacionados com os QTLs. Por esta razão, r^2 tem sido a medida preferida nos estudos de associação.

Comparado ao mapeamento via análise de ligação, o mapeamento por associação apresenta as seguintes vantagens: 1) grande flexibilidade na escolha da população de mapeamento que pode ser constituída por qualquer coleção de genótipos, ou seja, podem ser utilizadas populações derivadas de cruzamentos, coleções de bancos de germoplasma ou de melhoramento, linhagens em desenvolvimento e populações naturais; 2) por não depender de cruzamentos biparentais, esta abordagem aproveita todo o histórico de recombinações ocorridas ao longo de várias gerações de cruzamentos, aumentando o poder na detecção de associações (RAFALSKI, 2002); 3) as estimativas dos efeitos dos QTLs são mais realistas uma vez que eles são avaliados num conjunto de genótipos relevantes para o melhoramento (MALOSETTI et al., 2007). Uma desvantagem freqüentemente citada para esta abordagem é a necessidade de uma alta saturação de marcadores no genoma da espécie em estudo. No entanto, essa justificativa só é válida para espécies cujo desequilíbrio de ligação se estende por regiões muito curtas no genoma. Outra desvantagem ou dificuldade potencial está relacionada aos modelos estatísticos usados para as análises de associação. Uma vez que num estudo de mapeamento por associação uma grande quantidade de dados moleculares e fenotípicos está geralmente disponível, faz-se necessária a adoção de métodos robustos, capazes de lidar não só com o volume de informações, mas que sejam capazes, principalmente, de extrair as informações de natureza genética neles contidas.

O mapeamento por associação foi originalmente desenvolvido para mapear genes relacionados a doenças em humanos (KEREM et al., 1989). Em plantas, os estudos iniciais foram realizados em espécies modelo como milho (REMINGTON et al., 2001; THORNSBERRY et al., 2001; WHITT et al., 2002) e *Arabidopsis thaliana* (NORDBORG, 2000; NORDBORG et al., 2002). No entanto, houve uma resistência inicial à sua adoção em plantas devido à falta geral de conhecimentos básicos sobre os níveis de desequilíbrio de ligação nas espécies vegetais e aos efeitos gerados pela existência de estruturação dentro das populações (FLINT-GARCIA; THORNSBERRY; CUCKLER, 2003). Nos últimos anos, porém, foram desenvolvidos diversos métodos estatísticos que têm facilitado sua aplicação e, como conseqüência, diversos estudos nesta área têm sido publicados (WILSON et al., 2004; SCHNELL et al., 2005; BRESEGHELLO; SORRELS, 2006; TOMMASINI et al., 2007; AGRAMA; EISENGA; YAN, 2007; JUN et al., 2008).

A caracterização do desequilíbrio de ligação dentro e entre populações, assim como em diferentes regiões do genoma é crucial para o mapeamento por associação. A distância ao longo da qual o desequilíbrio se estende no genoma irá influenciar o número e a densidade de marcadores e aspectos do desenho experimental adotado nos estudos. Além disso, diversos fatores podem explicar a existência de desequilíbrio e, portanto, as associações entre o marcador e o caráter de interesse. Entre estes fatores estão ligação física entre os alelos, mistura resultante de migração ou introdução de novos genótipos, mutação, seleção natural e artificial, sistema reprodutivo, efeito fundador e deriva genética (RAFALSKI; MORGANTE, 2004; ORAGUZIE et al., 2007).

No contexto do mapeamento por associação aplicado ao melhoramento o que se busca é detectar associações decorrentes de ligação física entre o marcador e o caráter, uma vez que estas são as associações realmente úteis ao melhoramento. Portanto, o sucesso na aplicação desta abordagem depende em grande parte da capacidade de separar o desequilíbrio decorrente de ligação daquele devido aos outros fatores (MALOSETTI et al., 2007). Estes fatores podem se manifestar através de estruturação na população, ou seja, pela presença de grupos dentro da população. A estruturação é um fenômeno bastante comum e pode resultar em falsas associações entre o marcador e o gene que afeta o caráter, devendo, portanto, ser avaliada e incorporada como um fator de correção na análise de associação. O complicador é que ela se manisfesta em diversos níveis, os quais nem sempre são evidentes ou simples de identificar. Um desses níveis é conseqüência da distribuição desigual dos alelos na população e é decorrente das relações de parentesco entre os indivíduos, misturas, sistema de acasalamento ou deriva. Em outras situações ela é o resultado da seleção, natural ou artificial, praticada durante a evolução, domesticação e melhoramento da espécie. Na prática, a estruturação numa população é resultado de uma combinação de vários desses fenômenos, tornando complexo o seu entendimento.

Os níveis de estruturação na população podem ser estimados através do uso de marcadores independentes cobrindo todo o genoma (PRITCHARD; ROSENBERG, 1999; PRITCHARD; STEPHENS; DONNELLY, 2000b; BRADBURY et al., 2007). Entre os principais métodos disponíveis para sua avaliação estão análises baseadas em modelo de genética de populações (PRITCHARD et al., 2000a), análise de componentes principais (PRICE et al.,

2006; ZHAO et al., 2007, KANG et al., 2010) e análise de coancestralidade (HARDY; VEKEMANS, 2002).

A análise implementada no programa STRUCTURE (PRITCHARD; STEPHENS; DONNELLY, 2000b; FALUSH; STEPHENS; PRITCHARD, 2003, 2007; HUBISZ et al., 2009) tem sido uma das mais utilizadas para a determinação da presença de subgrupos. O programa utiliza um método de agrupamento baseado em estatística Bayesiana para inferir o número de subgrupos presentes na amostra experimental. Para tanto, o método exige o uso de marcadores não ligados ou fracamente ligados. Segundo Falush, Stephens e Pritchard (2003), marcadores com forte ligação poderiam fazer com que o programa superestime o número de subgrupos, mas os autores não definiram um nível para tal. De forma resumida, o programa busca um modelo em que existem K sub-populações ou subgrupos (K é desconhecido *a priori*), onde cada uma delas é caracterizada por freqüências alélicas específicas em equilíbrio de ligação e de Hardy-Weinberg. Uma vez atingidas essas condições, são estimadas freqüências alélicas e a ancestralidade, ou seja, a proporção do genoma de cada indivíduo que teve, eventualmente, origem numa das outras sub-populações (HUBISZ et al., 2009).

Com relação à estruturação devido às relações de parentesco, os métodos se baseiam essencialmente no cálculo de índices de coancestralidade, que refletem o grau de covariância genética entre os indivíduos. O coeficiente de coancestralidade (F_{ij}) entre dois indivíduos *i* e *j* é definido como a probabilidade de dois genes homólogos, tomados aos acaso, serem idênticos por descendência (IBD). Coeficientes de coancestralidade podem ser calculados a partir dos registros de *pedigree* ou a partir dos marcadores genotipados na população. No entanto, tem sido sugerido que a segunda abordagem seria mais eficaz em explicar eventuais desvios de contribuição dos parentais para a progênie, decorrentes de seleção e deriva (BERNARDO; MURIGNEUX; KARAMAN, 1996; YU et al., 2006). Alguns autores têm chamado a atenção para o fato de que estas estimativas dos coeficientes de coancestralidade são obtidas a partir de pressuposições arbitrárias e, em alguns casos, não realistas no contexto do mapeamento por associação (ZHAO et al., 2007; STICH et al., 2008; ZHANG et al., 2009; KANG et al., 2010). Por exemplo, a definição de como os indivíduos estão relacionados uns aos outros em termos de suas genealogias nem sempre é uma tarefa simples e a própria definição do termo “relacionado” é arbitrária. Alternativamente, tem sido sugerido o uso de coeficientes baseados na proporção de alelos

compartilhados (ZHAO et al., 2007; KANG et al., 2008), Alelos Idênticos por Estado (IBS) (KANG et al., 2010) ou até mesmo coeficientes baseados em índices de similaridade genética como Jaccard, Dice, Gower, entre outros (BUIL et al., 2009).

1.5.3 Estudos de desequilíbrio de ligação e Mapeamento por associação em cana

Em cana-de-açúcar, há evidências de que o desequilíbrio de ligação está presente de forma extensiva no genoma desta espécie, o que é explicado por sua história de melhoramento, derivada de um número limitado de genótipos e gerações, combinado com forte efeito fundador. Num estudo envolvendo 59 cultivares representando os clones comerciais mais importantes nas ilhas Maurícius e algumas cultivares tradicionais envolvidas em suas genealogias, Jannoo et al. (1999a) investigaram os níveis de desequilíbrio com o objetivo de verificar se haplótipos estabelecidos no início do melhoramento das variedades modernas de cana ainda poderiam ser detectados atualmente. Para tanto, eles utilizaram como referência um mapa genético da cultivar 'R570' (GRIVET et al., 1996). Os autores observaram que a maioria das associações significativas encontrava-se a distância de até 10 cM e envolvia marcadores presentes numa freqüência de 30 a 40 % nas variedades analisadas. Segundo estes autores o nível de desequilíbrio detectado seria uma indicação do tamanho aproximado dos blocos cromossômicos herdados pelas variedades modernas a partir dos genótipos utilizados no início do melhoramento. Raboin et al. (2008) realizaram um segundo estudo com o objetivo de estabelecer um meio prático para uso do desequilíbrio na identificação de alelos importantes para o melhoramento. Neste estudo foi utilizado como referência um mapa genético mais saturado da cultivar 'R570' (HOARAU et al., 2001; RABOIN et al., 2006). Esses autores detectaram níveis significativos de desequilíbrio em distâncias de até 40 cM, embora a maior parte tenha se concentrado nos primeiros 30 cM, sendo que os níveis mais expressivos de desequilíbrio foram observados nos primeiros 5 cM, com queda progressiva a partir desta distância.

Os resultados desses dois estudos ressaltam o valor prático do uso desequilíbrio de ligação não apenas no mapeamento de QTLs, mas também para entender melhor o processo de formação do genoma das variedades modernas de cana. Além disso, pelo menos do ponto de vista dos níveis de desequilíbrio observados, seria possível identificar marcas associadas a caracteres importantes em cana, mesmo em situações de baixa cobertura do genoma. Considerando por exemplo o nível de 5 cM, o número mínimo de marcadores necessários para atingir a densidade

de uma ou duas marcas estaria entre 300 e 600, assumindo o tamanho do genoma haplóide de cana em 1.500 cM (RABOIN et al., 2008).

Seguindo essa possibilidade, Wei et al. (2006) realizaram um estudo numa coleção de 154 variedades, buscando identificar associações entre marcadores moleculares (1.068 marcadores AFLP e 141 SSR) e resistência aos patógenos *Pachymetra chaunorhiza*, *Xanthomonas albilineans*, *Fiji disease virus* e *Ustilago scitaminea*. Os autores realizaram análise de marcas simples como uma evidência inicial de associação e, em seguida, análise de regressão simples, onde também foi avaliada a contribuição dos efeitos de estrutura da população. As marcas significativamente associadas resultantes das análises preliminares foram, então, usadas numa análise de regressão múltipla. Mesmo não tendo sido possível determinar com precisão o número exato de subgrupos na população, por meio desta estratégia foram identificadas onze marcas associadas com a resistência a *Ustilago scitaminea*, quatro para *Pachymetra chaunorhiza*, *Xanthomonas albilineans* e *Fiji disease virus*.

Butterfield (2007) realizou um estudo para identificar marcadores associados com a resistência ao fungo causador do carvão (*Ustilago scitaminea*) e à broca do colmo (*Eldana saccharina* Walker) (Lepidoptera: Pyralidae), uma das pragas mais destrutivas da cana-de-açúcar na África do Sul. Para tanto, o autor utilizou 77 genótipos de cana-de-açúcar que foram genotipados com 275 marcadores RFLP e 1.056 marcadores AFLP. A análise de regressão dos dados fenotípicos revelou 64 marcadores significativamente associados à resistência ao carvão e 115 à broca do colmo. Utilizando o método de regressão *stepwise*, Buttefield foi capaz de identificar modelos, contendo os seis marcadores com maior efeito sobre a variação para as características estudadas, que explicaram 54% da variação para a resistência ao carvão e 62% da resistência à broca. Além de definir um modelo de associação útil para a identificação de marcadores associados com a resistência ao carvão e à broca, este estudo teve como objetivos validar a estratégia de mapeamento baseada no desequilíbrio de ligação e determinar se os níveis de desequilíbrio observados seriam úteis para aplicação no contexto do melhoramento molecular; e desenvolver um mapa de desequilíbrio para identificar haplótipos contendo marcadores associados com as características estudadas. Um mapa contendo 841 marcadores reunidos em 231 haplótipos foi obtido, permitindo que o autor pudesse estudar a origem destes haplótipos presentes nas cultivares modernas, a partir de genótipos parentais ancestrais. As análises de

desequilíbrio mostraram que 58% dos haplótipos identificados se estendiam por distâncias maiores que 10 cM, confirmando o potencial de uso do desequilíbrio como uma ferramenta útil para o melhoramento de cana. Além disso, foram identificados haplótipos significativamente associados com a resistência ou suscetibilidade (*Ustilago scitaminea*) e à broca do colmo (*Eldana saccharina* Walker).

Num estudo mais amplo, Wei et al. (2010) utilizaram um modelo de efeitos mistos para avaliar simultaneamente os efeitos de estruturação na população decorrente das relações de parentesco, interação genótipo por ambientes e variações espaciais associadas aos experimentos. Cerca de 480 genótipos foram avaliados em três locais para os caracteres tonelada de cana por hectare e conteúdo de açúcar. Foram utilizados 1.531 marcadores DArT classificados como discretos, isto é, genotipados como presença e ausência e 15.360 marcadores DArT denominados contínuos, isto é, relacionados ao número de cópias de cada alelo por genótipo. Para cada local ficou evidenciado que quando um ou mais dos efeitos anteriores não foi contemplado no modelo, o número de marcas associadas aumentava. Os autores observaram também que a variação espacial dentro dos experimentos teve um impacto bem menor do que a interação genótipo por ambientes e que a estrutura da população (avaliada a partir das informações de *pedigree*) foi o fator com efeitos mais pronunciados nas associações. Usando o modelo que contemplava todos os efeitos foi possível detectar várias marcas associadas com os dois caracteres estudados. Neste estudo a avaliação da estrutura do desequilíbrio de ligação no genoma de cana utilizou como referência o mapa de ligação derivado do cruzamento ‘IJ76-514’ x ‘Q165’ (AITKEN; JACKSON; McINTYRE, 2005)

Portanto, estes trabalhos sugerem que esta é uma abordagem promissora para a identificação de marcadores associados a caracteres de importância agronômica em cana-de-açúcar e que, nesse sentido, sua aplicação poderia contribuir de forma considerável para os programas de melhoramento.

1.6 Modelos mistos em estudos de associação

Modelos mistos são modelos estatísticos que incorporam ao mesmo tempo efeitos fixos e aleatórios. Sua função primária é descrever as relações entre uma variável de resposta (fenótipo) e algumas covariáveis nos dados, as quais foram medidas ou observadas junto com a variável de resposta. A crescente popularidade desses modelos é explicada pela grande flexibilidade oferecida na modelagem de correlações intra-grupo presente nos dados e na manipulação de dados balanceados (PINHEIRO; BATES, 2000). Em função destas características, os modelos mistos proporcionam uma excelente estratégia para detectar associações entre marcadores e caracteres fenotípicos, uma vez que eles podem lidar de forma adequada com os vários níveis de estruturação numa população, diminuindo os erros tipo I e tipo II. De acordo com Henderson (1975), a equação do modelo misto tem a seguinte expressão geral:

$$y = \mathrm{X}\beta + \mathrm{Z}u + e$$

onde y é o vetor de observações fenotípicas (variável de resposta); β é um vetor de efeitos fixos; u é um vetor de efeitos aleatórios não observados, e é o vetor aleatório de efeitos residuais; X e Z são matrizes de incidência, relacionando y a β e u, respectivamente. É assumido que os vetores aleatórios u e e são normalmente distribuídos com média zero e matrizes de variância-covariância G e R, respectivamente. Diferentes estruturas para estas matrizes de variância-covariância são possíveis.

Há na literatura diversas versões do modelo básico adaptados para o contexto do mapeamento por associação (YU et al., 2006; CROSSA et al., 2007; MALOSETTI et al., 2007; AULCHENKO; KONING; HALEY, 2007; STICH et al., 2008). O método mais popular foi proposto por Yu et al. (2006), o qual foi implementado no programa TASSEL (BRADBURY et al., 2007). Basicamente, o programa visa corrigir num único modelo os efeitos do desequilíbrio de ligação causados pela presença de estruturação na população ou decorrentes de efeitos de parentesco entre os indivíduos, uma vez que estas têm sido apontadas como as duas principais causas de falsas associações. Através do uso de um grande número de marcadores independentes

e distribuídos aleatoriamente no genoma são obtidas previamente as matrizes Q, com os efeitos de estrutura da população, e K, expressando as relações de parentesco dos genótipos. Tais efeitos são incluídos como covariáveis para ajustar um modelo de associação que explique eventuais correlações entre os marcadores e os caracteres fenotípicos. Uma vantagem é a possibilidade de usar matrizes Q e K obtidas através de diferentes métodos. Diversos estudos publicados nos últimos anos têm comprovado a eficácia desta abordagem na identificação de associações (ROSTOKS et al., 2006; AGRAMA; EIZENGA; YAN, 2007; TOMMASINI et al., 2007; ZHAO et al., 2007; REIMER et al., 2008; ABDURAKHMONOV et al., 2009; YAO et al., 2009).

Portanto, considerando que os estudos de desequilíbrio de ligação realizados em cana-de-açúcar até o momento foram realizados tendo como referência, basicamente as populações de mapeamento derivadas das cultivares ‘R570’ e ‘Q165’, e, tendo em vista o portencial de aplicação da abordagem de mapeamento por associação na detecção de regiões no genoma associados com caracteres quantitativos, os objetivos deste trabalho foram: 1) avaliar os níveis e extensão do desequilíbrio de ligação no genoma da cana-de-açúcar, utilizando como referência um mapa de ligação construído a partir de cultivares brasileiras; e, 2) identificar associações entre marcadores moleculares SSR e AFLP com caracteres agronômicos em cana-de-açúcar, aplicando o conceito de mapeamento por associação através da abordagem de modelos lineares mistos.

2 MATERIAL E MÉTODOS

2.1 Material vegetal

Os 103 genótipos que constituem a população (painel de diversidade) deste estudo foram selecionados no banco de germoplasma do Centro de Cana do Instituto Agronômico de Campinas (IAC). A seleção foi feita de modo que a população fosse composta por cultivares comerciais recentes e tradicionais e clones em avaliação. Foi utilizado também como critério de seleção a importância de alguns genótipos como genitores nos cruzamentos realizados pelo programa de melhoramento do IAC. Além disso, o contraste para as características fenotípicas de interesse também foi considerado. Três genótipos selvagens, 'Badilla de Java' (*Saccharum officinarum*), 'SES 205A' (*Saccharum spontaneum*) e 'Chunnee' (*Saccharum barberi*) também foram selecionados para inclusão nas análises de diversidade. Informações sobre os 103 genótipos selecionados podem ser encontradas na Tabela 1.

2.2 Caracteres fenotípicos

Dados de avaliações fenotípicas para os 103 genótipos foram obtidos no banco de dados Caiana do Centro de Cana do IAC. Estes dados são oriundos de diversos experimentos conduzidos em quatro regiões: Ribeirão Preto, Jaú e Piracicaba em São Paulo e Goianésia em Goiás no período entre 1990 e 2009. Todos os experimentos consistiram de delineamento em Blocos ao Acaso com três repetições. Os caracteres avaliados foram: Altura, Diâmetro e Número de Colmos; Percentual de Fibra na Cana (Fibra % Cana); Porcentagem em massa de sacarose (Pol % cana) e Tonelada de Cana por Hectare (TCH). O número de anos em que cada genótipo foi avaliado variou, uma vez que a cada ano novas variedades ou clones eram incluídos enquanto outros eram excluídos do sistema de avaliação. Da mesma forma, nem todos os genótipos foram avaliados em todas as regiões. O número de experimentos também variou de região para região. A Tabela 2 fornece uma caracterização geral das médias dos dados fenotípicos para as cultivares em todos os experimentos analisados.

Tabela 1 - Genótipos de cana-de-açúcar selecionados para compor a população de mapeamento por associação, com seus respectivos parentais

(continua)

Genótipo	Parentais		Genótipo	Parentais	
CTC1	LCP8110	?	IACSP95-3028	SP84-2189	RB85-5156
CTC2	SP82-1176	?	IACSP95-5000	SP84-2066	SP80-185
CTC3	SP76-52	?	IACSP96-2019	SP81-3251	SP84-1192
CTC7	RTT86-18	?	IACSP96-2022	SP84-5019	IAC82-3092
CTC8	SP81-3493	?	IACSP96-2037	SP80-3390	SP83-1483
CTC11	SP84-3478	?	IACSP96-2041	SP80-1842	IAC87-3396
CTC12	SP80-1842	SP81-1747	IACSP96-2042	SP81-5193	SP77-5181
IAC48-65	CP27108	?	IACSP96-2086	SP84-1192	SP83-1483
IAC64-257	Co419	IAC49-131	IACSP96-2100	F175	IAC87-3396
IAC82-2045	IAC65-113	IAC52-150	IACSP96-3057	SP82-6108	?
IAC82-3092	CB4176	IAC68-12	IACSP96-7586	SP84-7017	?
IAC83-4157	IAC68-12	SP70-1143	IACSP97-2000	SP80-185	SP82-3530
IAC86-2210	CP5248	Co798	IACSP97-2025	CTC90222	SP84-2268
IAC86-2480	US71399	?	IACSP97-2053	SP84-2066	IAC86-3154
IAC87-3396	Co740	SP70-1143	IACSP97-2055	SP84-2066	IAC86-3154
IAC91-1099	RB78-5148	?	IACSP97-2109	SP80-185	SP82-3530
IAC91-2195	RB78-5148	?	IACSP97-3354	SP80-185	SP82-3530
IAC91-2205	RB85-5035	?	IACSP97-3402	SP84-2066	IAC86-3154
IAC91-2218	RB85-5035	?	IACSP97-3406	SP84-2066	IAC86-3154
IAC91-3093	SP80-1520	?	IACSP97-6682	SP84-2066	SP84-7017
IAC91-4168	SP81-3137	?	IACSP97-7065	SP84-2066	IAC86-3154
IAC91-5155	SP80-3212	?	IACSP97-7077	SP80-1816	IAC86-2210
IACSP93-1042	SP80-5483	?	IACSP98-2071	IAC91-3093	CTC88-13
IACSP93-2060	RB76-5418	?	IACSP99-2033	SP81-5450	SP88-720
IACSP93-3046	SP79-1011	?	IACSP99-2058	SP80-3280	?
IACSP93-3050	SP79-1011	?	NA56-79	Co419	Co419
IACSP93-6006	SP79-1011	?	POJ2878	POJ2364	EK28
IACSP94-2094	SP84-7017	?	RB72-454	CP5376	?
IACSP94-2101	SP77-5181	RB78-5148	RB83-5054	RB72-454	NA56-79
IACSP94-2180	SP84-2029	SP77-5181	RB83-5486	L6014	?
IACSP94-4002	SP82-6108	SP77-5181	RB84-5210	RB72-454	SP70-1143
IACSP94-4004	SP82-6108	SP77-5181	RB84-5257	RB72-454	SP70-1143
IACSP95-2048	SP82-6108	SP80-3280	RB85-5035	L6014	SP70-1284
IACSP95-2078	SP80-185	?	RB85-5113	SP70-1143	RB72-454
IACSP95-2213	SP80-1079	SP80-3390	RB85-5156	RB72-454	TUC717
IACSP95-2289	SP80-3280	RB83-5486	RB85-5453	TUC717	?
IACSP95-3018	SP84-2189	SP80-1842	RB85-5536	SP70-1143	RB72-454

Tabela 1 - Genótipos de cana-de-açúcar selecionados para compor a população de mapeamento por associação, com seus respectivos parentais

(conclusão)

Genótipo	Parentais		Genótipo	Parentais	
RB86-7515	RB72-454	?	SP83-2847	HJ5741	SP70-1143
RB92-5211	RB85-5206	?	SP84-1201	CO62175	?
RB92-5268	RB85-5511	?	SP84-2025	CO1007	SP71-6180
RB92-5345	H59-1966	?	SP85-3877	H65606	?
RB92-579	nd	nd	SP85-5077	H393633	SP72-1861
RB92-8064	SP70-1143	?	SP86-155	SP78-3081	?
RB93-5744	RB83-5089	RB76-5418	SP86-42	SP70-1143	?
SP70-1143	IAC48-65	?	SP87-344	CO740	SP70-1143
SP71-1406	NA56-79	?	SP87-365	SP77-3067	?
SP79-1011	NA56-79	CO775	SP89-1115	CP731547	?
SP79-2233	H562954	?	SP90-1638	SP78-4601	?
SP80-1816	SP71-1088	H575028	SP90-3414	SP80-1079	SP82-3544
SP80-1842	SP71-1088	H575028	SP91-1049	SP80-3328	SP81-3250
SP80-3280	SP71-1088	H575028	SP91-3011	RB84-5257	?
SP81-3250	CP701547	SP71-1279			

Notas:
nd : Informação não disponível
? : Parental derivado de policruzamento

Tabela 2 - Valores médios dos caracteres fenotípicos nas quatro regiões onde as variedades foram avaliadas. Cada valor representa a média de todos os experimentos conduzidos em cada região

(continua)

Genótipo	Característica					
	Altura (cm)	Diâmetro (cm)	Nº Colmos	Fibra (%)	Pol % Cana	TCH
Mínimo	88,00	1,13	3,68	4,51	6,57	17,46
Média	267,17	2,60	12,30	12,10	14,81	96,13
Máximo	494,00	6,40	33,83	19,67	21,28	244,80
CTC1	271,55	2,99	12,36	11,40	14,36	101,08
CTC2	302,71	2,68	15,67	12,06	15,77	118,64
CTC3	280,84	3,00	15,06	11,25	15,55	131,74
CTC7	284,89	2,79	13,24	11,22	13,98	125,66
CTC8	309,83	2,71	14,87	12,67	15,37	123,22
CTC11	251,67	2,55	22,37	10,46	15,64	160,98
CTC12	248,35	2,64	14,86	13,10	14,32	106,95
IAC48-65	208,75	2,30	21,71	12,11	17,02	95,86
IAC64-257	257,71	3,18	13,11	11,65	14,13	96,24
IAC82-2045	239,06	2,72	11,98	11,57	13,87	92,72
IAC82-3092	281,24	2,51	10,97	11,15	13,80	96,10
IAC83-4157	241,92	2,71	11,71	10,54	14,83	89,93
IAC86-2210	243,03	2,44	11,68	11,98	14,75	85,55
IAC86-2480	226,08	2,59	12,47	11,68	15,09	89,51
IAC87-3396	288,22	2,63	11,74	12,92	15,65	102,55
IAC91-1099	259,99	4,18	14,31	11,90	15,19	118,06
IAC91-2195	249,90	2,42	14,25	11,19	14,58	96,71
IAC91-2205	263,36	2,59	12,70	11,36	15,27	103,2
IAC91-2218	230,84	2,31	16,12	12,11	15,39	91,94
IAC91-3093	198,25	2,26	10,85	13,35	12,94	60,24
IAC91-4168	252,20	2,24	14,67	13,26	14,38	97,47
IAC91-5155	282,43	2,38	14,18	11,60	15,00	94,28
IACSP93-1042	244,78	2,66	11,46	11,70	15,09	89,78
IACSP93-2060	238,85	2,43	13,85	11,42	14,60	89,92
IACSP93-3046	266,93	3,38	13,92	11,56	15,47	121,31
IACSP93-3050	245,60	2,53	13,46	11,36	14,23	105,27
IACSP93-6006	256,65	2,69	12,95	10,89	13,93	102,73
IACSP94-2094	262,71	2,30	15,53	12,08	15,23	106,55
IACSP94-2101	228,49	2,76	13,87	11,55	15,44	112,13
IACSP94-2180	219,09	2,39	12,08	11,80	15,52	66,06
IACSP94-4002	246,90	2,69	11,78	10,94	15,04	91,62
IACSP94-4004	266,98	2,88	13,75	10,82	14,97	129,18
IACSP95-2048	222,13	2,57	10,99	12,09	14,94	76,89
IACSP95-2078	225,84	2,42	14,01	11,68	15,23	87,71

Tabela 2 - Valores médios dos caracteres fenotípicos nas quatro regiões onde as variedades foram avaliadas. Cada valor representa a média de todos os experimentos conduzidos em cada região

(continuação)

Genótipo	Característica					
	Altura (cm)	Diâmetro (cm)	Nº Colmos	Fibra (%)	Pol % Cana	TCH
IACSP95-2213	246,64	2,39	13,48	12,10	15,16	93,35
IACSP95-2289	272,30	3,03	12,30	12,45	14,84	92,18
IACSP95-3018	233,82	2,59	12,56	11,69	16,39	86,87
IACSP95-3028	248,58	2,59	12,51	11,86	15,74	89,18
IACSP95-5000	278,43	2,95	13,02	11,46	15,64	115,41
IACSP96-2019	245,23	2,635	11,47	13,10	13,60	98,88
IACSP96-2022	257,94	2,71	11,37	11,32	13,67	86,17
IACSP96-2037	247,89	2,75	11,24	11,10	15,01	89,12
IACSP96-2041	242,25	2,56	11,72	11,97	15,49	85,01
IACSP96-2042	255,09	2,72	13,78	10,77	15,17	108,13
IACSP96-2086	258,19	2,83	9,40	11,23	13,97	68,12
IACSP96-2100	283,85	2,91	11,47	12,41	12,93	100,34
IACSP96-3057	231,80	2,44	11,39	12,98	14,77	71,56
IACSP96-7586	263,59	2,67	12,21	11,90	14,60	94,61
IACSP97-2000	258,58	2,43	13,96	12,44	14,48	95,49
IACSP97-2025	266,25	3,04	15,11	11,95	15,11	117,97
IACSP97-2053	255,41	3,01	12,22	10,97	13,69	109,38
IACSP97-2055	261,11	2,80	14,35	11,26	14,09	107,68
IACSP97-2109	275,85	2,93	11,31	12,07	13,48	100,94
IACSP97-3354	248,75	2,50	16,50	11,89	13,74	162,21
IACSP97-3402	277,62	2,61	13,00	11,45	12,82	114,09
IACSP97-3406	252,62	3,16	10,73	11,17	13,47	103,52
IACSP97-6682	231,06	2,70	13,42	11,58	15,62	97,92
IACSP97-7065	257,87	2,86	12,12	11,12	13,92	101,38
IACSP97-7077	261,63	2,56	11,80	11,76	14,16	90,92
IACSP98-2071	282,50	2,94	11,85	13,16	15,95	100,73
IACSP99-2033	288,75	2,55	11,70	13,12	16,83	115,18
IACSP99-2058	377,50	2,79	13,91	13,10	14,68	120,52
NA5679	207,25	2,52	19,46	11,15	15,93	122,31
POJ2878	191,00	2,62	16,33	12,00	15,78	84,29
RB72-454	266,36	2,70	11,77	11,57	15,25	99,06
RB83-5054	272,89	2,62	12,34	11,33	13,43	82,38
RB83-5486	261,24	2,60	11,23	11,96	14,57	88,56
RB84-5210	269,20	2,55	14,34	11,96	15,14	106,68
RB84-5257	267,55	2,28	13,46	13,37	16,30	94,49
RB85-5035	301,48	2,50	10,77	12,24	12,52	88,34

Tabela 2 - Valores médios dos caracteres fenotípicos nas quatro regiões onde as variedades foram avaliadas. Cada valor representa a média de todos os experimentos conduzidos em cada região

(conclusão)

Genótipo	Característica					
	Altura (cm)	Diâmetro (cm)	Nº Colmos	Fibra (%)	Pol % Cana	TCH
RB85-5113	267,32	2,53	11,96	12,91	16,96	97,80
RB85-5156	255,69	2,66	13,69	10,84	14,53	106,41
RB85-5453	244,89	2,92	11,74	11,41	13,60	101,76
RB85-5536	296,73	2,67	12,65	12,74	16,02	107,37
RB86-7515	295,29	3,04	11,20	12,22	14,84	111,02
RB92-5211	273,82	2,73	13,25	11,20	13,70	106,46
RB92-5268	284,70	2,75	14,40	11,79	16,23	117,06
RB92-5345	295,65	2,69	12,65	12,31	13,66	106,87
RB92-579	300,53	2,82	12,68	11,34	15,15	119,00
RB92-8064	303,33	2,65	14,11	11,77	15,42	119,70
RB93-5744	323,73	2,89	12,34	11,67	14,96	134,68
SP70-1143	239,22	2,90	16,12	12,61	15,42	90,01
SP71-1406	211,92	2,82	9,61	13,82	14,40	52,08
SP79-1011	237,01	2,34	12,66	12,74	15,41	84,76
SP79-2233	291,67	2,68	14,92	9,84	14,48	149,51
SP80-1816	267,23	2,55	12,67	12,69	16,00	97,65
SP80-1842	280,55	2,44	11,63	12,46	14,61	89,13
SP80-3280	301,13	2,88	12,08	12,25	15,52	114,08
SP81-3250	278,53	2,63	12,67	12,06	14,99	107,78
SP83-2847	301,45	2,85	12,31	12,79	15,64	112,07
SP84-1201	190,00	2,43	22,46	10,56	16,46	128,68
SP84-2025	295,75	2,97	12,68	11,56	14,78	108,67
SP85-3877	242,44	2,72	13,78	10,84	16,43	160,56
SP85-5077	325,25	3,00	13,94	10,47	12,74	117,50
SP86-155	264,24	2,35	15,08	12,05	14,58	103,77
SP86-42	260,19	2,51	13,68	12,41	16,25	94,38
SP87-344	235,08	2,52	11,93	11,61	13,89	91,33
SP87-365	256,24	2,75	15,23	11,92	15,87	114,86
SP89-1115	275,52	3,58	11,24	10,14	13,89	113,73
SP90-1638	210,25	2,47	25,75	11,99	15,88	131,51
SP90-3414	285,10	2,96	17,05	11,64	15,33	155,90
SP91-1049	266,66	2,74	12,06	11,38	13,87	104,74
SP91-3011	238,99	2,32	13,86	11,70	14,30	86,06

2.3 Extração de DNA

O DNA total das plantas foi extraído a partir do meristema apical, conforme metodologia específica para cana-de-açúcar proposta por Al-Janabi, Forget e Dookun (1999). Três gramas de tecido meristemático foram macerados em almofariz contendo 6 mL do tampão de homogeneização (200 mM Tris-HCl pH 8,0; 50 mM EDTA pH 8,0; 2,2 M NaCl; 2% CTAB; 0,06% sulfito de sódio). Em seguida, o extrato foi transferido para tubo de polipropileno de 50 mL tipo Falcon ao qual se adicionou 2 mL de 5 % N-lauril-sarcosina; 2 mL de solução 10% PVP; 2 mL de solução 20% CTAB pré-aquecida. Os volumes dos tubos foram misturados por inversão e incubados a 65°C por 60 min. As amostras foram então resfriadas à temperatura ambiente e um volume igual da solução de fenol:clorofórmio:álcool isoamílico (25:24:1) foi adicionado. A mistura foi agitada manualmente por 2 min e, em seguida, centrifugada a 10.000 *g* por 10 min. A fase aquosa foi transferida para outro tubo e volume igual de isopropanol gelado, acrescido de 2 mL de uma solução 5 M NaCl. Esta mistura foi incubada a –20ºC por algumas horas e, em seguida, o DNA foi "pescado", transferido para tubos de microcentrífuga e lavado duas vezes com 1 mL de etanol a 70%. O *pellet* foi dissolvido em 500 µL de TE (10 mM Tris; 1 mM EDTA, pH 8,0) contendo RNAse (50 µg/mL). Os tubos foram incubados a 37ºC para homogeneização completa do DNA e em seguida foram armazenados a –20ºC. A qualidade do DNA total foi verificada em gel de 0,8% agarose e a concentração foi estimada em fluorômetro DyNA Quant 200 (GE Healthcare BioSciences).

2.4 Dados moleculares

2.4.1 Marcadores SSR

Quarenta e quatro locos microssatélites ou SSR (*Simple Sequence Repeat*) foram utilizados na genotipagem da população experimental. O critério de seleção levou em consideração informações da literatura, indicando uma associação dos locos com caracteres quantitativos ou com base em sua capacidade para a discriminação de variedades em estudos de diversidade. Do total, 31 locos foram selecionados entre os locos gerados pelo Consórcio Internacional de Biotecnologia de Cana-de-açúcar (ICSB) (CORDEIRO et al., 1999; CORDEIRO; TAYLOR; HENRY, 2000), sendo 20 da série 'CIR' e 11 do tipo 'SMC'. Treze locos foram obtidos a partir do banco de seqüências expressas do SUCEST, sendo que quatro ('*ESTB41*', '*ESTB65*', '*ESTB92*' e '*ESTC44*') foram desenvolvidos por Pinto et al. (2004) e

Oliveira et al. (2007, 2009) e nove locos (‘*CV29*’, ‘*CV37*’, ‘*CV38*’, ‘*CV60*’, ‘*CV78*’, ‘*CV79*’, ‘*CV94*’, ‘*CV104*’ e ‘*CV106*’) foram desenvolvidos por Jordão Jr. (2009). Os locos do ICBS foram desenvolvidos a partir bibliotecas genômicas (gSSR), enquanto que os do SUCEST foram obtidos de bibliotecas de seqüências expressas (EST-SSR).

Na extremidade 5’ do iniciador *Forward* de cada loco foi adicionada uma das seguintes caudas fluorescentes: HEX (5’ *GGC TAG GAA AGG TTA GTG GC* 3’), NED (5’ *ACC AAC CTA GGA AAC ACA G* 3’) ou 6-FAM (5’ *GAC TAT GGG CGT GAG TGC AT* 3’), de acordo com Missiaggia e Grattapaglia (2006). As Tabelas 3 e 4 resumem as principais informações e características dos locos selecionados.

As reações de amplificação foram realizadas num volume de 15 µl contendo: 20 ng de DNA, tampão da enzima *Taq* DNA polimerase (75 mM Tris-HCl; 20 mM $(NH_4)_2SO_4$ pH 8,8; 0,01% (v/v) *Tween* 20); 1 U de *Taq* DNA polimerase; 2,0 mM $MgCl_2$; 100 µM de cada dNTP; 0,2 µM de cada iniciador e 0,03 µM de cauda fluorescente. Para evitar amplificações inespecíficas, os locos foram amplificados através de PCR *touchdown*, nas seguintes condições: um ciclo inicial a 94 ºC por 5 min, temperatura de *melting* do iniciador com decréscimo de 1ºC por ciclo até atingir a temperatura de anelamento específica de cada iniciador; a partir desse ponto seguiu-se 30 ciclos a 94 ºC por 45 s, temperatura de anelamento do iniciador por 35 s, 72ºC por 45 s; um ciclo de extensão final de 72 ºC por 5 min. As reações para cada loco foram realizadas separadamente e uma alíquota de 1 µL de reação para um mesmo genótipo amplificado com iniciadores fluorescentes distintos, totalizando 3 µL, foram misturadas a 12 µL de formamida e 0,5 µL do marcador de peso molecular GeneScan ROX-500 (35-500 pb) (Applied Biosystems). Essa mistura foi desnaturada a 94ºC por 5 minutos e mantida a 4ºC. A eletroforese dos produtos amplificados foi realizada em seqüenciador automático ABI PRISM 310 (Applied Biosystem). Os eletroferogramas resultantes foram analisados através dos programas GeneScan 3.7 e Genotyper 3.7 (Applied Biosystems) para a identificação dos alelos em cada genótipo.

Tabela 3 - Relação dos locos microssatélite selecionados, indicando o critério de seleção e os caracteres aos quais se mostraram associados, seus grupos de homologia e/ou co-segregação e os respectivos trabalhos de referências usados para a seleção

(continua)

Loco Microssatélite	Critério de seleção / Característica associada	GH / GC[1]	Referência
CIR1	N°/Altura de colmos / TCH / Pol	II / nl / IV-3	Aitken et al., 2008; Aitken et al., 2005
CIR4	-	VII-1a	Rossi et al., 2003
CIR8	Pol / Brix	I / 19, 42	Aitken et al., 2006; Aitken et al., 2005
CIR10	Pol	VI	Piperidis et al., 2008
CIR14	Pol / Brix	IV	Aitken et al., 2006; Aitken et al., 2005
CIR15	TCH	VI / 71	Aitken et al., 2008; Aitken et al., 2005
CIR18	Brix	VIII	Piperidis et al., 2008
CIR19	Caracterização de germoplasma	-	Souza, 2003
CIR21	Pol / Brix	VII-8	Piperidis et al., 2008
CIR27	N° e Altura de colmos / TCH	IV / 47	Aitken et al., 2005, Aitken et al., 2008
CIR34	Pol / Brix	V / 55, 57	Aitken et al., 2006, Aitken et al., 2005
CIR42	Brix	I-4	Piperidis et al., 2008
CIR43	N° colmos / TCH	I / 27	Aitken et al., 2008; Aitken et al., 2005
CIR44	Ligado ao SMC1825LA	II / 35	Aitken et al., 2006; Aitken et al., 2005
CIR51	N° colmos	X	Jordan et al., 2004; Grivet et al., 1996
CIR56	Caracterização de germoplasma	-	Souza, 2003
CIR64	Pol / Brix	VI	Aitken et al., 2006, Aitken et al., 2005
CIR66	-	VII / 55	Reffay et al., 2005
CIR67	-	VII-1b	Rossi et al., 2003
CIR69	-	VII-1b	Rossi et al., 2003
SMC17CG	Caracterização de germoplasma	-	Brown et al., 2007
SMC36BUQ	Caracterização de germoplasma	-	Souza, 2003
SMC119CG	Caracterização de germoplasma	-	Souza, 2003
SMC179SA	Pol / Brix	I / 50	Aitken et al., 2006; Aitken et al., 2005
SMC221MS	Caracterização de germoplasma	-	Brown et al., 2007
SMC334BS	Caracterização de germoplasma	-	Brown et al., 2007
SMC1047HA	Caracterização de germoplasma	-	Luciana Rossini Pinto, comunicação pessoal
SMC1149HA	Ligado ao CIR34	V / 55	Aitken et al., 2006; Aitken et al., 2005
SMC1608CL	-	III	Aitken et al., 2005
SMC1825LA	Pol / Brix	II / 8, 35	Aitken et al., 2006; Aitken et al., 2005
SMC2055FL	Diâmetro e N° de colmos	III / 28	Aitken et al., 2008; Aitken et al., 2005

Tabela 3 - Relação dos locos microssatélite selecionados, indicando o critério de seleção e os caracteres aos quais se mostraram associados, seus grupos de homologia e/ou co-segregação e os respectivos trabalhos de referências usados para a seleção

(conclusão)

Loco Microssatélite	Critério de seleção / Característica associada	GH / GC[1]	Referência
ESTB41[2]	Ligado ao CIR67	X-4	Oliveira et al., 2009; Oliveira et al., 2007
ESTB65[2]	-	XIV-1b	Oliveira et al., 2009; Oliveira et al., 2007
ESTB92[2]	-	IX-6	Oliveira et al., 2009; Oliveira et al., 2007
ESTC44[2]	-	XIV-1b	Oliveira et al., 2009; Oliveira et al., 2007
CV29	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV37	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV38	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV60	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV78	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV79	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV94	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV104	Caracterização de germoplasma	-	Jordão Júnior, 2009
CV106	Caracterização de germoplasma	-	Jordão Júnior, 2009

Notas:

1 - GH: Grupode Homologia; GC: Grupo de Co-segregação; nl: não ligado

2 - ESTB41 (Proteína hipotética);
ESTB65 (Provável proteína ribossomal L32)
ESTB92 (Provável transportador de açúcar)
ESTC44 (B1066G12.15)
Informações obtidas de Oliveira et al. (2009)

Tabela 4 - Sequências dos iniciadores dos locos microssatélites com as respectivas caudas fluorescentes, temperaturas de anelamento e tamanhos observados dos alelos neste trabalho

(continua)

Loco SSR	Iniciador Forward (5' - 3')	Iniciador *Reverse* (5' - 3')	Temp. anel. (°C)	Fragmentos observ. (pb)
CIR1	HEX/CTTGTGGATTGGATTGGAT	AGGAAATGGATTGCTCAGG	55	147 - 203
CIR4	NED/TTCCAGCAGCAGCATCAAT	CCCACTAGGAGAAGCAATAACT	55	259 - 278
CIR8	NED/ACCAGCACACGGAAACGGAAC	ATGCGCAGGGATAGGATAGGAG	55	217 - 303
CIR10	6-FAM/ACACCACTCACATCCACTTG	TGATACACCATTGTTGATGC	55	185 - 204
CIR14	NED/GATTGTTTTTCCCCCACTA	CACCTTGTTCTTGCTTTACTC	55	223 - 265
CIR15	NED/CTTGGACCCGTTCTTGGATG	AGCACTGAGGCGACTTACCC	55	183 - 359
CIR18	HEX/GGGTGTTCTGTTGAGCA	GAGGTAGGAGGGAGTGTT	55	177 - 234
CIR19	6-FAM/GGTTCCAAAATACACAAA	CAATCTTATCTACGCACTT	55	148 - 203
CIR21	6-FAM/CGCCAGCCACATAAAAGG	CGACCAGGAGTTCATCAA	55	138 - 194
CIR27	HEX/CCCTATCACTGTTCTTCAT	ATTTGTCCTTGCGAGAG	55	195 - 232
CIR34	6-FAM/ATCGCCTCCACTAAATAAT	TTGTCTTTGCTTCCTCCTC	54	132 - 186
CIR42	NED/GTCTCTTGGGTGGCGTAT	TTTTTTGTTTTTCTCTTTTG	54	178 - 240
CIR43	NED/ATTCAACGATTTTCACGAG	AACCTAGCAATTTACAAGAG	55	137 - 269
CIR44	HEX/TCCCTCCTCATCACTCTG	AAAATAAGCACCAAAAGC	55	224 - 238
CIR51	6-FAM/GCAAGAAGAACCAACAGTAG	GTATTTACCGCAACATTGAT	55	250 - 301
CIR56	6-FAM/ATTTGACGCTACGATGGTG	ATCCGTTTTTCAGCAGAGC	55	143 - 204
CIR64	NED/ATTGGATTCCTTCTGCTA	CATCACAACAGGTTTCAG	55	132 - 280
CIR66	HEX/AGGTGATTTAGCAGCATA	CACAAATAAACCCAATGA	55	147 - 165
CIR67	6-FAM/GACTGAGTATTGGCTTGG	GCTTCTTCTTTGGTCGTT	55	107 - 257
CIR69	HEX/GATTGGATTTGTGATGTAA	CCTTCCTGATTTCTGATT	54	164 - 172
SMC17CG	HEX/AAGGTAGCACGAAACACGTCGAT	AACCCCAGCGGCGATCTCT	55	125 - 200
SMC36BUQ	6-FAM/GGGTTTCATCTCTAGCCTACC	TCAGTAGCAGAGTCAGACGCTT	50	112 - 270
SMC119CG	6-FAM/TTCATCTCTAGCCTACCCCAA	AGCAGCCATTTACCCAGGA	50	107 - 174
SMC179SA	HEX/CATTTGACCAACCATGCACAGC	GGCTTGGCAGGATTGGAAAC	55	121 - 180
SMC221MS	HEX/CATGCCAACTTCCAAACAGACT	GGTGATGCGAAGAGATTGGA	50	129 - 173
SMC334BS	HEX/CAATTCTGACCGTGCAAAGAT	CGATGAGCTTGATTGCGAATG	50	120 - 183
SMC1047HA	6-FAM/TGAGCCTAAGCCAGAAAGAAG	GGAACTAATTTCCTACGAGAACAC	50	109 - 187
SMC1149HA	HEX/CAGCATTGATGTTGCTCAGG	CACTTCCATACACAGTTCCTGC	50	110 - 264
SMC1608CL	HEX/ACTGTAGGTTCACCGAAGCAAGAG	CAGCCACTTTCATCCACAGAGG	50	153 - 185

Tabela 4 - Sequências dos iniciadores dos locos microssatélites com as respectivas caudas fluorescentes, temperaturas de anelamento e tamanhos observados dos alelos neste trabalho

(conclusão)

Loco SSR	Iniciador Forward (5' - 3')	Iniciador *Reverse* (5' - 3')	Temp. anel. (°C)	Fragmentos observ. (pb)
SMC1825LA	6-FAM/CACGTCCTTCCGCCTTGA	TCATCGTTCGTCGCACTG	50	113 - 124
SMC2055FL	6-FAM/TGAGCACATGGTTCTGCAAG	TCCACTCCCTCACCATCAG	55	262 - 337
ESTB41	6-FAM/CATGGAGAGCTGGGCGACCTG	GGCGGCGGCGAGGATGA	63	142 - 179
ESTB65	HEX/CGATCCCAGGCCAGTTCAT	CTTGTTCGCTACTGCTACCGTGTT	57	205 - 254
ESTB92	NED/TCTGAATGGATGTCGCCCTGTG	TTTGCGGGCTTCTCTGCTTTCT	59	222 - 258
ESTC44	6-FAM/CATGGTGCTGTGGGGTTCTG	GGTGGGCAAAGCTGTCCTGA	63	110 - 175
CV29	6-FAM/TCGCGTCCACCAATGTAACC	GCGTGCATCGCTTGTGTCTT	64	103 - 157
CV37	HEX/GGATGGACGACGTGTCCTGG	ATAAAGTGGCCGCTTGGATTGA	64	115 - 188
CV38	NED/GAAGCAGGGGCCTCAAGTTG	GTCAAACAGGCGATCTGGCTC	64	121 - 205
CV60	HEX/AATCTGCACCCTGCCCTCTC	CAGCTGGAGCATGGATGGAG	64	155 - 201
CV78	6-FAM/ACGAGGCCACCATAGAACATG	GCAATTGGGAGGAGAGGAATG	64	138 - 224
CV79	HEX/GGCACTGCTGGTGGTTGATTG	TCCCACATCAAGAGGCAGCTA	64	137 - 196
CV94	6-FAM/GGCAGGCCAAGATGAATGAAG	AGCACAGCGGAGGGTACGG	64	180 - 221
CV104	6-FAM/GATTTTCGACTGTGCGCTTGG	AAGTTCTCTGCCGGAGCAAAC	64	151 - 177
CV106	6-FAM/AAACAGAGCATACTCGAGGCC	ACGTTGCTGACGAGGTTTTCC	64	149 - 176

2.4.2 Marcadores AFLP

2.4.2.1 *Restrição do DNA*

Na etapa de restrição 400 ng de DNA total foram digeridos numa reação de 25 µL de volume final contendo 2,5 U de *Eco*RI (Invitrogen); 3,0 U de *Mse*I (New England BioLabs); tampão de reação *NEBuffer* 3 (10 mM NaCl; 5 mM Tris-HCl; 1mM $MgCl_2$; 0,1 mM DTT; pH 7,9) (New England BioLabs). A reação foi realizada em termociclador a 37ºC por 4 h, seguida de uma etapa de inativação das enzimas a 65ºC por 20 min.

2.4.2.2 *Ligação dos adaptadores*

Antes da reação de ligação foram preparadas soluções contendo adaptadores para as extremidades coesivas de suas respectivas enzimas (*Eco*RI *Forward*: 5´ CTC GTA GAC TGC GTA CC 3´; *Eco*RI *Reverse*: 5´ AAT TGG TAC GCA GTC TAC 3´; *Mse*I *Forward*: 5´ GAC GAT GAG TCC TGA G 3´; *Mse*I *Reverse*: 5´ TAC TCA GGA CTC AT 3´). Para a enzima *Eco*RI foi preparada uma solução contendo 5 µM de cada adaptador e para *Mse*I a concentração foi de 50 µM. Antes da reação de ligação as soluções foram tratadas em termociclador a 65ºC por 10 min, 37ºC por 10 min e 25ºC por 10 min e mantidas a 4ºC para uso imediato na reação de ligação.

A reação de ligação foi realizada num volume final de 30 µL contendo 15 µL da reação de restrição; 1x de tampão para a enzima T4 DNA Ligase (50 mM Tris-HCl; 10 mM $MgCl_2$; 10 mM DTT; 1,0 mM ATP); 0,25 µM de adaptadores *Eco*RI; 2,5 µM de adaptadores *Mse*I e 1 U de enzima T4 DNA Ligase (New England BioLabs). A ligação foi realizada em termociclador a 16ºC por 16 h e finalizada por inativação da enzima a 65ºC por 20 min. Ao final a reação foi armazenada a -20ºC.

2.4.2.3 *Reação de Pré-amplificação*

Nesta reação foram utilizados iniciadores complementares aos adaptadores de *Eco*RI e *Mse*I porém, tais iniciadores não continham nucleotídeos seletivos na extremidade 3´ (*Eco*RI+0: 5´ GAC TGC GTA CCA ATT C 3´; *Mse*I+0: 5´ GAT GAG TCC TGA GTA A 3´). A reação foi realizada num volume final de 15 µL contendo 2 µL da reação de ligação diluída seis vezes; tampão da enzima *Taq* DNA Polimerase (75 mM Tris-HCl; 20 mM $(NH_4)_2SO_4$ pH 8,8; 0,01%

(v/v) *Tween* 20); 2,5 U de enzima *Taq* DNA Polimerase (Fermentas); 2 mM de $MgCl_2$; 0,5 mM de cada dNTP; 3,3 µM de cada iniciador *Eco*RI e *Mse*I. A reação de PCR foi realizada de acordo com o seguinte programa: um primeiro ciclo a 94ºC por 2 min; seguido de 29 ciclos a 94ºC por 1 min; 60ºC por 30 s e 72ºC por 1 min; finalizando com um ciclo a 72ºC por 5 min. Os produtos desta reação foram diluídos 10x em água ultrapura e armazenados a -20ºC. Uma alíquota de 5 µL foi utilizada para averiguar a qualidade da reação em gel de 1% agarose.

2.4.2.4 *Reação de Amplificação Seletiva*

Os iniciadores utilizados nesta reação continham três nucleotídeos seletivos na extremidade 3´ do iniciador. Os iniciadores de *Eco*RI (5´ *GAC TGC GTA CCA ATT C* 3´) foram: E-*ACG*, E-*AAG*, E-*AAA*, E-*AGA*, E-*AGC*, E-*AGG*, E-*AAC*, E-*ACA*; os de *Mse*I (5´ *GAT GAG TCC TGA GTA A* 3´) foram: M-*CGT*, M-*CGC*, M-*CTT*, M-*CTG*, M-*CTC*, M-*CAC*, M-*CAT*, M-*CAG*, M-*CTA*. Dois iniciadores de *Eco*RI (E-*ACG* e E-*AAG*) e dois de *Mse*I (M-*CAT* e M-*CAG*) foram marcados com os fluoróforos *IRDye* 700 e *IRDye* 800, respectivamente. A reação de amplificação seletiva foi feita num volume final de 10 µL contendo 2 µL da reação de pré-amplificação diluída 10x; tampão da enzima *Taq* DNA Polimerase (75 mM Tris-HCl; 20 mM $(NH_4)_2SO_4$ pH 8,8; 0,01% (v/v) *Tween* 20); 1 U de enzima *Taq* DNA Polimerase (Fermentas); 2 mM de $MgCl_2$; 0,2 mM de cada dNTP; 0,05 µM do iniciador de *Eco*RI e 0,3 µM do iniciador de *Mse*I. As reações para as diversas combinações de iniciadores *Eco*RI e *Mse*I foram realizadas de modo a conter sempre um iniciador marcado e outro não marcado. Um total de 18 combinações com os iniciadores acima foram usadas na genotipagem da população (Tabela 5). A reação de PCR ocorreu nas seguintes condições: um ciclo inicial a 94ºC por 2 min; seguido de 13 ciclos a 94ºC por 30 s; 65ºC por 30 s (com decréscimo de 0,7ºC por ciclo) e 72 ºC por 1 min; mais 23 ciclos a 94ºC por 30 s; 56ºC por 30 s e 72ºC por 1 min e finalizando com um ciclo a 72ºC por 2 min.

Tabela 5 - Combinações de iniciadores AFLP usados na genotipagem da população. Estão indicados apenas os três nucleotídeos seletivos posicionados na extremidade 3´ dos iniciadores *Eco*RI e *Mse*I. Os iniciadores marcados estão indicados com seus respectivos fluoróforos

N° da Combinação	Seqüência 3´ do iniciador *EcoRI*	Seqüência 3´ do iniciador *MseI*	N° da Combinação	Seqüência 3´ do iniciador *EcoRI*	Seqüência 3´ do iniciador *MseI*
1	ACG-IRD700	CGT	10	AAG-IRD800	CAC
2	ACG-IRD700	CGC	11	AAG-IRD800	CTC
3	ACG-IRD700	CTT	12	AAG-IRD800	CAG
4	ACG-IRD700	CTC	13	AAA	CAT-IRD700
5	ACG-IRD700	CTA	14	AGA	CAT-IRD700
6	ACG-IRD700	CAT	15	AAA	CAG-IRD800
7	AAG-IRD800	CGT	16	ACA	CAG-IRD800
8	AAG-IRD800	CGC	17	AGG	CAG-IRD800
9	AAG-IRD800	CTG	18	AGA	CAG-IRD800

Notas:
Seqüências não marcadas são indicadas apenas pelos três nucleotídeos
IRD700: marcado com fluoróforo IRDye 700
IRD800: marcado com fluoróforo IRDye 800

2.4.2.5 *Análise dos fragmentos AFLP*

Os produtos das reações de amplificação seletiva foram analisados no analisador automático de DNA Li-cor modelo 4300 (LI-COR Biosciences). Para tanto, foram feitas misturas numa proporção de 1:1,5 µL de reação do iniciador fluorescente *IRDye* 700 e *IRDye* 800, respectivamente. A um volume de 1,5 µL desta mistura foi adicionado 2 µL de azul de bromofenol contendo 90% de formamida. Em seguida, esta mistura foi desnaturada em termociclador a 95°C por 5 min e mantida a 4°C. Um volume de 0,6 µL de cada amostra foi aplicado em gel de 6,5% poliacrilamida pré-aquecido por 30 min. Foi aplicado também 0,8 µL de marcador de peso molecular contendo fragmentos de 50 a 700 pb marcados com as fluorescências *IRDye* 700 e *IRDye* 800 (LI-COR Biosciences). Durante a eletroforese imagens dos géis em formato eletrônico foram automaticamente geradas. Estas imagens foram posteriormente analisadas para obtenção das marcas AFLP através do programa SAGA^{MX} (LI-COR Biosciences).

2.5 Análise dos dados

A natureza complexa do genoma da cana-de-açúcar com sua alta poliploidia e ocorrência concomitante de aneuploidia torna difícil determinar a dosagem em que um alelo está presente no genoma. Desta forma, independentemente da natureza do marcador, em cana, todos são tratados como dominantes. Portanto, para cada alelo somente dois padrões podem observados nos indivíduos da população: presença ou ausência. Neste trabalho foram utilizados dois sistemas distintos para a genotipagem da população de modo que, no caso dos AFLPs os alelos foram visualizados como bandas e no caso dos microssatélites como picos. Para facilitar a compreensão ao longo do texto, as bandas e picos serão referidos de forma comum como marcas ou alelos. Por fim, todos os alelos assim obtidos foram reunidos numa matriz de dados binária onde a presença, representando o alelo dominante, foi codificada como "1" e a ausência, representando o alelo nulo, foi codificada como "0". Foram excluídos alelos com freqüência maior que 0,95 e menor que 0,05. Desta forma, um total de 850 alelos, sendo 370 SSR e 480 AFLP foi obtido.

2.5.1 Análises de diversidade

2.5.1.1 *Polimorfismo e Diversidade alélica*

Como medida do nível de informação das marcas AFLP e SSR calculou-se o Conteúdo de Informação Polimórfica (PIC) conforme descrito por Vuylsteke (2000):

$$PIC_i = 1 - \sum_{j=1}^{n} p_{ij}^2$$

onde p_{ij} é a freqüência do *j-ésimo* padrão do alelo *i* e *n* é o número de padrões para cada alelo marcador que, neste caso, assume apenas dois valores (0 e 1). As freqüências dos padrões de cada alelo foram obtidas por meio do programa AFLP-SURV (VEKEMANS, 2002). Para permitir uma comparação da eficiência dos marcadores usados em detectar polimorfismos na população, foi calculada uma média dos valores de PIC anteriores, conforme a expressão:

$$PIC_m = \sum_{i=1}^{N} PIC_i/\mathrm{N}$$

onde PIC_i é o PIC do *i-ésimo* alelo e *N* é o número de alelos originados de um mesmo loco SSR ou de uma mesma combinação de iniciadores AFLP.

2.5.1.2 *Análise de similaridade genética*

Com o objetivo de explorar as relações genéticas entre os genótipos, estimativas de similaridade genética baseadas no coeficiente de Dice (NEI; LI, 1979) foram geradas de acordo com a expressão:

$$SG_{ij} = \frac{2a}{2a+b+c} \quad \text{onde,}$$

SG_{ij}: Similaridade genética entre os genótipos *i* e *j*;

a: número de alelos comuns aos genótipos *i* e *j*;

b: número de alelos presentes no genótipo *i* e ausentes no genótipo *j*;

c: número de alelos ausentes no genótipo *i* e presentes no genótipo *j*;

Coeficientes de dissimilaridade foram obtidos ($DG_{ij} = 1 - SG_{ij}$) a partir da expressão anterior e usados na construção de um dendrograma pelo método *Neighbor-Joining* modificado conforme Gascuel (1997). Análise de *bootstrap* com 2.000 repetições foi realizada para testar a robustez com que os grupos foram formados. O grau de correlação entre e as distâncias representadas no dendrograma e a matriz de dissimilaridade original foi avaliado calculando-se uma matriz de coeficientes de correlação cofenéticos (r_{cof}) e comparando esta matriz com a matriz de dissimilaridade original. Nesta análise foram incluídos os três genótipos selvagens ‘Badilla de Java’ (*Saccharum officinarum*), ‘SES 205A’ (*Saccharum spontaneum*) e ‘Chunnee’ (*Saccharum barberi*). Todas as análises foram realizadas por meio dos programas Darwin 5.0 (PERRIER; JACQUEMOUD-COLLER, 2006) e MEGA 5.0 (TAMURA et al., 2011).

2.5.2 Desequilíbrio de ligação

O desequilíbrio de ligação foi avaliado através da medida r^2 que representa o coeficiente de correlação entre dois alelos ou marcas. Para tanto, os alelos foram organizados na forma de uma tabela de contingência, que foi usada também para obtenção de estimativas não viesadas dos *p* valores através do teste exato de Fisher com 1.000 permutações.

Foram obtidos coeficientes de correlação r^2 para dois conjuntos de marcas. O primeiro conjunto consistiu de marcas oriundas de um mapa genético de ligação desenvolvido no Laboratório de Melhoramento de Plantas do CENA/USP, tendo como base o mapa previamente publicado por Garcia et al. (2006) para uma população de 100 indivíduos do cruzamento SP80-180 x SP80-4966 (cultivares pré-comerciais pertencentes ao programa de melhoramento do Centro de Tecnologia Canavieira). No entanto, o mapa desenvolvido para este trabalho incluiu 719 marcas AFLP (KIDO, 2003), 209 SSR (SOUZA, 2003) e 197 RFLP (ALBINO, 2006), desenvolvidos no Laboratório de Melhoramento de Plantas do CENA/USP, e mais 798 marcas TRAP desenvolvidas no IAC por Creste (Dados não publicados), totalizando 1923 marcas. A população de mapeamento consistiu de 92 indivíduos da população original, os quais estavam disponíveis no Laboratório de Melhoramento do CENA/USP. Utilizou-se o programa JoinMap 3.0 (VAN OOIJEN; VOORRIPS, 2001) para a obtenção do mapa de ligação. As marcas foram agrupadas usando-se LOD mínimo igual a 5 e fração máxima de recombinação igual a 0,35. A função de mapeamento de Kosambi foi utilizada para conversão das freqüências de recombinação em distâncias de mapa (KOSAMBI, 1944).

O mapa obtido foi formado por 185 grupos de co-segregação, dos quais 24 grupos entre os de maior densidade de marcas foram selecionados para as análises de desequilíbrio de ligação (Anexo A). Estes 24 grupos totalizaram 243 marcas, incluindo os quatro tipos de marcadores usados (AFLP, RFLP, SSR, TRAP). A análise com estas marcas do mapa de ligação teve como objetivo caracterizar a extensão dos níveis de desequilíbrio no genoma de cana e, para tanto, os valores de r^2 foram visualizados num gráfico em função da distância genética em Centimorgan (cM). Uma curva de regressão foi ajustada a partir dos dados usando a função LOESS (BRESEGHELLO; SORRELS, 2006; ROBBINS et al., 2011). Foi determinado um nível crítico de r^2 a partir do qual o desequilíbrio foi declarado como significativo e, a princípio, é resultante apenas de ligação física. Para tanto, utilizou-se as marcas localizadas em grupos de co-

segregação diferentes (marcas não ligadas) para calcular uma distribuição de valores de r^2, tomando-se o valor referente ao percentil 97,5% desta distribuição como limite crítico. Os coeficientes de correlação das marcas não ligadas foram incluídos também no mesmo gráfico onde foram plotados os coeficientes das marcas ligadas. A significância das correlações entre as marcas foi avaliada após correção erro tipo I para múltiplos testes pelo método de Bonferroni.

O segundo conjunto de marcas usadas para análise do desequilíbrio de ligação consistiu de 405 marcas selecionadas entre as 850 desenvolvidas neste trabalho. A seleção destas marcas está justificada e descrita no item 2.5.3. O objetivo com esse procedimento foi comparar os níveis de desequilíbrio entre estas 405 marcas na população de mapeamento por associação com os níveis observados entre as marcas ligadas e não ligadas do mapa genético de ligação. Os cálculos dos coeficientes r^2 e dos *p* valores foram os mesmos descritos anteriormente.

As análises foram realizadas através dos programas PowerMarker 3.0 (LIU; MUSE, 2005) e R (R DEVELOPMENT CORE TEAM, 2010).

2.5.3 Análise de estruturação da população

A presença de subgrupos ou estruturação na população foi avaliada inicialmente utilizando um método multivariado baseado em Análise de Componentes Principais (PCoA). Para tanto, a matriz de dissimilaridade original (*d*) precisou ser transformada numa matriz de distâncias Euclideanas (*d'*), o que foi feito estimando-se um índice *p* (índice Euclideano) de modo que $d' = d^p$. As análises foram realizadas através do programa Darwin 5.0 (PERRIER; JACQUEMOUD-COLLER, 2006).

Uma segunda abordagem baseada em estatística Bayesiana e disponível através do programa STRUCTURE (PRITCHARD; STEPHENS; DONNELLY, 2000b; FALUSH; STEPHENS; PRITCHARD, 2003, 2007; HUBISZ et al., 2009) foi utilizada. Seguindo recomendação de Falush, Stephens e Pritchard (2003) foi realizada uma seleção prévia entre as 850 marcas SSR e AFLP, excluindo-se aquelas com forte ligação entre si. Uma vez que a maioria dos mapas genéticos de ligação foram construídos adotando-se o limite de 35 cM para as frações de recombinação entre marcas e com base em alguns dos principais mapas que usaram os mesmos marcadores SSR e AFLP deste trabalho (AITKEN; JACKSON; McINTYRE, 2005; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2007), estabeleceu-se o limite de 35

cM para declarar duas marcas como fortemente ligadas. Este valor está alinhado também com os resultados de Raboin et al. (2008) que observaram alto nível de desequilíbrio a distâncias de até 40 cM.

Na determinação dos parâmetros de análise no STRUCTURE, nenhuma informação *a priori* sobre as cultivares foi usada para a determinação dos subgrupos, ou seja, a princípio elas não pertenciam a nenhum grupo conhecido. Os genótipos foram tratados como haplóides e avaliou-se a hipótese de até sete subgrupos, ou seja, K variou de 1 a 7, com 10 repetições para cada K estimado. O modelo de ancestralidade adotado foi o de mistura (*Admixture model*), ou seja, cada genótipo pode ter uma fração de seu genoma oriundo em um ou mais dos grupos identificados. Considerou-se o modelo de freqüências alélicas correlacionadas baseado no fato de que há uma ancestralidade comum para muitos dos genótipos da população. Os dados foram coletados com 200.000 repetições MCMC (*Monte Carlo Markov Chain*) após um período de 100.000 repetições iniciais (*burnin*). O método de Evanno, Regnault e Goudet (2005) foi utilizado para determinar o número ótimo de subgrupos (K) na população através do programa STRUCTURE HARVESTER v. 0.6 (EARL, 2011). A partir destes resultados foi obtida a matriz de estrutura populacional (matriz Q) que contém a proporção do genoma de cada indivíduo com origem em algum dos K subgrupos identificados.

2.5.4 Coeficientes de coancestralidade

Duas estratégias distintas foram usadas para determinar as relações genealógicas entre os genótipos. Na primeira, o programa SPAGeDI (HARDY; VEKEMANS, 2002) foi utilizado para calcular a matriz de coancestralidade (matriz K ou *Kinship*) de acordo com um método específico para marcadores dominantes proposto por Hardy (2003). As estimativas dos coeficientes de coancestralidade entre pares de indivíduos foram calculadas segundo a expressão:

$$F_{ij} = (Q_{ij} - Q_m)/(1 - Q_m)$$

onde *Qij* é a probabilidade de dois genes tomados ao acaso nos indivíduos *i* e *j* serem idênticos por estado (IBS) e *Qm* é a probabilidade média de desses mesmos genes serem idênticos por

estado, tendo sido amostrados em dois indivíduos tomados ao acaso na população de onde *i* e *j* foram retirados. Para computar as estimativas de F_{ij} considerou-se o coeficiente de endogamia em cada individuo igual a zero (*f* = 0).

Alternativamente, foi calculado o índice de similaridade genética de Gower (GOWER, 1971) com base na observação de que este índice corresponde a uma medida da proporção de alelos idênticos por estado (IBS) (WESSEL; SCHORK, 2006; BUIL et al., 2009). O índice de Gower tem a seguinte expressão:

$$Gower_{ij} = \frac{a}{a+b+c+d} \quad \text{onde,}$$

$Gower_{\mathrm{ij}}$: Similaridade genética entre os genótipos *i* e *j*;

a: número de alelos comuns aos genótipos *i* e *j*;

b: número de alelos presentes no genótipo *i* e ausentes no genótipo *j*;

c: número de alelos ausentes no genótipo *i* e presentes no genótipo *j*;

d: número de alelos ausentes nos genótipos *i* e *j*;

2.5.5 Análise dos dados fenotípicos

Os dados fenotípicos foram previamente analisados utilizando uma abordagem de modelos mistos, através do pacote estatístico *lme4* (BATES; MAECHLER, 2010) do programa R (R DEVELOPMENT CORE TEAM, 2010). O modelo testado considerou as regiões como efeitos fixos e anos e variedades como efeitos aleatórios. Interações entre anos e regiões e entre regiões e variedades foram incluídas no modelo. Com base nos resultados obtidos e a partir dos dados fenotípicos observados foram calculados BLUPs (*Best Linear Unbiased Predictors*) das variedades em cada região onde elas foram avaliadas e também uma estimativa geral para todas as regiões, os quais foram usados no modelo de associação.

2.5.6 Análise de associação

As análises de associação foram realizadas no programa TASSEL (BRADBURY et al., 2007), utilizando o método proposto por Yu et al. (2006). Através dos marcadores foram obtidas as matrizes Q e K conforme descrito nos itens 2.5.3 e 2.5.4, respectivamente. Seus efeitos foram incluídos como covariáveis para ajustar um modelo de associação capaz de explicar associações entre os marcadores e os caracteres fenotípicos. A equação do modelo misto tem a seguinte forma:

$$y = X\beta + S\alpha + Qv + Zu + e$$

Neste modelo y é o vetor de observações fenotípicas; β é um vetor de efeitos fixos (outros efeitos diferentes dos marcadores e estrutura populacional); α é um vetor de efeitos fixos dos marcadores; v é um vetor de efeitos fixos de estrutura da população; u é um vetor de efeitos poligênicos aleatórios desconhecidos e e é o vetor aleatório de efeitos residuais; Q é a matriz do STRUCTURE relacionando y a v; X, S e Z são matrizes de incidência (compostas por zeros e uns), relacionando y a β, α e u, respectivamente.

Assumiu-se que os vetores de efeitos aleatórios u e e são normalmente distribuídos com média zero e variâncias: $Var(u) = \mathrm{K}\sigma_g^2$, sendo σ_g^2 a variância genética e K uma matriz (n x n) contendo os coeficientes de coancestralidade calculados no SPAGeDI ou as proporções de alelos idênticos por estado (IBS) baseadas no índice de Gower; $Var(e) = \mathrm{I}\sigma_e^2$, onde σ_e^2 é a variância residual e I é uma matriz identidade.

Com o objetivo de testar o impacto de cada um dos efeitos de estruturação no poder do modelo em detectar associações, foram testadas diversas variações do modelo de associação. Como havia duas matrizes K diferentes (K_1 para a matriz calculada pelo programa SPAGeDI, e K_2 obtida através do índice de Gower) foram testadas, inicialmente, duas variações do modelo básico: Q + K_1 e Q + K_2. Em seguida foram removidos os efeitos de estruturação da matriz Q, mantenho apenas os efeitos de estruturação resultantes das relações de parentesco, modelos K_1 e K_2. Portanto, foram testadas quatro variações do modelo de associação.

Utilizou-se o método FDR (*False Discovery Rate*) para controle do erro tipo I associado a múltiplos testes e para declarar um marcador como significativamente associado a uma característica fenotípica (BENJAMINI; HOCHBERG, 1995). O valor FDR (q*) representa a taxa de falsos positivos entre testes significativos, ou seja, é a taxa de associações do marcador com o caráter, declaradas significativas, quando na realidade, não há associação. O teste foi realizado do seguinte modo: considerou-se inicialmente as hipóteses para m testes como sendo $H_1, H_2, H_3, \ . \ . \ . \ , H_m$; cujos respectivos p valores são $P_1, P_2, P_3, \ . \ . \ . \ , P_m$. Os p valores foram então organizados em ordem crescente de modo que $P_1 \leq P_2 \leq \ . \ . \ . \leq P_m$. Dessa forma, haverá uma proporção de hipóteses rejeitadas $H_1, H_2, H_3, \ . \ . \ . \ , H_k$; quando k for o maior valor de i que satisfaça a seguinte condição: $P_i \leq (i/m) \times q^*$. Os p valores dos testes foram obtidos no programa TASSEL.

3 RESULTADOS E DISCUSSÃO

3.1 Seleção dos genótipos para compor a população de mapeamento

Em estudos de associação, a escolha da população experimental é bem mais flexível quando comparada a estudos de mapeamento de ligação. Neste trabalho, o principal critério de seleção se baseou nos dados fenotípicos para as características avaliadas, de modo que houvesse o máximo possível de variabilidade fenotípica entre os indivíduos selecionados (Figura 1). Deste modo, os 103 genótipos que compõem a população de mapeamento incluem cultivares que são ou foram utilizadas comercialmente em áreas de produção importantes ou que possuem grande importância como parentais de variedades recentes e clones em desenvolvimento no programa de melhoramento do IAC. Além destes, três genótipos selvagens foram incluídos em algumas das análises realizadas: 'Badilla de Java' (*Saccharum officinarum*), 'Chunnee' (*Saccharum barberi*) e 'SES 205A' (*Saccharum spontaneum*). Na Figura 1 encontram-se os histogramas com as distribuições de freqüência para os caracteres fenotípicos de interesse nesse estudo. Observa-se que todos os caracteres apresentaram distribuição normal dos dados e mostraram também considerável variabilidade na população.

3.2 Análises de diversidade

Os 44 locos SSR utilizados revelaram um total de 370 alelos polimórficos com média de 8,4 alelos por loco (Tabela 6). O número de alelos variou de 2 (locos *CIR69* e *SMC1825LA*) a 17 (loco *CV78*). A partir das 18 combinações de iniciadores AFLP foram obtidos um total de 480 alelos polimórficos (Tabela 7). O número de alelos para cada combinação variou de 14 (*AAA_CAT*) a 43 (*ACG_CTT*) com média de 26,7 por combinação. A análise da distribuição do número de alelos presentes nos genótipos fornece uma medida indireta do nível de heterozigosidade e de diversidade presente na população (AITKEN et al., 2006). Deste modo, quando se analisou o número de alelos presentes nos genótipos cultivados, considerando-se cada tipo de marcador individualmente (SSR ou AFLP) e em conjunto (Total), foi observada grande variação nas distribuições desses valores (Figura 2). Entre os 103 genótipos cultivados e considerando o total de marcadores, observou-se uma média de 381 alelos por genótipo com mínimo de 284 para 'IAC86-2480' e máximo de 469 para 'IACSP96-2041'.

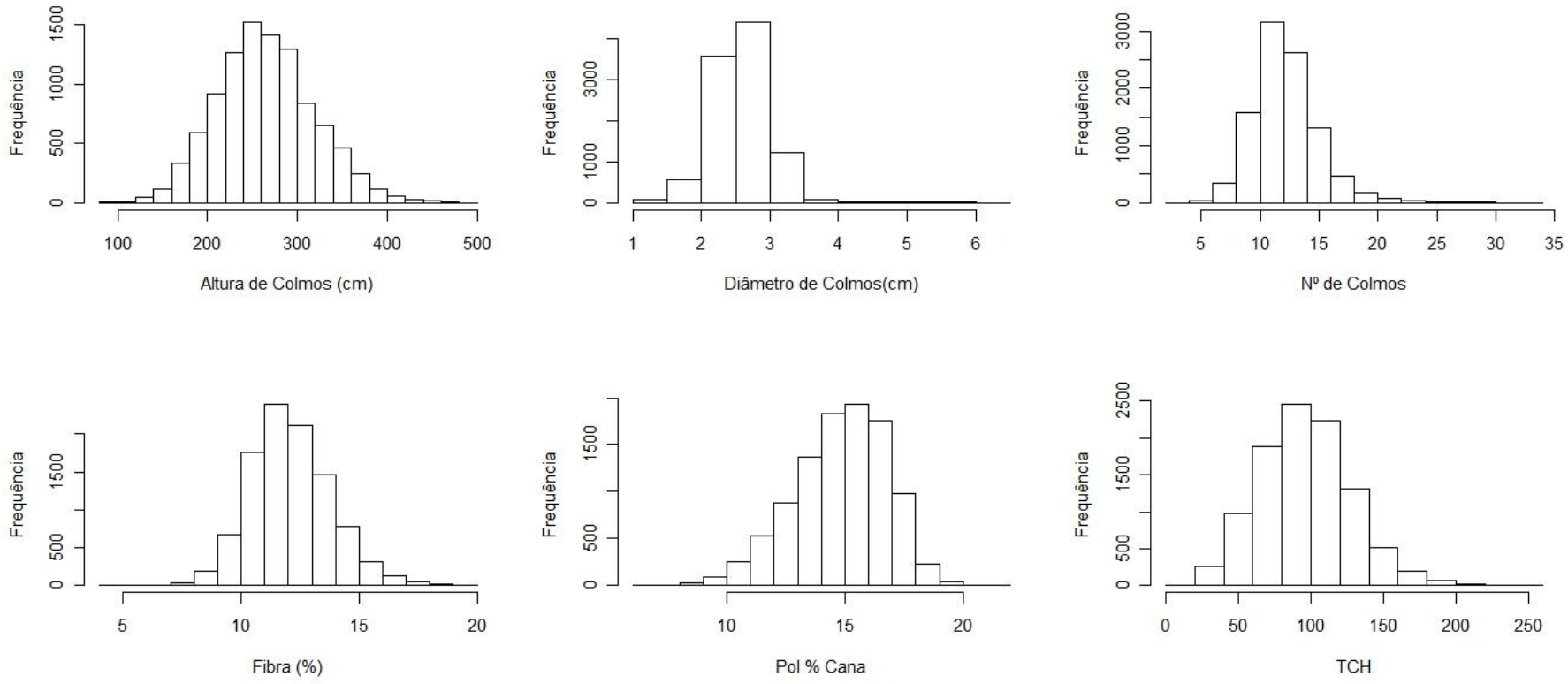


Figura 1 - Histogramas dos valores fenotípicos para os caracteres analisados

Tabela 6 - Total de alelos polimórficos e Conteúdo de Informação Polimórfica médio (PIC*m*) para os locos SSR. Os alelos correspondem aos picos observados nos eletroferogramas produzidos no analisador genético ABI PRISM 310, e analisados através dos programas GeneScan 3.7 e Genotyper 3.7 (Applied Biosystems).

Loco SSR	N° de alelos	PIC*m*	Loco SSR	N° de alelos	PIC*m*
CIR1	7	0,39	SMC119CG	15	0,33
CIR4	3	0,37	SMC179SA	7	0,31
CIR8	9	0,33	SMC221MS	7	0,36
CIR10	6	0,28	SMC334BS	7	0,33
CIR14	6	0,28	SMC1047HA	16	0,35
CIR15	15	0,22	SMC1149HA	11	0,34
CIR18	6	0,43	SMC1608CL	5	0,38
CIR19	6	0,43	SMC1825LA	2	0,37
CIR21	13	0,42	SMC2055FL	10	0,31
CIR27	8	0,33	ESTB41	8	0,33
CIR34	9	0,35	ESTB65	11	0,28
CIR42	6	0,35	ESTB92	4	0,27
CIR43	9	0,34	ESTC44	8	0,29
CIR44	5	0,26	CV29	15	0,36
CIR51	6	0,33	CV37	15	0,34
CIR56	10	0,34	CV38	16	0,32
CIR64	9	0,36	CV60	6	0,44
CIR66	3	0,40	CV78	17	0,29
CIR67	8	0,29	CV79	7	0,43
CIR69	2	0,44	CV94	6	0,35
SMC17CG	10	0,29	CV104	6	0,33
SMC36BUQ	8	0,36	CV106	7	0,37
Total de alelos: 370		Média de alelos: 8,4		Média dos PIC*m*: 0,34	

Notas:
$PIC_m = \sum_{i=1}^{N} PIC_i / N$
$PIC_i = 1 - \sum_{j=1}^{n} p_{ij}^2$

Tabela 7 - Total de bandas polimórficas e Conteúdo de Informação Polimórfica médio (PIC*m*) para as combinações de iniciadores AFLP. As bandas foram reveladas no analisador automático de DNA Li-cor modelo 4300 e analisadas através do programa SAGA^{MX} (LI-COR Biosciences)

N°	Combinação AFLP	N° de Bandas	PIC*m*	N°	Combinação AFLP	N° de Bandas	PIC*m*
1	ACG_CGT	37	0,36	10	AAG_CAC	39	0,38
2	ACG_CGC	34	0,32	11	AAG_CTC	22	0,31
3	ACG_CTT	43	0,42	12	AAG_CAG	18	0,32
4	ACG_CTC	32	0,33	13	AAA_CAT	14	0,33
5	ACG_CTA	25	0,37	14	AGA_CAT	21	0,29
6	ACG_CAT	32	0,33	15	AAA_CAG	25	0,41
7	AAG_CGT	35	0,39	16	ACA_CAG	18	0,34
8	AAG_CGC	28	0,38	17	AGG_CAG	17	0,37
9	AAG_CTG	22	0,36	18	AGA_CAG	18	0,39
Total de alelos: 480			Média de alelos: 26,7			Média dos PIC*m*: 0,36	

Notas:
A numeração das combinações é a mesma adotada na Tabela 5
$PIC_m = \sum_{i=1}^{N} PIC_i / N$
$PIC_i = 1 - \sum_{j=1}^{n} p_{ij}^2$

Cinco genótipos (‘IAC86-2480’, ‘IAC82-2045’, ‘CTC1’, ‘SP86-42’ e ‘IACSP96-2041’) apresentaram número discrepante de alelos de acordo com a distribuição dos números de alelos para os locos SSR (Figura 2). Quatro deles (‘IAC86-2480’, ‘IAC82-2045’, ‘CTC1’ e ‘SP86-42’) apresentaram número de alelos abaixo do valor mínimo da distribuição. Uma inspeção individual nesses genótipos mostrou que eles apresentaram uma proporção maior de dados perdidos quando comparados com os demais genótipos analisados, especialmente no caso da cultivar ‘IAC86-2480’ que apresentou apenas 36 alelos. Por causa deste baixo número de alelos SSR esta cultivar aparece também de forma discrepante na distribuição dos valores totais. Em relação aos marcadores AFLP todos os genótipos apresentaram valores dentro da distribuição.

Os três genótipos selvagens avaliados apresentaram um número menor de alelos em relação aos genótipos cultivados: 174 alelos para ‘SES 205A’ (*S. spontaneum*), 253 para ‘Chunnee’ (*S. barberi*) e 316 para ‘Badilla de Java’ (*S. officinarum*). Aitken et al. (2006) utilizando marcadores AFLP observaram resultados semelhantes em que clones de *S. officinarum* com 2*n*=80 cromossomos apresentaram um número médio de alelos de 214 e entre clones da mesma espécie com mais de 80 cromossomos a média foi de 243 alelos. Já entre a coleção de cultivares a média foi de 276 alelos variando de 229 a 309. Da mesma forma, Brown et al.

(2007) usando marcadores SSR, verificaram uma média de 67,1 alelos entre os clones de *S. officinarum*, 64,8 em *S. spontaneum* e 81,7 para *S. barberi*, enquanto que a média entre os clones comerciais foi de 78,2. Os resultados obtidos no presente trabalho não permitem uma comparação mais consistente dos níveis de heterozigosidade entre genótipos cultivados e selvagens, uma vez que um único representante de cada espécie selvagem foi utilizado. No entanto, considerando as devidas especificidades de cada estudo em relação aos genótipos usados, pode-se observar grande concordância entre os resultados aqui apresentados e entre os dois estudos citados. Tomados conjuntamente, os dados sugerem a existência níveis consideráveis de heterozigosidade e de diversidade entre os genótipos cultivados.

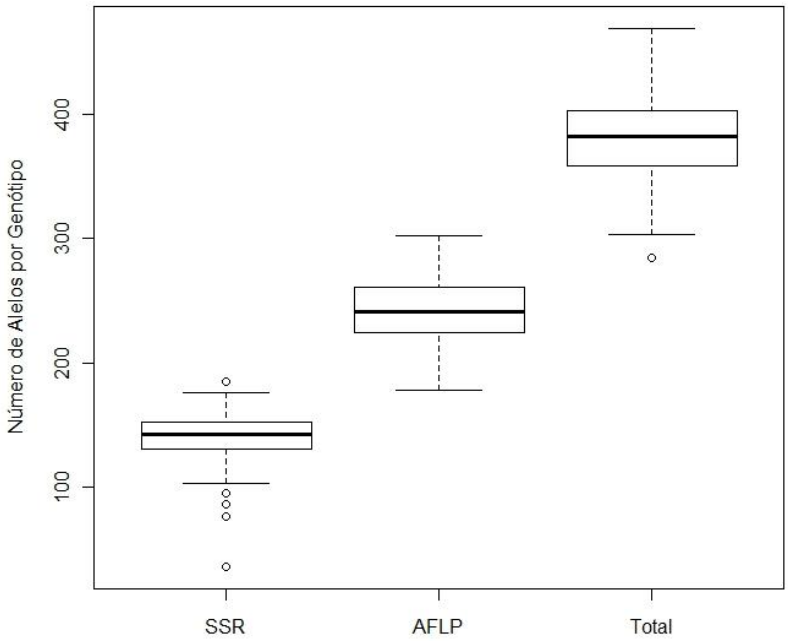


Figura 2 - Distribuição do número de alelos para os marcadores SSR, AFLP e para o Total (soma de ambos) avaliada nos 103 genótipos cultivados

O conteúdo de informação polimórfica dos locos SSR variou de 0,22 (loco *CIR15*) a 0,44 (locos *CIR69* e *CV60*) com média de 0,34 para os locos analisados. Para os locos de AFLP, os valores de PIC variaram de 0,29 para a combinação *AGA_CAT* a 0,42 para a combinação *ACG_CTT*, que também mostrou o maior número de alelos. Uma vez que os dois sistemas de

marcadores utilizados neste trabalho foram considerados como dominantes, cada alelo SSR e AFLP corresponde a um “loco marcador” com apenas dois estados alélicos possíveis: presença ou ausência. Esses padrões representam, na verdade, fenótipos moleculares de presença ou ausência de um fragmento amplificado. Essa compreensão é importante, pois determina a forma como os dados foram analisados. Assim, com base nos dois estados alélicos possíveis e considerando a expressão utilizada para o seu cálculo, o valor máximo que pode ser obtido para o PIC é de 0,50. Os valores médios tanto para os SSR (0,34) quanto para os AFLP (0,36) evidenciam a boa capacidade destes marcadores na discriminação das cultivares. Este resultado é mais uma confirmação da qualidade destes locos SSR e combinações AFLP, os quais têm sido amplamente utilizados em estudos de diversidade em cana (BESSE et al., 1998; LIMA et al., 2002; CORDEIRO; PAN; HENRY, 2003; SELVI et al, 2006; AITKEN; JACKSON; McINTYRE, 2006; BROWN et al., 2007).

3.3 Análise de similaridade genética

A análise da matriz de dissimilaridade genética baseada no coeficiente de Dice revelou que a média de dissimilaridade foi de 0,37, ou seja, o índice médio de similaridade entre as variedades analisadas foi de 63%. As medidas de similaridade individuais variaram de 0,33 entre a cultivar ‘CTC3’ e a espécie selvagem ‘SES 205A’ (*S. spontaneum*) a 0,92 entre as cultivares ‘IACSP96-2037’ e ‘IACSP96-2041’. O dendrograma construído com base na matriz de dissimilaridade genética, cujo coeficiente de correlação cofenético foi r_{cof} = 0,97, indicou forte correlação entre a matriz de dissimilaridade e as distâncias no dendrograma (Figura 3). Observa-se que ‘Badilla de Java’, ‘Chunnee’ e ‘SES 205A’ foram os que apresentaram maiores distâncias em relação aos demais genótipos, uma vez que se tratam de genótipos selvagens de cana-de-açúcar. Pode-se observar também que dos três, ‘Badilla de Java’ foi o mais próximo dos genótipos cultivados, seguido de ‘Chunnee’ (*S. barberi*) e, por último, ‘SES 205A’. O arranjo está de acordo com o processo de melhoramento da cana, no qual as cultivares modernas se originaram por nobilização, ou seja, sucessivos retrocruzamentos para *S. officinarum*. Uma hipótese postula que *S. barberi* pode ter sido originada através de hibridações entre *S. officinarum* e *S. spontaneum* e isso explicaria sua localização entre as duas espécies no dendrograma. *Saccharum spontaneum* é considerada uma das espécies com maior diversidade genética entre as espécies do gênero *Saccharum*, o que justifica a alta dissimilaridade do genótipo ‘SES 205A’ em

relação aos demais genótipos. Selvi et al. (2003) utilizando marcadores SSR de milho foram capazes de estabelecer relações genéticas entre espécies pertencentes aos gêneros *Saccharum* e *Erianthus*, identificando a presença de cinco grupos entre 33 genótipos selvagens e comerciais de cana-de-açúcar. Os resultados mostraram *S. spontaneum* formando um grupo separadamente dos demais; *S. barberi* e *S. sinensi* agrupando juntas; *S. officinarum* e *S. robustum* formaram um grupo separado; um grupo reuniu apenas as cultivares comerciais e outro apenas as espécies de *Erianthus*. O agrupamento de clones das espécies *S. officinarum* e *S. robustum* tem sido relatado frequentemente e reflete a origem de *S. officinarum* a partir de *S. robustum* (GLASZMANN et al., 1989; BURNQUIST; SORRELLS; TANKSLEY, 1992; NAIR et al., 1999). O grupo formado pelas espécies *S. barberi* e *S. sinensi* posicionou-se entre o grupo de *S. officinarum* e o de *S. spontaneum*, observação que, segundo os autores, é condizente com suas prováveis origens a partir dessas espécies. Brown et al. (2007) realizaram um estudo para avaliar as relações entre cinco espécies de *Saccharum*: *S. officinarum* (SO), *S. spontaneum* (SP), *S. barberi* (SB), *S. sinensi* (SS) e *S. robustum* (SR). Utilizando métodos multivariados os diversos clones das espécies avaliadas formaram dois grupos principais sendo que o grupo I envolveu clones de SO, SB, SS, SR e clones comercias, e o grupo II conteve apenas clones de SP. Dentro do grupo I ocorreram sete subgrupos, assim compostos: pelo menos três subgrupos (1 e 3a e 3b) foram formados basicamente por clones de SB e SS, com a presença de alguns SR; o subgrupo 2 conteve apenas clones SB; o subgrupo 3c conteve uma mistura de clone SB e clones comerciais; o subgrupo 4 conteve principalmente clones SR e o subgrupo 5 foi formado por clones de SO e SB. Ressalta-se nesses resultados a clara separação de *S. spontaneum*, sugerindo maior divergência desta espécie em relação às demais e o agrupamento dos clones de *S. barberi* e *S. officinarum* com clones comerciais. No entanto, baseados nos agrupamentos do dendrograma e em análise de coordenadas principais (PCA), os autores chamam a atenção para o fato de que nenhum dos genótipos de *S. barberi* apareceu como progênie das supostas espécies parentais o que indicaria que outra espécie pertencente a algum gênero relacionado, *Erianthus* ou *Miscanthus*, poderia ter contribuído também para a formação do genoma de *S. barberi*.

Métodos multivariados baseados em dendrogramas são úteis para a análise das relações entre os indivíduos de uma população e podem ser usados para identificar grupos de indivíduos. Apesar da alta similaridade média observada entre as cultivares, uma análise atenta do

dendrograma da Figura 3 sugere a existência de três grupos de cultivares na população. No entanto, essa divisão é arbitrária uma vez que não há um critério específico para discriminação destes grupos com base apenas na informação gerada pelo dendrograma. Observa-se que, em linhas gerais, houve agrupamento das cultivares de acordo com o programa de melhoramento ao qual pertencem. O grupo 1 é constituído, basicamente, por cultivares IACSP e algumas RB. Destaca-se nesse grupo a presença de 'POJ2878', 'NA56-79' e 'RB72-454', que são genótipos importantes na genealogia das diversas variedades brasileiras. Os genótipos identificados com a sigla IACSP são cultivares desenvolvidas através de uma colaboração entre os programas do IAC e da antiga Copersucar. No grupo 2 estão presentes cultivares IAC e CTC e três IACSP. O grupo 3 reuniu todas as cultivares SP e mais algumas RB. Pode-se observar ainda que os grupos 1 e 2 partem de um ramo comum no dendrograma, sugerindo maior relacionamento entre as cultivares pertencentes a estes dois grupos em relação às pertencentes ao grupo 3.

Os principais programas de melhoramento de cana-de-açúcar no Brasil utilizaram, basicamente, os mesmos grupos de genitores no desenvolvimento de suas variedades comerciais, de modo que a grande similaridade entre os genótipos aqui observada é considerada normal. Alta similaridade genética entre genótipos de cana-de-açúcar tem sido relatada em diversos estudos (SELVI et al., 2003; AITKEN; JACKSON; McINTYRE, 2006; CRESTE et al., 2010). Uma análise na genealogia das variedades até a terceira geração, realizada no sistema Caiana que gerencia o banco de germoplasma do IAC revelou que há recorrência de um número restrito de genótipos parentais das cultivares usadas neste trabalho. De modo geral, os genótipos 'Co419', 'NA56-79', 'POJ2878', 'RB72-454', 'SP70-1143', 'SP70-1284', 'SP71-1088', 'SP84-2066' e 'SP71-345' apareceram de forma bastante freqüente como parentais das cultivares pertencentes aos três grupos aqui identificados.

Analisando-se apenas os parentais de primeira geração em cada grupo, observou-se que o genótipo 'RB72-454' aparece com freqüência na genealogia das demais cultivares da série RB. Curiosamente, essas cultivares se dividiram entre os grupos 1 e 3, embora fosse esperado que elas estivessem apenas no grupo 3 com base nos registros genealógicos ao longo das três gerações. Considerando somente as cultivares IACSP do grupo 1, observou-se que a maioria são meios-irmãos e irmãos completos, derivados de cruzamentos biparentais. Entre essas cultivares destaca-se a ocorrência dos genótipos 'SP77-5181', 'SP80-185', 'SP82-6108' e 'SP84-2066' como

parentais de primeira geração em 44% dos casos. As cultivares do grupo 2 praticamente não apresentaram parentais de primeira geração em comum. No entanto, entre os parentais de segunda e terceira gerações, destacam-se os genótipos ‘Co419’, ‘POJ2878’, ‘SP70-1143’, ‘SP70-1284’ e ‘SP71-1088’. Outra característica das cultivares desse grupo é que a maioria (73%) é resultante de policruzamentos (Tabela 1). No entanto, a análise das três gerações de parentais das cultivares do grupo 2 não sugere nenhuma explicação razoável para o agrupamento das cultivares da séria CTC junto com cultivares IAC. Por sua vez, os parentais de primeira geração mais comuns nas cultivares do grupo 3 foram ‘RB72-454’, ‘SP70-1143’ e ‘SP71-1088’. Neste grupo, especialmente entre as variedades da série SP, destaca-se a forte ocorrência de variedades Co (originárias do programa de melhoramento de Coimbatore), CP (Canal Point) e H (Hawaí) na genealogia dessas cultivares em todas as gerações analisadas.

Embora uma análise aprofundada das relações genealógicas esteja fora do escopo deste trabalho, observou-se de um modo geral que as cultivares relacionadas em termos genealógicos localizaram-se próximas umas das outras no dendrograma. Por exemplo, no grupo 1 existe um agrupamento reunindo as cultivares ‘IACSP95-5000’, ‘IACSP97-6682’, ‘IACSP97-7065’, ‘IACSP97-2053’, ‘IACSP97-2055’, ‘IACSP97-3402’ e ‘IACSP97-3406’, todos possuindo a cultivar ‘SP84-2066’ como parental feminino de primeira geração. Outro agrupamento é representado pelas cultivares ‘IACSP96-2042’, ‘IACSP94-2094’, ‘IACSP94-2101’, ‘IACSP94-2180’, ‘IACSP94-4002’ e ‘IACSP94-4004’, as quais possuem ‘SP77-5181’ como parental de primeira geração, com exceção do genótipo IACSP94-2094 que pelos registros genealógicos não compartilha nenhum parental em comum com as demais cultivares até a terceira geração. Outro exemplo ainda no grupo 1 é o das cultivares ‘RB83-5054’, ‘RB84-5210’ e ‘RB84-5257’ que agruparam quase todas próximas à ‘RB72-454’ que é seu progenitor feminino. A exceção foi dos genótipos ‘RB85-5035’ e ‘RB83-5486’, os quais não possuem a ‘RB72-454’ em sua genealogia, mas são meio-irmãos por compartilharem o parental feminino ‘L6014’. Curiosamente estas cultivares se posicionaram distantes no dendrograma. No grupo 2 o estabelecimento dos agrupamentos menores considerando as informações genealógicas foi menos evidente, a não ser pela presença de parentais em comum nas gerações mais avançadas, conforme já comentado anteriormente. No grupo 3 houve uma agrupamento envolvendo 13 cultivares RB e SP. Neste agrupamento todas as cultivares são derivadas de ‘RB72-454’ ou de ‘SP70-1143’ e em alguns

casos derivam do cruzamento direto entre esses dois parentais. Destacam-se ainda o agrupamento das cultivares ‘SP80-3280’, ‘SP80-1842’ e ‘SP80-1816’, todas filhas da cultivar ‘SP71-1088’. As demais cultivares, de modo geral, não apresentaram genótipos parentais em comum nas primeiras gerações. No entanto, entre eles é comum a ocorrência de genótipos CP em gerações mais avançadas.

Como visto, a compreensão das relações genéticas dentro do gênero *Saccharum* é bastante complexa. Abordagens baseadas em marcadores moleculares devem considerar entre outras coisas uma boa cobertura do genoma, embora não se possa determinar com exatidão a quantidade mínima necessária para tal; o tipo de marcador a ser usado e as regiões do genoma atingidas por esse marcador; o tamanho da população e, se envolver genótipos selvagens, a origem desses materiais; e métodos estatísticos adequados. Mesmo assim, nem sempre os resultados obtidos correspondem às informações fornecidas pelos registros genealógicos. Embora os registros atuais sejam confiáveis, os mais antigos podem não ter a mesma confiabilidade o que é, sem duvida, um complicador na determinação exata dos ancestrais das cultivares atuais. A natureza dominante dos marcadores também contribui para a perda de poder na discriminação das cultivares. Além disso, a poliploidia do genoma de cana e, especialmente a aneuploidia, são fatores que certamente contribuem para algumas das distorções observadas.

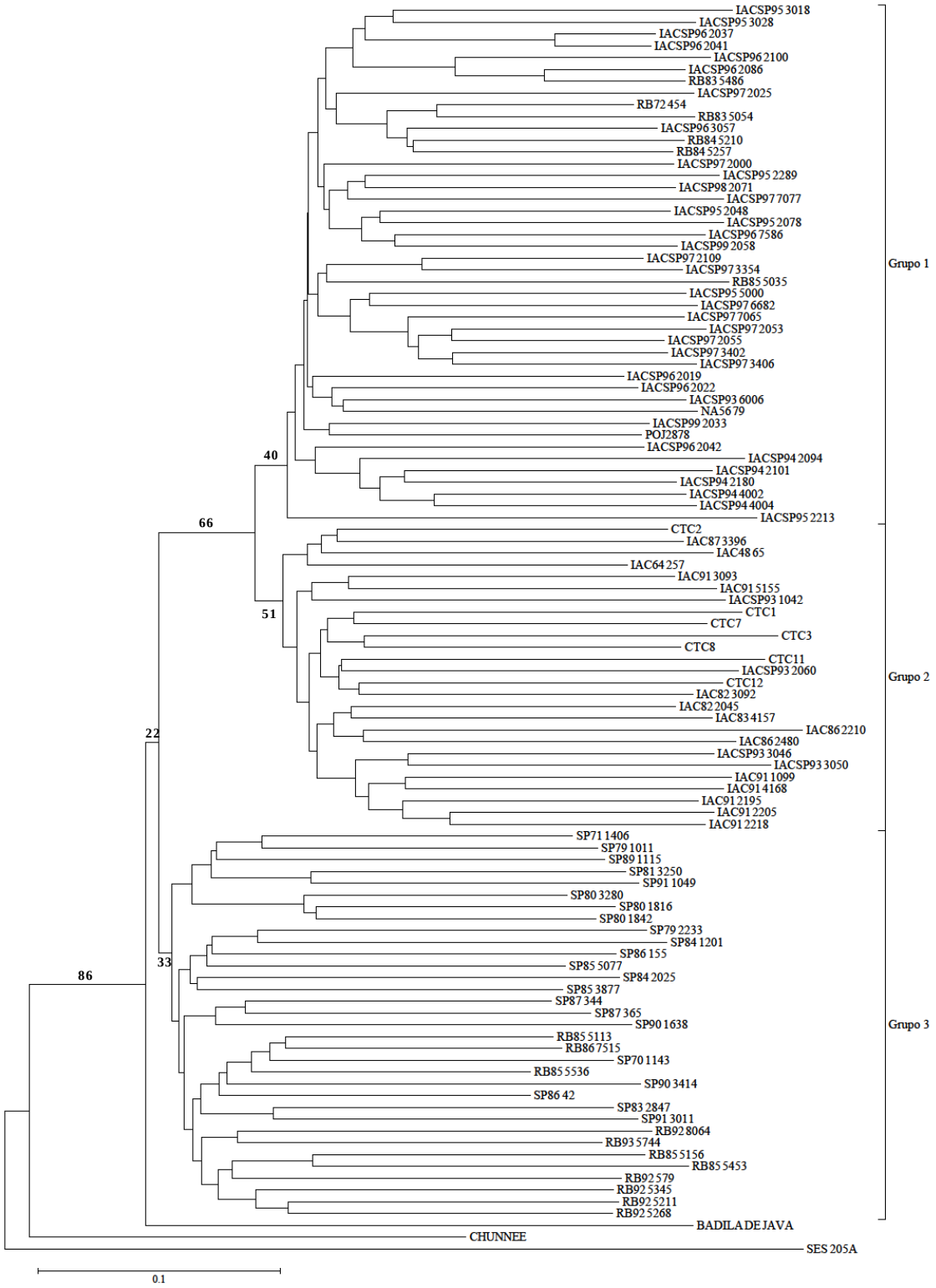


Figura 3 - Dendrograma mostrando as relações genéticas entre os 103 genótipos cultivados de cana e 3 genótipos selvagens. O método de agrupamento foi *Neighbor-Joining* modificado segundo Gascuel (1997) utilizando a matriz de dissimilaridade genética de Dice. São mostrados apenas os bootstraps dos principais grupos. A barra indica a distância entre os genótipos

A diferenciação observada entre os genótipos pertencentes a diferentes programas de melhoramento é de certa forma, surpreendente, no sentido que houve uma profusa mistura de alguns genótipos parentais na formação das cultivares, conforme evidenciada na análise genealógica. Jannoo et al. (1999a) verificaram que o agrupamento das cultivares de acordo com o programa de melhoramento em seu estudo se deveu ao fato de alguns marcadores serem específicos ou apresentarem maior freqüência em certos grupos de cultivares. Eles observaram que estes marcadores também eram específicos ou predominantes no genoma de *S. spontaneum* e ao analisar a genealogia daquelas cultivares foi possível determinar que elas herdaram tais marcadores de híbridos interespecíficos que continham *S. spontaneum* em sua genealogia.

Creste et al. (2010) utilizaram marcadores TRAP específicos para genes do metabolismo de açúcar e de resistência à seca, visando estimar a variabilidade genética entre 53 genótipos cultivados pertencentes a diferentes programas de melhoramento. Também foram incluídas quatro espécies selvagens do gênero *Saccharum* e uma espécie de *Erianthus*. No entanto, um agrupamento claro envolvendo genótipos em função do programa de melhoramento não foi observado, provavelmente porque os genes usados expressaram polimorfismos de regiões especificas do genoma de cana e não puderam refletir de forma mais ampla a variabilidade existente na população. Mesmo assim, uma análise do nível de variabilidade determinou que a maior parte da variação para os caracteres estudados estava dentro de cada programa e somente uma fração pequena estava distribuída entre eles.

Para complementar a análise de agrupamento realizada através do dendrograma um segundo método baseado em análise de componentes principais (PCoA) foi utilizado. Enquanto os métodos baseados em distâncias genéticas tendem a representar as relações entre os indivíduos em termos de sua genealogia, os métodos baseados em análise de componentes principais representam estas relações em termos da diversidade ou da variação presente na população. Além de fornecer uma estimativa dos níveis de variabilidade na população este método de análise permite determinar com mais clareza a ocorrência de subgrupos na população. Neste sentido, a realização desta análise forneceu uma primeira indicação do número de subgrupos presentes na população de mapeamento por associação. Uma vez que os genótipos selvagens não fazem parte do painel de associação e para evitar eventuais distorções na distribuição dos grupos introduzida por sua presença, estes genótipos não foram considerados na análise.

O resultado dessa análise mostrou que 23,4% da variação na população foi explicada pelos quatro componentes principais. Resultado semelhante foi relatado por Brown et al. (2007) em que os três primeiros eixos explicaram 22,5% da variação. Segundo os autores esse valor representa um bom ajuste do modelo, dada a variabilidade existente na população e o grande número de marcadores usado. A Figura 4 mostra o agrupamento das cultivares resultante da análise dos dois primeiros componentes principais, os quais explicaram 16,6% da variação.

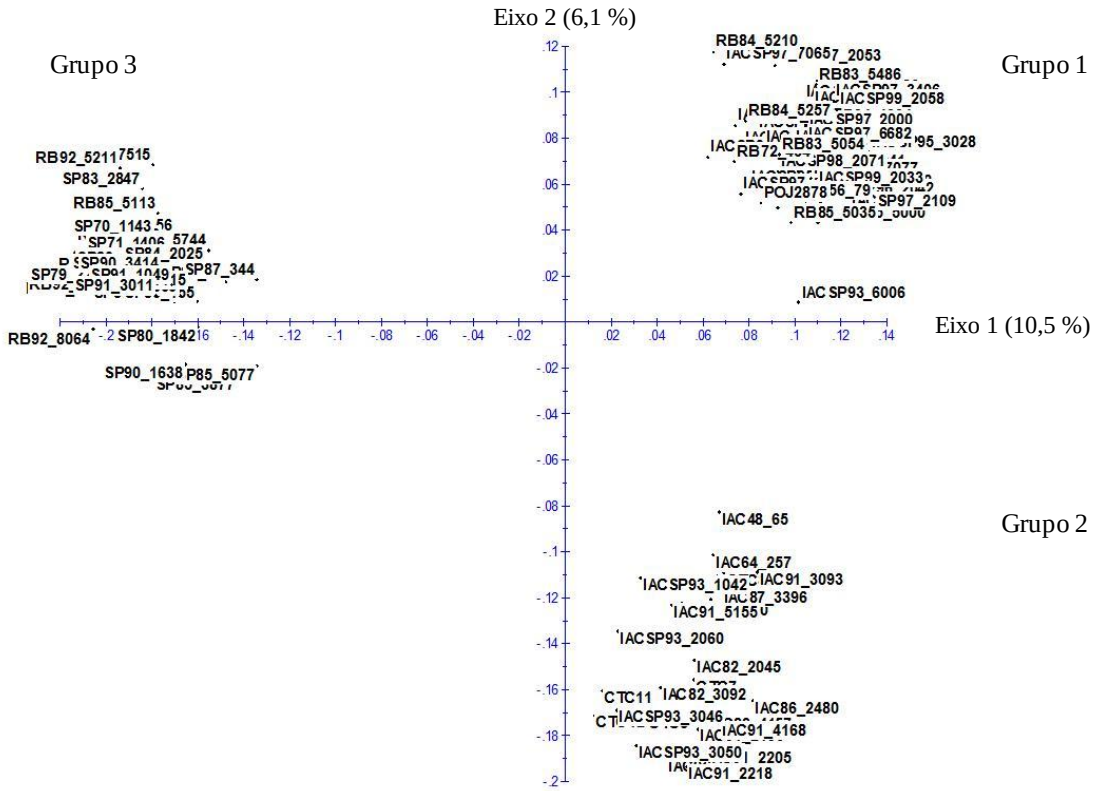


Figura 4 - Agrupamento dos 103 genótipos cultivados de cana revelados pelo método PCoA. A numeração dos grupos corresponde à mesma numeração adotada no dendrograma. Os dois primeiros eixos ou componentes principais estão indicados na figura com os respectivos valores de variação distribuídos em cada eixo

Houve uma completa correspondência entre os grupos formados através deste método e os grupos formados no dendrograma. Isso se deve em parte ao fato dos dois métodos serem baseados em distâncias genéticas. Observa-se que o eixo 1 foi responsável por 10,5% da

variação e separou o grupo 3 dos demais grupos, enquanto que o eixo 2 explicou 6,1% da variação e separou os grupos 1 e 2, resultado que reforça uma maior similaridade entre esses dois grupos. Nota-se também que os grupos 2 e 3 apresentam dispersão ligeiramente maior que o grupo 1. Uma análise da similaridade genética média dentro de cada grupo foi realizada como forma de medir indiretamente a diversidade genética dentro de cada grupo. Os valores obtidos foram: SG_1 = 0,70, SG_2 = 0,67 e SG_3 = 0,66. As diferenças não foram muito grandes, mas confirmam a tendência e haver maior variação nos grupos 2 e 3 em relação ao grupo 1.

Portanto, os resultados apresentados no dendrograma, na análise de componentes principais, e considerando ainda a análise da genealogia e os valores de similaridade média dentro dos grupos, quando analisados conjuntamente sugerem que os grupos 2 e 3 são os que apresentam maior variabilidade. O fato do grupo 1 ter apresentado o maior valor de similaridade genética confirma a análise de genealogia discutida anteriormente que mostrou que a maioria das variedades são de irmãos completos e de meio-irmãos.

3.4 Níveis e estrutura do desequilíbrio de ligação no genoma de cana

A avaliação dos níveis de desequilíbrio é uma etapa crítica num estudo de associação, pois a extensão desse desequilíbrio no genoma da espécie determina a resolução da análise de mapeamento por associação. Tem sido observado que a estrutura do desequilíbrio é uma característica especifica da população ou dos locos em estudo. Portanto, a correlação entre locos numa população pode ser influenciada por diversos fatores como pressão de seleção, migração, deriva, mutação e ligação física, de modo que uma distinção entre o desequilíbrio de ligação resultante de ligação física entre os locos das demais causas é essencial para a compreensão de sua estrutura no genoma e também para se compreender como essa estrutura influencia os níveis de associação entre QTLs e caracteres fenotípicos.

A partir das 243 marcas oriundas do mapa genético da população ‘SP80-180’ x ‘SP80-4966’ foram obtidas 29.403 correlações, sendo 2.157 exclusivamente entre marcas ligadas, ou seja, dento dos grupos de ligação e 27.246 entre marcas de grupos diferentes ou marcas não ligadas. Os valores dos coeficientes de correlação r^2 para cada grupo de marcas foram plotados num único gráfico para permitir uma análise da estrutura do desequilíbrio de ligação no genoma de cana. No caso das marcas ligadas os valores foram expressos em função da distância genética

que as separa. Como o maior dos grupos de ligação possui comprimento de 93 cM, adotou-se a distância de 100 cM para representar as marcas não ligadas (Figura 5).

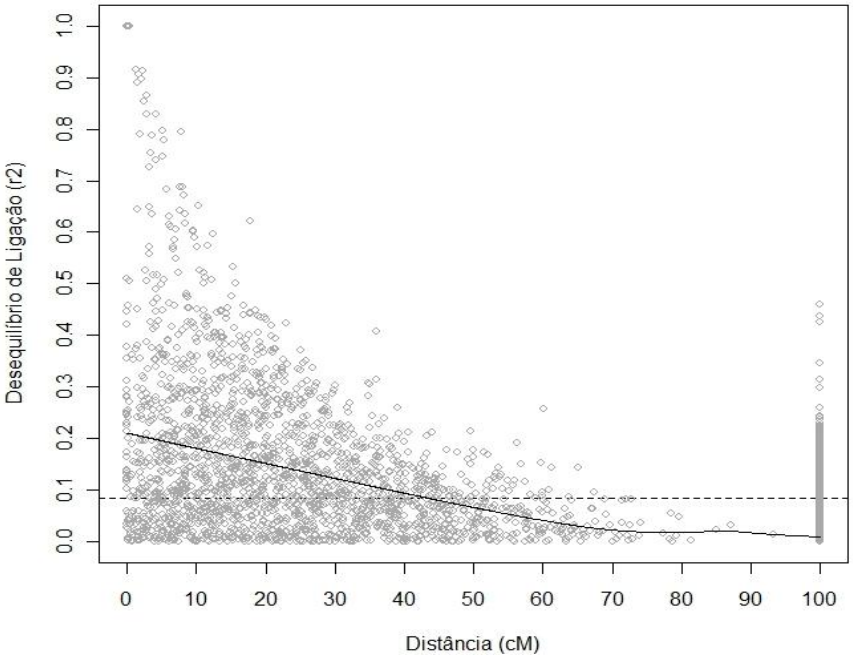


Figura 5 - Estrutura do desequilíbrio de ligação em função da distância genética para 243 marcas AFLP, RFLP, SSR e TRAP do mapa genético construído a partir de uma população de 92 indivíduos do cruzamento entre os clones 'SP80-180' x 'SP80-4966'. A linha horizontal tracejada representa o percentil 97,5% da distribuição de r^2 para marcas localizadas em grupos de co-segregação diferentes (marcas não ligadas) no mapa genético. A linha sólida representa a curva de regressão para os dados ajustada com a função de LOESS

Entre as marcas ligadas as estimativas de desequilíbrio variaram de $4{,}0 \times 10^{-6}$ a 1,0 com uma média de 0,1490. Para as marcas não ligadas os valores variaram de zero a 0,4596 com valor médio de 0,0161. Para diferenciar associações verdadeiras de associações falsas dois métodos distintos foram usados. O primeiro consistiu no estabelecimento de um nível crítico de significância (5%) para comparação com os *p* valores associados às estimativas dos coeficientes de desequilíbrio de ligação, utilizando para tal a correção de Bonferroni. Em relação ao número total de correlações observou-se que 402 (1,36%) apresentaram desequilíbrio significativo após a correção de Bonferroni ($p < 1{,}70 \times 10^{-6}$) sendo que a imensa maioria destas associações (396) foram entre marcas ligadas e 6 entre marcas não ligadas (0,002%). Considerando apenas as marcas ligadas e a distribuição de suas associações em relação às distâncias genéticas observou-se que 174 (43,9%) envolviam marcas separadas por distâncias menores que 10 cM. Este

percentual atinge o valor de 98% quando se analisa as associações até 30 cM no mapa. Houve, no entanto, um par de marcas (*DIR2A3.2* e *DIR1A1.2*) separadas por 60 cM que exibiram associação significativa. O *p* valor associado à estimativa da correlação entre essas marcas foi $1{,}19 \times 10^{-6}$. Mesmo próximo do limite crítico estabelecido pela correção de Bonferroni, este valor foi baixo o suficiente para descartar um provável escape do teste estatístico. Embora não seja possível com base apenas nessa evidência determinar a causa exata de tal resultado, sabe-se que diversos fenômenos genéticos poderiam criar situações como essa. O próprio histórico de melhoramento da cana pode ter favorecido o estabelecimento de grandes blocos cromossômicos em elevado desequilíbrio conforme tem sido demonstrado (JANOO et al., 1999a; RABOIN et al., 2008; WEI et al., 2010).

Analisando de forma mais cuidadosa estas duas marcas, verificou-se que ambas estavam presentes num grupo de co-segregação que reuniu 52 marcas do tipo TRAP, todas relacionadas ao metabolismo de açúcar e fibra (CG 3 - Anexo A). As próprias marcas são membros de uma família de genes que codificam proteínas dirigentes (*DIR*), as quais estão relacionadas aos processos de defesa e síntese de fibras (CALSA Jr., 2005; CASU et al., 2004; DAMAJ et al., 2010). Portanto, este grupo de co-segregação parece abrigar grupos de genes envolvidos com processos fisiológicos relacionados o que é um fenômeno relativamente comum. Em cana-de-açúcar isso poderia ter acontecido nas fases iniciais de melhoramento com o estabelecimento de alguma combinação favorável de genes em uma ou poucas cultivares híbridas. O processo de fixação desta combinação poderia ter se intensificado a partir do uso freqüente deste(s) genótipo(s) nas fases seguintes de melhoramento ou pelo uso recorrente de genótipos aparentados. Raboin et al. (2008) relataram que cromossomos híbridos entre *S. officinarum* e *S. spontaneum* parecem possuir segmentos grandes em elevados níveis de desequilíbrio e que tais cromossomos identificados na cultivar ‘R570’ demonstram uma taxa de recombinação menor, contribuindo desta forma para a manutenção dessas combinações. No caso da população de mapeamento brasileira o cruzamento foi realizado a partir dos clones elite SP80-180, parental feminino com alta produção de colmos e baixo teor de açúcar, com SP80-4966, parental masculino com baixa produção de colmos e alto teor de açúcar (GARCIA et al., 2006). Uma análise da genealogia dessas duas cultivares (Figura 6) evidencia que a cultivar ‘POJ100’ desempenhou um papel importante na formação dos dois progenitores da população de mapeamento.

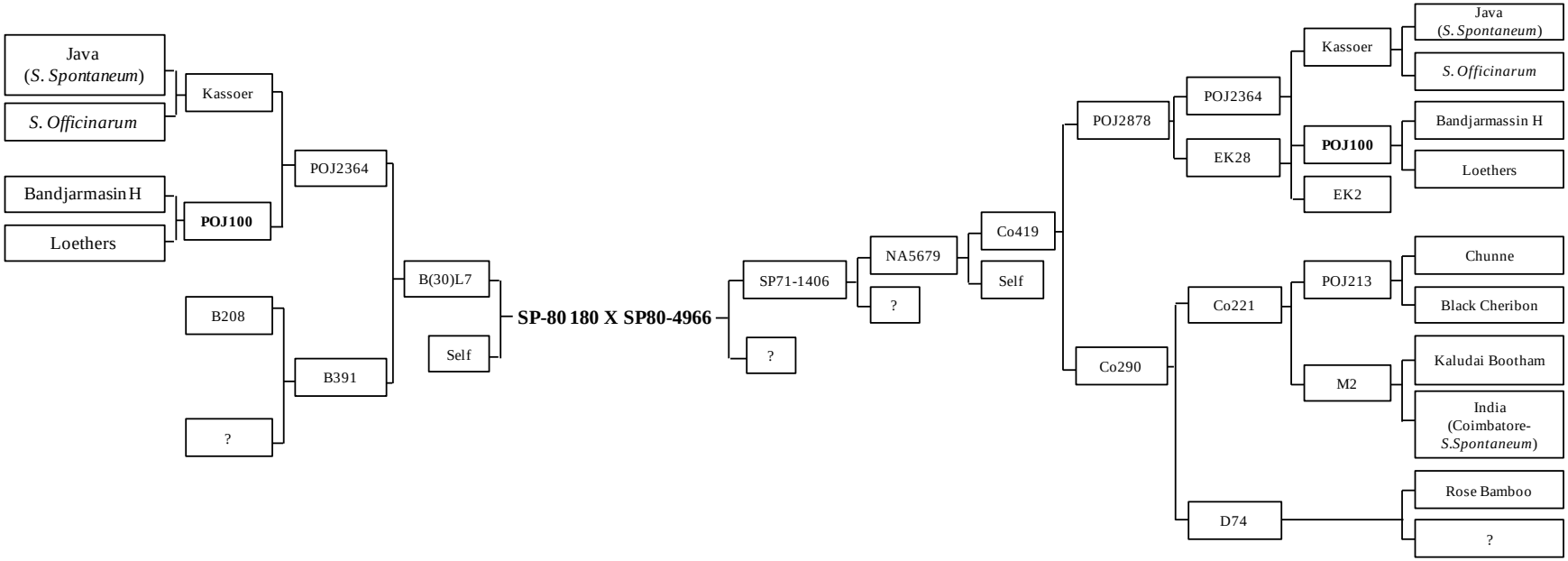


Figura 6 - Genealogia das cultivares SP80-180 e SP80-4966 usadas como parentais para obtenção do mapa genético de ligação brasileiro. Em negrito está destacado também o genótipo POJ100. O símbolo "?" indica que o parental desconhecido (policruzamento). O termo "Self" significa parental autofecundado

A cultivar ‘POJ100’ participou, por exemplo, como parental feminino das duas cultivares que originaram a cultivar ‘POJ2878’. Esta, por sua vez, foi o parental feminino da cultivar ‘Co419’ que foi autofecundada para obtenção da ‘NA56-79’. A partir desta última cultivar foram realizados dois ciclos de policruzamento até a obtenção do genótipo ‘SP80-4966’. Do lado da cultivar ‘SP80-180’, ‘POJ100’ aparece como parental masculino de terceira geração. Portanto, o tamanho efetivo da população de mapeamento se restringe a poucos genótipos e esse histórico endossa a origem de uma combinação de genes relacionada ao metabolismo de açúcar em alto desequilíbrio, tendo a cultivar ‘POJ100’ como o genótipo onde esta combinação provavelmente se originou ou através do qual ela foi passada às demais gerações.

Outra situação ocorrida, porém bem mais comum do que a anterior, foi a de marcas separadas no mapa de ligação por distâncias pequenas, mas que não se mostraram significativamente associadas. Obviamente uma distância menor não implica necessariamente em desequilíbrio significativo. Sendo assim, verificou-se, por exemplo, que nos primeiros 5 cM houve 165 (66,8%) pares de marcas de um total de 247 com associações não significativas.

Sabe-se que o método de Bonferroni é extremamente conservativo e que associações verdadeiras podem ser deixadas de fora em função do rigor do teste. Por isso um segundo método baseado numa análise de caráter gráfico foi adotado (BRESEGHELLO; SORRELS, 2006; ROBBINS et al., 2011). Neste método é estabelecido um nível crítico de desequilíbrio a partir da distribuição dos valores das correlações entre as marcas não ligadas, tomando-se um valor de corte nesta distribuição que corresponde, no mínimo, a 95% da distribuição. Uma curva de regressão dos dados é calculada e o ponto de interseção entre esta curva e o limite crítico anterior determina no eixo das distâncias a extensão do desequilíbrio. A análise da Figura 5 mostra o resultado obtido através desta estratégia. Em linhas gerais o gráfico evidencia uma queda relativamente suave no nível de desequilíbrio. Observa-se que houve poucos valores de desequilíbrio com r^2 = 1 ou próximos, fazendo com que a curva tivesse início num valor pouco maior que 0,2. Considerando o ponto de interseção entre a linha que representa o limite crítico (r^2 = 0,0883) e a curva de regressão, verifica-se que o desequilíbrio de ligação foi bastante comum em distâncias de até 40 cM. Diferentemente da primeira abordagem que é usada para determinar o número de associações estatisticamente significativas, esta outra permite avaliar a extensão em que o desequilíbrio está presente no genoma. É claro que os dois métodos estão

relacionados, conforme foi discutido anteriormente que a maioria das associações significativas, detectadas pela primeira abordagem, situam-se nos limites estabelecidos pela segunda, ou seja, 30 cM.

Analisando de forma mais detalhada a curva de regressão percebe-se que os valores médios dos coeficientes de correlação foram relativamente constantes à medida que se avança no mapa até o limite de 40 cM. Esse resultado contrasta um pouco com os resultados já relatados na literatura demonstrando forte concentração de desequilíbrio em distâncias mais curtas. O primeiro estudo que analisou os níveis de desequilíbrio no genoma de cana foi realizado por Jannoo et al. (1999a) numa população de cultivares originárias principalmente do programa de melhoramento das ilhas Maurícius. Este trabalho usou como referência o mapa da cultivar 'R570' e foi observado que os níveis mais significativos de desequilíbrio envolviam marcas distanciadas a menos de 10 cM e ainda que dois terços das marcas em desequilíbrio eram derivadas de *S. spontaneum*. Os autores relatam que, embora somente cerca de 20% do genoma das cultivares modernas seja derivado dessa espécie, esta observação pode estar relacionada à maior resolução dos marcadores moleculares nesta fração do genoma e ao maior polimorfismo também associado a ele. Raboin et al. (2008) e Wei et al. (2010) por sua vez, obtiveram resultados muito parecidos em que os níveis mais expressivos de desequilíbrio se concentraram nos primeiros 5 cM, caindo rapidamente, mas com valores ainda significativos até 25-30 cM e, ocasionalmente, atingindo 40 cM. Apesar da discordância no nível observado nas distâncias mais curtas os resultados aqui apresentados e os da literatura estabelecem a distância até 40 cM como o limite para a maior parte do desequilíbrio de ligação significativo no genoma de cana. Essa diferença pode estar relacionada ao número de marcas utilizadas. Neste estudo foram usadas 243 marcas, contra 1.057, 396 e 310 utilizadas nos trabalhos citados anteriormente nesta mesma ordem. Embora houvesse disponível 1.923 marcas o objetivo era somente obter uma caracterização geral dos níveis de desequilíbrio numa população oriunda de cultivares brasileiras, já que dos outros trabalhos, dois foram realizados com a cultivar francesa 'R570' e o outro com a cultivar australiana 'Q165'. Deste modo, considerando que o desequilíbrio é uma característica que depende, entre outros fatores, da origem da população, pode ser que parte dessas diferenças esteja associada ao histórico de melhoramento particular de cada população. Por exemplo, Raboin et al. (2008) discutem que a maioria dos blocos em desequilíbrio observados na cultivar

‘R570’ seriam herdados da ‘Co213’ e principalmente de ‘POJ2878’, dada a elevada recorrência dessas cultivares tradicionais na genealogia de ‘R570’.

O segundo grupo de marcas analisado, representado por um subconjunto de 405 alelos genotipados na população de mapeamento por associação, apresentou um valor médio de correlação de 0,0153, muito semelhante ao valor médio das marcas não ligadas do mapa genético. No entanto, as correlações apresentaram uma distribuição de valores bem mais ampla, variando de zero a 0,8535. Este valor máximo sugeriu a existência de desequilíbrio significativo entre tais marcas, o que foi confirmado após comparação dos *p* valores associados a cada estimativa de r^2 com o nível de significância corrigido por Bonferroni. Desta forma, foram observadas 56 correlações significativas ($p < 6{,}11 \times 10^{-7}$) correspondendo a 0,068% das 81.810 correlações testadas neste conjunto de marcas. Comparando-se o coeficiente médio de correlação destas marcas ($r^2 = 0{,}400$) com o das marcas ligadas do mapa de ligação que também exibiram desequilíbrio significativo ($r^2 = 0{,}3929$), verifica-se que os valores foram também muito semelhantes. A mesma situação acontece quando se compara as menores correlações significativas que foram $r^2 = 0{,}2523$ para as marcas do mapa genético e $r^2 = 0{,}2463$ para as da população de mapeamento por associação. Portanto, pode-se inferir a partir destas comparações que níveis idênticos de desequilíbrio existem nas duas populações.

Apesar de ser uma espécie alógama e com elevada heterozigosidade nos locos, a estrutura do desequilíbrio no genoma das cultivares modernas de cana se aproxima mais da estrutura de espécies com autofecundação. A origem a partir de um número limitado de genótipos fundadores; o sistema de reprodução assexuada; o uso recorrente de genótipos aparentados nos cruzamentos e o número limitado de gerações transcorridas desde as primeiras hibridações que originaram as cultivares modernas, contribuíram para que os elevados níveis de desequilíbrio presentes no genoma da cana-de-açúcar (HUANG; AITKEN; GEORGE, 2010). Os resultados apresentados no presente estudo estão em linha com aqueles apresentados por Jannoo et al. (1999), Raboin et al. (2008) e Wei et al. (2010), demonstrando que o desequilíbrio de ligação é um fenômeno comum no genoma de cana e que uma caracterização aprofundada de sua estrutura pode fornecer informações relevantes não só para a identificação de alelos associados a caracteres importantes, mas relevantes também para elucidar a evolução das cultivares modernas de cana-de-açúcar.

3.5 Análise de estruturação pelo programa STRUCTURE

Uma das principais preocupações no mapeamento por associação é a ocorrência de associações falsas resultantes da existência de subgrupos na população de mapeamento (PRITCHARD; ROSENBERG, 1999; PRITCHARD; STEPHENS; DONNELLY, 2000; YU et al., 2006; ZHAO et al., 2007). Mesmo que os indivíduos selecionados para compor a população sejam considerados não relacionados *a priori*, sempre haverá alguma estruturação subjacente, relacionada a algum fator desconhecido ou não considerado. Por exemplo, se uma dada característica for bem mais comum em alguns indivíduos na população e, se neste mesmo grupo a freqüência de um alelo for também mais alta, haverá associação estatística positiva entre este alelo e a característica sem que, na verdade, haja uma relação causal do ponto de vista genético. Portanto, a estruturação é resultante de freqüências alélicas diferenciais entre grupos de genótipos que compõem uma população sendo que cada grupo possui sua própria composição genética e esta pode não estar relacionada à característica que está sendo considerada no estudo. Todas as populações estão sujeitas a fatores estruturantes. Populações de melhoramento, por exemplo, passam por um continuado processo de seleção artificial, intrínseco à atividade. Populações naturais estão sujeitas, por exemplo, à seleção natural, deriva e migração. Assim, cada um destes fatores atua isoladamente ou em conjunto para moldar os diversos níveis de estruturação que podem ocorrer numa população.

Existem diversos métodos para lidar com esta questão como, por exemplo, Controle Genômico (DEVLIN; ROEDER, 1999), Análise de Componentes Principais (PCA) (PRICE et al., 2006), e métodos baseados em estatística bayesiana como o proposto por Corander et al., (2008) e implementado no programa BAPS e o método de PRITCHARD et al. (2000) implementado no programa STRUCTURE. A finalidade básica de todos estes métodos é reduzir ou eliminar a proporção de testes falsos positivos na análise de associação. Consideradas as diferenças metodológicas de cada um, não é tarefa simples determinar qual deles é o mais apropriado para cada situação, embora este seja um ponto de importância crítica numa análise de associação. Por exemplo, o Controle Genômico é bastante utilizado em estudos de associação em humanos devido à sua facilidade de aplicação em dados fenotípicos de natureza binária, comum em esquemas de análise conhecidos como TDT (*Transmission Desequilibrium Tests*) (KANG et al., 2008). Na abordagem de Componentes Principais os autores propõem o uso de uma matriz

contendo os principais eixos que explicam a maior parte da variação presente na população. Uma vantagem do método proposto por Corander et al. (2008) é que ele pode usar informações de seqüências de DNA ou dados de marcadores moleculares. Em linhas gerais este método é muito semelhante ao proposto por Pritchard et al. (2000). Atualmente o método implementado no programa STRUCTURE é, sem dúvida, o mais usado para o controle dos efeitos de estruturação. Neste trabalho, além da análise de agrupamento pelo método de DICE e da Análise de Componentes Principais (PCoA), ambos baseados em distâncias genéticas, utilizou-se também o programa STRUCTURE para avaliar a presença de estruturação na população de mapeamento por associação.

Uma etapa importante da análise é a definição do modelo de ancestralidade dos indivíduos na população, ou seja, deve ser informado ao programa se os indivíduos possuem origem dentro de um único grupo ou se eles possuem ancestralidade em vários grupos. Dado que as cultivares de cana-de-açúcar que compõem a população de mapeamento, independentemente do programa de melhoramento ao qual pertencem, compartilham genótipos ancestrais comuns definiu-se o modelo de ancestralidade mista (ad*mixed model*) como o mais adequado para a análise dos dados. Pela mesma razão, definiu-se o modelo de freqüências correlacionadas para o modelo de freqüências alélicas. Utilizando estes parâmetros e testando a hipótese de até sete grupos na população foram obtidos os resultados que estão apresentados nas Figuras 7 e 8, as quais mostram, respectivamente, o número e a composição dos grupos detectados. Os resultados foram bastante claros e evidenciam a presença de três grupos na população (K = 3). Para permitir uma comparação com os outros dois métodos a mesma nomenclatura dada aos grupos foi mantida. A determinação do número exato ou mais provável de grupos na população, a partir do resultado apresentado na Figura 7, consiste em identificar o valor de K para o qual o valor “Delta K” é máximo (EVANNO; REGNAULT; GOUDET, 2005).

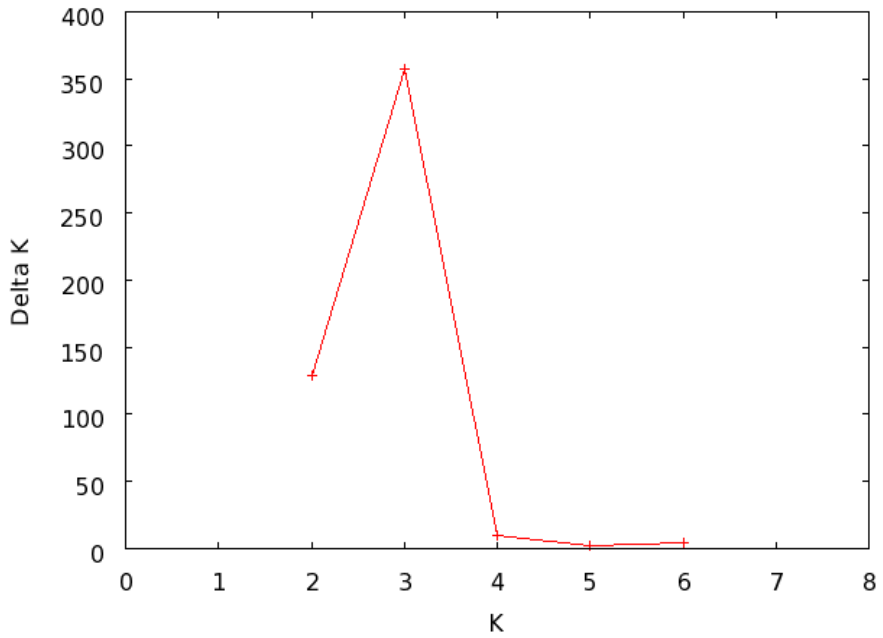


Figura 7 - Determinação do número de subgrupos na população de mapeamento por associação de acordo com o método proposto por Evanno, Regnault e Goudet (2005). No eixo x é representado o número provável de subgrupos K e no eixo y a estatística ΔK

A Figura 8 representa a distribuição das cultivares nos três grupos, bem como a ancestralidade expressa como a composição genômica de cada cultivar a partir dos K grupos identificados. Ela confirma também, com algumas exceções, a composição dos grupos em termos das cultivares presentes, conforme resultados apresentados no dendrograma (Figura 3) e Análise de Componentes Principais (Figura 4).

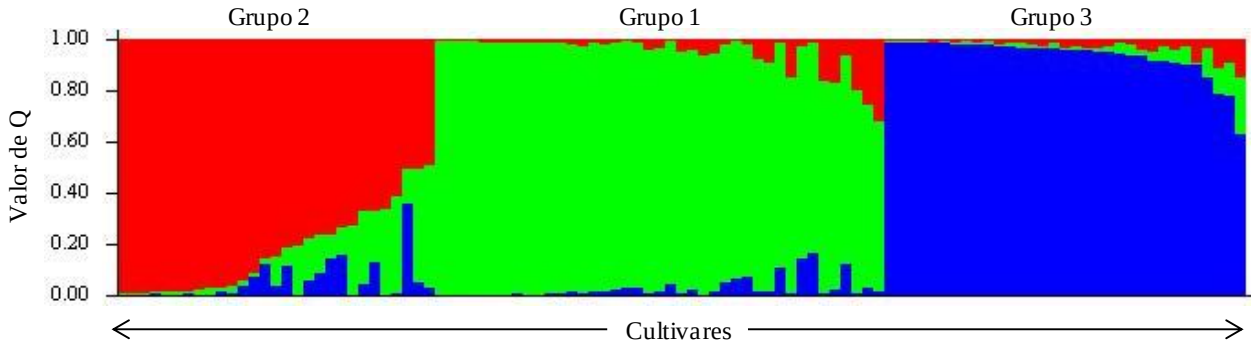


Figura 8 - Composição dos grupos identificados através do programa STRUCTURE. As cultivares estão dispostas ao logo do eixo horizontal e cada uma é representada por uma única linha vertical dividida em segmentos coloridos proporcionais à fração de seus genoma (valor Q) oriundo em um dos K grupos

Observa-se que todas as cultivares apresentaram alguma fração de seu genoma com origem em pelo menos um dos outros grupos, mas, obviamente, com a maior parte de seu genoma oriundo do próprio grupo ao qual foram alocadas. As diferenças na alocação de algumas cultivares em relação aos resultados das análises anteriores envolveram somente os grupos 1 e 2. Três cultivares, 'IACSP93-6006', 'IACSP95-2213' e 'RB85-5035', que haviam sido classificadas anteriormente no grupo 2, foram incluídas nesta análise no grupo 1. Duas delas apresentaram genomas fracionados quase pela metade entre os grupos 1 e 2: 'IACSP95-2213' (49%:47%) e 'RB85-5035' (50%:44%). Em alguns trabalhos é estabelecido um nível mínimo para declarar um genótipo como pertencente a determinado grupo (AGRAMA; EIZENGA; YAN, 2007; TOMMASINI et al., 2007; REIMER et al., 2008; ABDURAKHMONOV et al., 2009). Genótipos com valores abaixo do mínimo estabelecido são classificados como híbridos. Neste caso, as duas cultivares citadas seriam consideradas de origem híbrida. No entanto, neste estudo adotou-se como critério o maior valor Q apresentado pela cultivar para alocá-las em um dos grupos. O programa STRUCTURE utiliza uma abordagem baseada em freqüências alélicas para alocar os genótipos em cada grupo, enquanto que as análises do dendrograma e de componentes principais utilizam índices que expressam a distância genética entre os genótipos. Mesmo assim, as diferenças observadas não chegam a ser significativas o suficiente para alterar a estrutura já observada. Portanto, com base nesta e nas outras análises o número de grupos presentes na população de mapeamento foi igual a três.

Há poucos exemplos de aplicação deste método em cana-de-açúcar disponíveis na literatura. Wei et al. (2006) avaliaram a ocorrência de estruturação numa população de 154 clones do programa de melhoramento australiano BESES-CSIRO, mas relatam que não foi possível, com base apenas no resultado do STRUCTURE, definir o número exato de grupos. Os autores adotaram, então, utilizar a probabilidade posterior associada às estimativas de K ($ln\ Pr(X/K)$ que forneceu ao mesmo tempo a menor variação entre dois valores sucessivos de K e a matriz de ancestralidade Q que melhor correspondeu aos dados genealógicos, definindo o valor de K = 8. Segundo Wei et al. (2010) esta dificuldade pode estar relacionada à incapacidade do programa em lidar com as complexas inter-relações existentes nas populações de cana e que o uso de informações baseadas na genealogia das cultivares poderia ser suficiente para refletir esta complexidade.

3.6 Medidas dos coeficientes de coancestralidade

O método de associação de Yu et al. (2006) tem a vantagem de incorporar num único modelo diversos níveis de estruturação, sendo um deles as estruturas resultantes das relações de parentesco entre os indivíduos. A ocorrência de indivíduos relacionados genealogicamente resulta num tipo de estruturação difícil de determinar, principalmente quando os registros genealógicos não são precisos. Estas relações podem ser determinadas ou estimadas por meio do uso de marcadores para o cálculo de coeficientes de coancestralidade que refletem, na verdade, a proporção de alelos idênticos por descendência (IBD). Para os dados de cana-de-açúcar obtidos neste estudo foram calculados dois coeficientes expressando as relações de parentesco entre os indivíduos. A análise dos 850 alelos SSR e AFLP pelo programa SPAGeDI (HARDY; VEKEMANS, 2002) forneceu as estimativas dos coeficientes de coancestralidade (F_{ij}) entre os 103 genótipos do painel de associação de acordo com o método proposto por Hardy (2003) (Figura 9).

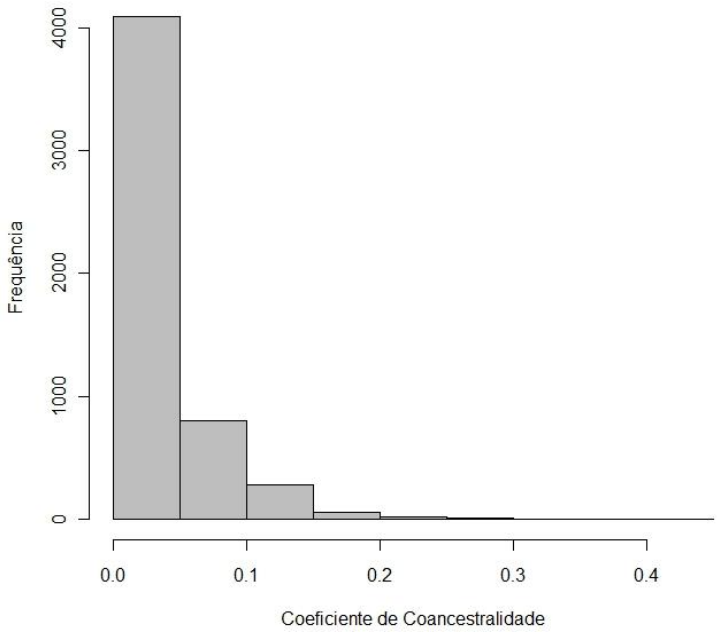


Figura 9 - Distribuição das estimativas dos coeficientes de coancestralidade (F_{ij}) calculados para os 103 genótipos de cana-de-açúcar pelo segundo o método proposto por Hardy (2003)

Os valores tendem a seguir um padrão contínuo de distribuição e evidenciam a existência de níveis elevados de relações de parentesco entre os indivíduos na população. Em termos numéricos, 41,4% das estimativas foram diferentes de zero, ou seja, houve um alto percentual de indivíduos mostrando algum nível de parentesco entre si. Este resultado está de acordo com as complexas relações de parentesco existente entre as cultivares de cana.

O cálculo dos coeficientes F_{ij} por meio do programa SPAGeDI parte de algumas pressuposições como, por exemplo, a de que genótipos parentais distantes são não relacionados. No contexto da cana isso pode não ser verdade na maioria dos casos, mas acredita-se que esta violação não causaria maiores conseqüências na análise de associação. Alternativamente, utilizou-se o coeficiente de Gower como uma medida das relações de parentesco entre os genótipos do painel de associação. Este índice é, na verdade, uma medida de similaridade genética entre dois genótipos e, neste caso, reflete a proporção de alelos idênticos por estado (IBS). Tem sido mostrado que em algumas situações o uso de medidas baseadas em distâncias genéticas é tão efetivo quanto as medidas obtidas pelo programa SPAGeDI na determinação das relações de parentesco entre os indivíduos (ZHAO et al., 2007; KANG et al., 2008).

3.7 Análises de associação

Uma das principais vantagens do mapeamento por associação é a possibilidade de utilizar as informações fenotípicas acumuladas ao longo de anos num programa de melhoramento. No entanto, pela própria natureza do processo as linhagens e as cultivares comerciais em avaliação mudam ao longo dos anos e nem sempre são avaliadas todas nos mesmos locais. Dados como esses são chamados não balanceados. Além disso, eles são afetados por interações com os ambientes bem como por variações espaciais nos experimentos. A obtenção de estimativas confiáveis dos efeitos genéticos é uma condição importante para se chegar a conclusões corretas a respeito das associações entre marcadores e caracteres. Modelos mistos são especialmente apropriados para essa finalidade uma vez que a variável de resposta pode ser expressa em função de fatores de efeitos fixos e aleatórios, utilizando diferentes estruturas de variância-covariância para ajustar o melhor modelo aos dados. Em função disso, tem sido sugerido o uso de BLUPs (*Best Linear Unbiased Predictors*) na análise de associação, em vez do uso direto dos dados fenotípicos brutos (ZHANG et al., 2009). Portanto, anteriormente à análise de associação os dados fenotípicos foram analisados usando a abordagem de modelos lineares mistos. As

estimativas dos BLUPs de cada cultivar nas quatro regiões e para cada um dos caracteres avaliados encontram-se nos Anexos B a G.

O modelo de associação entre as marcas e os caracteres fenotípicos utilizou o método proposto por Yu et al. (2006) e implementado no programa TASSEL (BRADBURY et al., 2007). Este método é baseado em modelos lineares mistos e sua grande vantagem é que ele é capaz de lidar com efeitos de múltiplos níveis de estruturação dentro da população. Estes efeitos foram representados pela matriz K, calculada no SPAGeDI pelo método de Hardy (2003) e no programa R pelo método de Gower (GOWER, 1971), e pela matriz Q de estrutura populacional, discriminando a estrutura de grupos da população, calculada no programa STRUCTURE. Como havia duas matrizes K, K_1 para a matriz calculada pelo programa SPAGeDI, e K_2 obtida através do índice de Gower, foram testadas, inicialmente, duas variações do modelo completo sendo a primeiro com os efeitos Q + K_1 e a segunda Q + K_2. Em seguida, os efeitos de estruturação Q foram removidos, testando-se modelos reduzidos contendo apenas os efeitos de parentesco, K_1 e K_2.

Associações significativas foram observadas em todos os modelos testados, para três dos seis caracteres avaliados (Nº de Colmos, Pol % Caldo e TCH) e em três das quatro regiões analisadas (Goiás, Piracicaba e Ribeirão Preto), incluindo três associações de caráter geral, ou seja, válida para todas as regiões (Tabela 8).

Considerando o modelo Q + K_1, observa-se que o caráter Número de Colmos mostrou-se positivamente associado com um alelo de 150 pb do loco *CIR56* na média das quatro regiões analisadas ($p < 0{,}0003529$). Deve-se ressaltar, no entanto, que não se trata de uma associação robusta visto que o valor q^* usado para comparação com o p valor admitiu uma proporção de falsos positivos de até 30%.

Ao utilizar a matriz de Gower para representar a estrutura de efeitos de parentesco no modelo Q + K_2, um segundo alelo de 283 pb, resultante de uma combinação de iniciadores AFLP (*ACG_CAT*), mostrou associação também com o Número de Colmos, mas somente para a região de Goiás. A robustez dessa associação ($p < 2{,}35 \times 10^{-4}$) foi embora baixa, mas ligeiramente mais forte do que a verificada anteriormente uma vez que se admitiu um valor $q^* \leq 20\%$.

A substituição da matriz de efeitos de parentesco K_1, calculada pelo programa SPAGeDI, para a matriz K_2, calculada com base no índice de similaridade de Gower, não teve nenhum impacto em relação à quantidade de associações detectadas. Além disso, os dois modelos completos foram capazes de identificar apenas associações moderadas entre os marcadores e os caracteres fenotípicos em cana-de-açúcar. Modelos que incorporam ao mesmo tempo as matrizes Q e K vêm sendo utilizados em estudos de associação com espécies alógamas e autógamas (SZALMA et al., 2005; AGRAMA; EIZENGA; YAN, 2007; CROSSA et al., 2007; GONZÁLEZ-MARTÍNEZ et al., 2007; IWATA et al., 2007; REIMER et al., 2008; YAO et al., 2009). No entanto, há relatos de que nem sempre seu desempenho é o melhor e que, em algumas situações, modelos reduzidos que utilizam apenas a matriz de efeitos de parentesco oferecem soluções melhores (CASA et al., 2008; STICH et al., 2008; WANG et al., 2008).

Tabela 8 - Marcadores SSR ou AFLP associados com caracteres fenotípicos em cana-de-açúcar de acordo com diferentes modelos de associação

Modelo	Loco/ Combinação	Alelo (pb)	Característica	Região	*p* valor	*q**
Q + K_1						
	CIR56	150	N° Colmos	Geral	$3{,}21 \times 10^{-4}$	0,30
Q + K_2						
	ACG_CAT	283	N° Colmos	Goíás	$2{,}00 \times 10^{-4}$	0,20
K_1						
	CV60	190	Pol % Cana	Ribeirão Preto	$1{,}10 \times 10^{-4}$	0,10
	CV106	156	Pol % Cana	Ribeirão Preto	$1{,}56 \times 10^{-3}$	0,20
	AAG_CAC	415	Pol % Cana	Ribeirão Preto	$1{,}12 \times 10^{-3}$	0,20
	AAG_CAC	207	Pol % Cana	Ribeirão Preto	$8{,}10 \times 10^{-4}$	0,20
	AAG_CAC	176	Pol % Cana	Ribeirão Preto	$1{,}16 \times 10^{-3}$	0,20
	AAG_CAG	439	Pol % Cana	Ribeirão Preto	$1{,}23 \times 10^{-3}$	0,20
	ACG_CTT	728	Pol % Cana	Ribeirão Preto	$4{,}97 \times 10^{-4}$	0,20
	ACG_CGC	462	TCH	Piracicaba	$1{,}82 \times 10^{-4}$	0,20
K_2						
	CV60	190	Pol % Cana	Ribeirão Preto	$2{,}06 \times 10^{-4}$	0,20
	ACG_CGT	324	N° Colmos	Geral	$8{,}99 \times 10^{-5}$	0,10
	AGG_CAG	212	N° Colmos	Geral	$1{,}95 \times 10^{-4}$	0,20

Dessa forma, quando se utilizou apenas a matriz de efeitos de parentesco do SPAGeDI (K_1) foram detectadas oito associações significativas envolvendo dois locos SSR e quatro diferentes combinações de iniciadores AFLP (Tabela 8). Sete das oito associações foram detectadas para a região de Ribeirão Preto, todas relacionadas à Porcentagem em massa de sacarose (Pol % cana). Para a região de Piracicaba houve uma associação significativa entre uma marca AFLP (*ACG_CGC* - 462 pb) e o caráter Tonelada de Cana por Hectare. De todas as associações, a que envolveu o loco *CV60* com o caráter PCC apresentou o *p* valor mais significativo, considerando o valor $q^* \leq 10\%$. Todas as demais associações foram válidas para valores de $q^* \leq 20\%$. Três alelos AFLP da combinação *AAG_CAC* (415, 207 e 176 pb) apresentaram associação com Porcentagem em massa de sacarose na região de Ribeirão Preto.

Quando foi utilizado o modelo que incorporou a matriz K_2, três associações significativas foram detectadas. Uma delas foi a mesma do modelo anterior, envolvendo o loco *CV60* e o caráter PCC na região de Ribeirão Preto. As outras duas associações envolveram dois alelos AFLP (*ACG_CGT* - 324 pb; *AGG_CAG* - 212 pb), ambas relacionadas ao caráter Número de Colmos para a média das quatro regiões.

Segundo Zhao et al. (2007), embora cada um dos dois termos do modelo completo proposto por Yu et al. (2006) (matrizes Q e K) se proponham a capturar diferentes efeitos de estruturação, não lhes parece clara a razão para uso de dois termos que, no final, são usadas com a mesma finalidade no modelo de associação. Para testar os efeitos de matrizes Q e K distintas no modelo de associação, estes autores obtiveram matrizes Q calculadas através do STRUCTURE e de análise de componentes principais (PCA) (PRICE et al., 2006) e também duas matrizes K alternativas, uma calculada pelo SPAGeDI e uma segunda baseada apenas na proporção de alelos compartilhados. A comparação entre os modelos completos do estudo de Zhao et al. (2007) com o modelo completo proposto por Yu et al. (2006) mostrou que, de modo geral, este modelo apresentou melhor desempenho, mas o uso de modelos utilizando apenas a matriz K calculada com base na proporção de alelos compartilhados forneceu resultados tão bons quanto os dos modelos completos. Em alguns casos, o modelo reduzido forneceu resultados inclusive melhores em termos de redução do número de associações falsas. Os resultados apresentados neste estudo, utilizando a proporção de alelos compartilhados por meio da matriz K_2 não se mostraram tão satisfatórios quanto os apresentados por Zhao et al. (2007). Stich et al.

(2008) recomendaram a utilização de uma matriz K expressando também a proporção de alelos compartilhados calculada conforme descrito por Bernardo et al. (1993). Os autores verificaram que o uso desta matriz K contribuiu para aumentar o poder na detecção de associações.

Por suas vantagens existe grande interesse na utilização do mapeamento por associação para identificar genes associados a caracteres agronomicamente importantes em cana-de-açúcar. Os elevados níveis de desequilíbrio de ligação observados neste estudo e nos relatados na literatura advogam a favor dessa metodologia na busca por tais associações. É claro que mais estudos são necessários para estabelecer se tais níveis são específicos para cada população ou se, no caso de cana, eles são válidos para toda a espécie. Esse conhecimento será fundamental para se dimensionar a saturação de marcas necessária para que associações robustas sejam identificadas. Além disso, a compreensão exata da estruturação na população é crítica para determinar o sucesso desta abordagem na identificação de associações verdadeiras. A contribuição de cada um dos fatores estruturantes deve ser cuidadosamente avaliada, principalmente porque, no caso da cana, o efeito fundador em sua história de melhoramento e a ampla utilização de genótipos com ascendência comum contribuíram para tornar bastante complexas as relações de parentesco entre as cultivares modernas. Esse talvez seja o principal fator que explica a dificuldade encontrada em alguns trabalhos para a determinação de grupos bem definidos (WEI et al., 2006; RABOIN et al., 2008; WEI et al., 2010). Além disso, a ocorrência de aneuploidia também pode contribuir para tornar ainda mais complexas essas relações, uma vez que cultivares que compartilham parentais em comum podem receber diferentes complementos cromossômicos como resultado desse fenômeno.

Embora os modelos mistos sejam considerados os mais robustos para a análise de associação genética e assumindo que há elevados níveis de desequilíbrio na população estudada, com base nos resultados apresentados para a população de mapeamento de ligação, o baixo número de associações significativas detectada entre os marcadores e os caracteres avaliados neste trabalho poderia ser explicado pelo baixo número de marcadores disponíveis. Embora o desequilíbrio de ligação estenda-se por distâncias de até 30 cM no genoma de cana (JANNOO et al., 1999a; RABOIN et al., 2008) ou mais (40 cM no caso deste trabalho), essas correlações caem de forma de forma significativa para níveis menores em distâncias em torno de 5 cM (RABOIN et al., 2008, WEI et al., 2010). Conforme mostrado por Wei et al. (2006) com esses níveis de

desequilíbrio uma quantidade moderada de marcadores (1.209) foi suficiente para permitir a identificação de um número grande de associações. Raboin et al. (2008), considerando que o genoma haplóide de cana possui cerca de 1.500 cM, estimaram que seriam necessários de 2.000 a 4.000 marcadores para se atingir uma densidade de 1 a 2 marcadores a cada 5 cM.

4 CONCLUSÕES

1) Foram observados níveis elevados de desequilíbrio de ligação (40 cM) no genoma das cultivares de cana-de-açúcar, compatíveis com os já relatados na literatura. A exploração desta característica deve permitir que genes associados com caracteres fenotípicos de interesse sejam identificados à medida que mais estudos forem realizados. Além disso, essa pode ser uma ferramenta útil para ajudar a esclarecer melhor a contribuição de cada espécie para a formação do genoma das variedades modernas de cana.

2) Apesar da dificuldade em determinar com precisão os efeitos de estruturação resultantes das relações de parentesco em cana-de-açúcar, o mapeamento por associação é factível nesta espécie, visto que houve indícios de associação envolvendo 10 alelos (*CIR56*, *ACG_CAT*, *CV60*, *CV106*, *AAG_CAC*, *AAG_CAG* , *ACG_CTT*, *ACG_CGC*, *ACG_CGT* e *AGG_CAG*).

REFERÊNCIAS

ABDURAKHMONOV, I.Y.; Saha S.; JENKINS, J.N.; BURIEV, Z.T.; SHERMATOV, S.E.; SCHEFFLER, B.E.; PEPPER, A.E.; YU, J.Z.; KOHEL, R.J.; ABDUKARIMOV, A. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. **Genetica**, Dordrecht, v. 136, n. 3, p. 401–417, 2009.

AGRAMA, H.A.; EIZENGA, G.C.; YAN, W. Association mapping of yield and its components in rice cultivars. **Molecular Breeding**, Dordrecht , v. 19, n. 4, p. 341–356, 2007.

AITKEN, K.S.; McNEIL, M. Diversity analysis. In: HENRY, R.; CHITTARANJAN, K. (Ed.). **Genetics, genomics and breeding of sugarcane**. New Hampshire: Science Publishers, 2010. chap. 3, p. 19-42.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in sugarcane cultivar. **Theoretical and Applied Genetics**, Berlin, v. 110, n. 5, p. 789-801, 2005.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar x *Saccharum officinarum* population. **Theoretical and Applied Genetics**, Berlin, v. 112, n. 7, p. 1306-1317, 2006.

AITKEN, K.S.; JACKSON, P.A.; McINTYRE, C.L. Construction of a genetic linkage map of *Saccharum officinarum* incorporarting both simplex and duplex markers to increase genome coverage. **Genome**, Ottawa, v. 50, n. 8, p. 742-756, 2007.

AITKEN, K.S.; LI, J.C.; JACKSON, P.A.; PIPERIDIS, G.; McINTYRE, C.L. AFLP analysis of genetic diversity within *Saccharum officinarum* and comparison with sugarcane cultivars. **Australian Journal of Agricultural Research**, Melbourne, v. 57, n. 11, p. 1167-1184, 2006.

AITKEN, K.S.; HERMANN, S.; KARNO, K.; BONNETT, G.D.; McINTYRE, C.L.; JACKSON, P.A. Genetic control of yield related stalk traits in sugarcane. **Theoretical and Applied Genetics**, Berlin, v. 117, n. 7, p. 1191-1203, 2008.

ALBINO, J.C. **Mapeamento genético em cana-de-açúcar (*Saccharum* spp.) utilizando sondas heterólogas de sorgo e genes análogos de resistência**. 2006, 121p. Dissertação (Mestrado em Ciências) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2006.

ALIX, K.; PAULET, F.; GLASZMANN, J.C.; D'HONT, A. Inter-Alu like species-specific sequences in the *Saccharum* complex. **Theoretical and Applied Genetics**, Berlin, v. 99, n. 6, p. 962-968, 1999.

ALIX, K.; BAURENS, F.C.; PAULET, F.; GLASZMANN, J.C.; D'HONT, A. Isolation and characterization of a satellite DNA family in the *Saccharum* complex. **Genome**, Ottawa, v. 41, n. 6, p. 854-864, 1998.

AL-JANABI, S.M.; FORGET, L.; DOOKUN, A. An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. **Plant Molecular Biology Reporter**, Netherlands, v. 17, n. 3, p. 1-8, 1999.

AL-JANABI, S.M.; HONEUCUTT, R.J.; McCLELLAND, M.; SOBRAL, B.W.S. A genetic linkage map of *Saccharum spontaneum* L. 'SES208'. **Genetics**, Baltimore, v. 134, n. 4, p. 1249-1260, 1993.

AL-JANABI, S.M.; PARMESSUR, Y.; KROSS, H.; DHAYAN, S.; SAUMTALLY, S.; RAMDOYAL, K.; AUTREY, L.J.C.; DOOKUN-SAUMTALLY, A. Identification of a major quantitative trait locus for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. **Molecular Breeding**, Berlin , v. 19, n. 1, p. 1-14, 2007.

ALWALA, S; KIMBENG, C.A. Molecular genetic linkage mapping in *Saccharum*: Strategies, resources and achievements. In: HENRY, R.; CHITTARANJAN, K. (Ed.). **Genetics, genomics and breeding of sugarcane**. New Hampshire: Science Publishers, 2010. chap. 5, p. 69-96.

ALWALA, S; KIMBENG, C.A.; VEREMIS, J.C.; GRAVOIS, K.A. Linkage mapping and genome analysis in a Saccharum interespecific cross using AFLP, SRAP and TRAP markers. **Euphytica**, Dordrecht, v. 164, n. 1, p. 37-51, 2008.

ALWALA, S.; SUMAN, A.; ARRO, J.A.; VEREMIS, J.C.; KINBENG, C.A. Target region amplification polymorphism for assessing genetic diversity in sugarcane germplasm collections. **Crop Science**, Madison, v. 46, n. 1, p. 448-445, 2006.

ARCENEAUX, G. Cultivated sugarcanes of the world and their botanical derivation. **Proceedings of the International Society of Sugar Cance Technologists**, San Juan, v. 12, p. 844-854, 1965.

ASNAGHI, C.; PAULET, F.; KAYE, C.; GRIVET, L.; DEU, M.; GLASZMANN, J.C.; D'HONT, A. Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. **Theoretical and Applied Genetics**, Berlin, v. 101, n. 5-6, p. 962-969, 2000.

AULCHENKO, Y.S.; KONING, D.; HALEY, C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. **Genetics**, Baltimore, v. 177, n. 1, p. 577-585, 2007.

BATES, D.M.; MAECHLER M. **lme4**: Linear mixed-effects models using S4 classes. R package version 2.12.1. Disponível em: <http://www.r-project.org/>. Acesso em: 15 fevereiro 2010.

BENJAMINI, Y.; HOCHBERG, Y. Controling the false discovery rate: a practical and powerful approach to multiple testing. **The Journal of the Royal Statistical Society**. **Series B**, Oxford, v. 57, n. 1, p. 289-300, 1995.

BERNARDO, R. Estimation of coefficient of coancestry using molecular markers in maize. **Theoretical and Applied Genetics**, New York, v. 85, n. 8, p. 1055-1062, 1993.

BERNARDO, R.; MURIGNEUX, A.; KARAMAN, Z. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. **Theoretical and Applied Genetics**, Berlin, v. 93, n. 1-2, p. 262-267, 1996.

BESSE, P.; McINTYRE, C.L.; BERDING, N. Characterization of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae Saccharinae) using RFLP markers. **Euphytica**, Dordrecht, v. 93, n. 3, p. 283-292, 1997.

BESSE, P.; TAYLOR, G.; CARROLL, B.; BERDING, N.; BURNER, D.; McINTYRE, C.L. Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis**. Genetica**, Dordrecht, v. 104, n. 2, p. 143-153, 1998.

BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVENS, T.M.; RAMDOSS, Y.; BUCKLER, E.S. **TASSEL**: software for association mapping of complex traits in diverse samples. **Bioinformatics**, Oxford, v. 23, n. 19, p. 2633-2635, 2007.

BRASIL. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. **Intercâmbio comercial do agronegócio**: principais mercados de destino. Brasília: MAPA/ACS, 2010. 443 p. Disponível em: <http://www.agricultura.gov.br/>. Acesso em: 15 março 2011a.

_____. **Anuário estatístico da agroenergia.** Brasília: MAPA/ACS, 2009. 160 p. Disponível em: <http://www.agricultura.gov.br/>. Acesso em: 07 fevereiro 2011b.

BRASIL. MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO. **Exportações brasileiras de etanol**. Disponível em: < http://www.mdic.gov.br/>. Acesso em: 12 março 2011c.

BRESSIANI, J.A. **Herdabilidade e repetibilidade dos componentes da produção na cultura da cana-de-açúcar**. 1993. Dissertação (Mestrado em Agronomia) - Universidade de São Paulo, 1993.

BRESEGHELLO, F.; SORRELLS, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics**, Baltimore, v. 172, n. 2, p. 1165-1177, 2006.

BROWN, J.S.; SCHENELL, R.J.; POWER, E.J.; DOUGLAS, S.L.; KUHN, D.N. Analysis of clonal germplasm from five *Saccharum* species: *S. barberi*, *S. robustum*, *S. officinarum*, *S.*

sinense and *S. spontaneum*. A study of inter- and intra species relationships using microsatellite markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 54, n. 3, p. 627–648, 2007.

BUIL, A.; MARTINEZ-PEREZ, A.; PERERA-LLUNA, A.; RIB, L.; CAMINAL, P.; SORIA, J.M. A new gene-based association test for genome-wide association studies. **BMC Proceedings**, St. Louis, v. 3, suppl. 7, S130, 15 p., 2009.

BURNQUIST, W.L.; SORRELLS, M.E.; TANKSLEY, S. Characterization of genetic variability in *Saccharum* germplasm by means of restriction fragment length polymorphism (RFLP) analysis. **Proceedings of the International Society of Sugar Cane Technologists**, Bangkok, v. 2, p. 355-365, 1992.

BUTTERFIELD, M.K. **Marker assisted breeding in sugarcane: a complex polyploid**. 2007, 162 p. Tese (PhD Thesis, Doutorado em Filosofia) - University of Stellenbosch, Matienland, Stellenbosch, South Africa, 2007.

BUTTERFIELD, M.K.; D'HONT, A.; BERDING, N. The sugarcane genome: a synthesis of current understanding and lessons for breeding and biotechnology. **Proceedings of the South African Sugarcane Technologists Association**, KwaZulu-Natal, v. 75, p.1-5, 2001.

CALSA Jr., T. **Análise transcricional de folha e parênquima de colmo de genótipos de cana-de-açúcar (*Saccharum* spp.) contrastantes para teor de sacarose e estágio de maturação por meio serial (SAGE)**. 2005, 209p. Tese (Doutorado em Ciências) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2005.

CASA, A.M.; PRESSOIR, G.; BROWN, P.J.; MITCHELL, S.E,; ROONEY, W.L.; TUINSTRA, M.R.; FRANKS, C.D.; KRESOVICH, S. Community resources and strategies for association mapping in sorghum. **Crop science**, Madison, v. 48, n. 1, p. 30-40, 2008.

CASU, R.E.; DIMMOCK, C.M.; CHAPMAN, S.C.; GROF, C.P.; McINTYRE, C.L.; BONNETT, G.G.; MANNERS, J.M. Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. **Plant Molecular Biology**, Dordrecht, v. 54, n. 4, p. 503-517, 2004.

CENTRO DE TECNOLOGIA CANAVIEIRA. Disponível em: <http://www.ctcanavieira.com.br/>. Acesso em: 14 abril 2011.

CESNIK, R.; MIOCQUE, J. **Melhoramento da cana-de-açúcar**. Brasília: Embrapa Informação Tecnológica, 2004. cap. 1, p. 23-30.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileira**: Cana-de-açúcar, 3° Levantamento, Brasília, 2011, 19 p.

CORANDER, J.; MARTTINEN, P.; SIREN, J.; TANG, J. Enhanced bayesian modelling in BAPS software for learning genetic structures of populations. **BMC Bioinformatics**, Helsinki, v. 9, artigo n. 539, 14 p., 2008.

CORDEIRO, G.M.; TAYLOR, G.O.; HENRY, R.J. Characterization of microsatellite markers from sugarcane (*Saccharum* spp.) a highly polyploid species. **Plant Science**, v. 155, n. 2, p. 161-168, 2000.

CORDEIRO, G.M.; PAN, Y.B.; HENRY, R.J. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. **Plant Science**, Madison, v. 165, n. 1, p. 181-189, 2003.

CORDEIRO, G.M.; MAGUIRE, T.L.; HENRY, R.J.; EDWARDS, K.J. Optimisation of a microsatellite enrichment technique in *Saccharum* spp. **Plant Molecular Biology Reports**, Piscataway, v. 17, n. 3, p. 225-229, 1999.

CRESTE, S.; ACCORONI, K.A.G.; PINTO, L.R.; VENCOVSKY, R.; GIMENES, M.A.; XAVIER, M.A.; LANDELL, M.G.A. Genetic variability among sugarcane genotypes based on polymorphisms in sucrose metabolism and drought tolerance genes. **Euphytica**, Dordrecht, v. 172, n. 3, p. 435-446, 2010.

CROSSA, J.; BURGUEÑO, J.; DREISIGACKER, S.; VARGAS, M.; HERRERA-FOSSEL, S.A.; LILLEMO, M.; SINGH, R.P.; TRETHOWAN, R.; WARBURTON, M.; FRANCO, J.; REYNOLDS, M.; CROUCH, J.H.; ORTIZ, R. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. **Genetics**, Baltimore, v. 177, n. 3, p. 1889-1913, 2007.

Da SILVA, J.A.G.; SORRELLS, M.E.; BURNQUIST, W.L.; TANKSLEY, S.D. RFLP linkage map and genome analysis of *Saccharum spontaneum*. **Genome**, Ottawa, v. 36, n. 4, p. 782-791, 1993.

Da SILVA, J.A.G.; HONEYCUTT, R.J.; BURNQUIST, W.L; Al-JANABI, S.M.; SORRELS, M.E.; TANKSLEY, S.D.; SOBRAL, B.W.S. *Saccharum spontaneum* L. 'SES208' genetic linkage map containing RFLP and PCR-based markers. **Molecular Breeding**, Berlin, v. 1, n. 2, p. 165-179, 1995.

DAMAJ, M.B.; BEREMAND, P.D.; BUENROSTRO-NAVA, M.T.; IVY, J.; KUMPATLA, S.P.; JIFON, J.; BEYENE, G.; YU, Q.; THOMAS, T.L.; MIRKOV, T.E. Isolating promoters of multigene family members from the polyploidy sugarcane genome by PCR-based walking in BAC DNA. **Genome**, Ottawa, v. 53, n. 10, p. 840-847, 2010.

DANIELS, J.; ROACH, B.T. Taxonomy and evolution. In: HEINZ D.J. (Ed.). **Sugarcane Improvement through Breeding**. Amsterdam: Elsevier Press, 1987. Chap. 2, p. 7-84.

DAUGROIS, J.H.; GRIVET, L.; ROQUES, D.; HOARAU, J.Y.; LOMBARDI, H.; GLASZMANN, J.C.; D'HONT, A. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. **Theoretical and Applied Genetics**, New York, v. 92, n. 8, p. 1059-1064, 1996.

DEVLIN, B.; ROEDER, K. Genomic control for association studies. **Biometrics**, Washington, v. 55, n. 4, p. 997-1004, 1999.

D'HONT, A.; GLASZMANN, J.C. Sugarcane genome analysis with molecular markers, a first decade of research. **Procedings of the International Society of Sugar Cane Technologists**, Brisbane, v. 24, p. 556-559, 2001.

D'HONT, A.; PAULET, F.; GLASZMANN, J.C. Oligoclonal interspecific origin of "North Indian" and "Chinese" sugarcanes. **Chromosome Research**, Dordrecht, v. 10, n. 3, p. 253-262, 2002.

D'HONT, A.; LU, Y.H.; FELDMANN, P.; GLASZMANN, J.C. Cytoplasmic diversity in sugarcane revealed by heterologous probes. **Sugar Cane**, v. 1, p. 12-15, 1993.

D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular and General Genetics**, Berlin, v. 250, n. 4, p. 405–413, 1996.

D'HONT, A.; RAO, P.; FELDMANN, P.; GRIVET, L.; ISLAM-FARIDI, N.; TAYLOR, P.; GLASZMANN, J.C. Identification and characterisation of intergeneric hybrids, *S. officinarum* x *Erianthus arundinaceus*, with molecular markers and in situ hybridization. **Theoretical and Applied Genetics**, Berlin, v. 91, n. 2, p. 320-326, 1995.

EARL, D.A. Structure harvester v 0.6.1. 2011. Disponível em: <http://taylor0.biology.ucla.edu/struct_harvest/>. Acesso em: 16 março 2011.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, Oxford, v. 14, n. 8, p. 2611-2620, 2005.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4 ed. Essex: Longmann, 1996. 464 p.

FALUSH, D.; STEPHENS, M.; PRITCHARD, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. **Genetics**, Baltimore, v. 164, n. 4, p. 1567-1587, 2003.

FALUSH, D.; STEPHENS, M.; PRITCHARD, J.K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. **Molecular Ecology Notes**, Oxford, v. 7, n. 4, p. 574-578, 2007.

FLINT-GARCIA, S.; THORNSBERRY, J.M.; CUCKLER, E.S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, Palo Alto, v. 54, p. 357-374, 2003.

GARCIA, A.A.F.; KIDO, E.A.; MEZA, N.A; 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. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross based on a maximum-likelihood approach. **Theoretical and Applied Genetics**, New York, v. 112, n. 2, p. 298-314, 2006.

GASCUEL, O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. **Molecular Biology and Evolution**, Lawrence, v. 14, n. 7, p. 685–695, 1997.

GLASZMANN, J.C.; LU, Y.H.; LANAUD, C. Variation of nuclear ribosomal DNA in sugarcane. **Journal of Genetics and Breeding**, Roma, v. 44, n. 3, p. 191-198, 1990.

GLASZMANN, J.C.; FAUTRET, A.; NOYER, J.L.; FELDMANN, P.; LANAUD, C. Biochemical genetic markers in sugarcane. **Theoretical and Applied Genetics**, New York, v. 78, n. 4, p. 537-543, 1989.

GONZÁLEZ-MARTÍNEZ, S.C.; WHEELER, N.C.; ERSOZ, E.; NELSON, C.D.; NEALE, D.B. Association genetics in *Pinus taeda* L. I. wood property traits. **Genetics**, Baltimore, v. 175, n. 1, p. 399-409, 2007.

GOWER, J.C. A general coefficient of similarity and some of its properties. **Biometrics**, Washington, v. 27, n. 4, p. 623-637, 1971.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, London, v. 5, n. 2, p. 122-127, 2001.

GRIVET, L.; DANIELS, C.; GLASZMANN, J.C.; D'HONT, A. A review of recent molecular genetics evidence for sugarcane evolution and domestication. **Ethnobotany Research and Applications**, Fort Worth, v. 2, p. 9-17, 2004.

GRIVET, L.; D'HONT, A.; ROQUES, D.; FELDMANN, P.; LANAUD, C.; GLASZMANN, J.C. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. **Genetics**, Bethesda, v. 142, n. 3, p. 987-1000, 1996.

GUIMARÃES, C.T.; HONEYCUTT, R.J.; SILLS, G.R.; SOBRAL, B.W.S. Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. ex grassl. **Genetics and Molecular Biology**, Ribeirão Preto, v. 22, n. 1, p. 125-132, 1999.

GUPTA, P.K.; RUSTGI, S.; KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. **Plant Molecular Biology**, Leiden, v. 57, n. 4, p. 461-485, 2005.

HARDY, O.J. Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. **Molecular Ecology**, Oxon, v. 12, n. 6, p. 1577-1588, 2003.

HARDY O.J.; VEKEMANS X. **SPAGeDi**: a versatile computer program to analyse spatial genetic structure at the individual or population levels. **Molecular Ecology Notes**, Oxford, v. 2, n. 4, p. 618-620, 2002.

HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Washington, v. 31, n. 2, p. 423-447, 1975.

HOARAU, J.Y.; OFFMAN, B.; D'HONT, A.; RISTERUCCI, A.M.; ROQUES, D.; GLASZAMNN, J.C.; GRIVET, L. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). I. Genome mapping with AFLP markers. **Theoretical and Applied Genetics**, New York, v. 103, n. 1, p. 84-97, 2001.

HOARAU, J.Y.; GRIVET, L.; OFFMAN, B.; RABOIN, L.M.; DIORFLAR, J.P.; PAYET, J.; HELLMANN, M.; D'HONT, A.; GLASZAMNN, J.C. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. **Theoretical and Applied Genetics**, New York, v. 105, n. 6-7, p. 1027-1037, 2002.

HUANG, E.; AITKEN, K.S.; GEORGE, A. Association studies. In: HENRY, R.; CHITTARANJAN, K. (Ed.). **Genetics, genomics and breeding of sugarcane**. New Hampshire: Science Publishers, 2010. chap. 4, p. 43-68.

HUBISZ, M.J.; FALUSH, D.; STEPHENS, M.; PRITCHARD, J.K. Inferring weak population structure with the assistance of sample group information. **Molecular Ecology Resources,** Malden, v. 9, n. 5, p. 1322-1332, 2009.

INSTITUTO AGRONÔMICO DE CAMPINAS. Disponível em: < http://www.iac.br/>. Acesso em: 14 abril 2011.

IWATA, H.; UGA, Y.; YOSHIOKA, Y.; EBANA, K.; HAYASHI, T. Bayesian association mapping of multiple quantitative loci and its application to the analysis of genetic variation among *Oryza sativa* L. germoplasms. **Theoretical and Applied Genetics**, New York, v. 114, n.8, p. 1437-1449, 2007.

JACKSON, P.A. Breeding for improved sugar content in sugarcane. **Field Crops Research**, Amsterdam, v. 92, n. 2-3, p. 277-290, 2005.

JANNOO, N.; GRIVET, L.; DOOKUN, A.; D'HONT, A.; GLASZMANN, J.C. Linkage disequilibrium among modern sugarcane cultivars. **Theoretical and Applied Genetics**, New York, v. 99, n. 6, p. 1053-1060, 1999a.

JANNOO, N.; GRIVET, L; SEGUIN, M.; PAULET, F.; DOMAINGUE, R.; RAO, P.S.; DOOKUN, A.; D'HONT, A.; GLASZMANN, J.C. Molecular investigation of the genetic base of sugarcane cultivars. **Theoretical and Applied Genetics**, New York, v. 99, n.1-2, p. 171-184, 1999b.

JORDÃO Jr., H. **Desenvolvimento de um sistema baseado em marcadores moleculares de DNA do tipo microssatélites para identificação de variedades de cana-de-açúcar**. 2009. 106p. Dissertação (Mestrado em Biologia Funcional e Molecular) - Universidade Estadual de Campinas, Campinas, 2009.

JUN, T-H.; VAN, K.; KIM, M.Y.; LEE, S.H.; WALKER, D.R. Association analysis using SSR markers to find QTL for seed protein content in soybean. **Euphytica**, Dordrecht, v. 162, n. 2, p. 179-191, 2008.

KANG, H.M.; ZAITLEN, N.A.; WADE, C.M.; KIRBY, A.; HECKERMAN, D.; DALY, M.J.; ESKIN, E. Efficient control of population structure in model organism association mapping. **Genetics**, Madison, v. 178, n. 3, p. 1709-1723, 2008.

KANG, H.M.; SUL, J.H.; SERVICE, S.K.; ZAITLEN, N.A.; KONG, S-Y.; FREIMER, N.B.; SABATTI, C.; ESKIN, E. Variance component model to account for sample structure in genome-wide association studies. **Nature Genetics**, Stanford, v. 42, n. 4, p. 348-356, 2010.

KEREM, B.S.; ROMMENS, J.M.; BUCHANAN, J.A.; MARKIEWCZ, D.; COX, T.K.; CHAKRAVARTI, A.; BUCHWALD, M. Identification of the cystic fibrosis gene: genetic analysis. **Science**, Washington, v. 245, n. 4922, p. 1073-1080, 1989.

KIDO, E.A. **Mapeamento de marcadores moleculares AFLP em população derivada de clones elite de cana-de-açúcar (*Saccharum* spp.) e suas associações com caracteres agronômicos**. 2003**.** 164p. Tese (Doutorado em Ciências) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2003.

KOSAMBI, D.D. The estimation of map distance from recombination values. **Annual Eugene**, London v. 12, p. 172-175, 1944.

LAKSHMANAN, P.; GEIJSKES, J.; AITKEN, K.S.; GROF, C.L.P.; BONNETT, G.D.; SMITH, G.R. Sugarcane biotechnology: the challenges and opportunities. **In Vitro Cellular & Developmental Biology**, Oxon, v. 41, n. 4, p. 345-363, 2005.

LANDELL, M.G.A.; BRESSIANI, J.A. Melhoramento genético, caracterização e manejo varietal. In: DINARDO-MIRANDA, L.L.; VASCONCELOS, A.C.M.; LANDELL, M.G.A. (Ed.). **Cana-de-açúcar**. Campinas: Instituto Agronômico, 2008. cap. 5, p. 101-155.

LIMA M.L.A.; GARCIA A.A.F; OLIVEIRA K.M; MATSUOKA S.; SOUZA Jr. C.L.; SOUZA A.P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum spp.*). **Theoretical and Applied Genetics**, New York, v. 104, n. 1, p. 30-38, 2002.

LIU, K.; MUSE, S.V. **PowerMarker v. 3.0**: integrated analysis environment for genetic marker data**. Bioinformatics**, Oxford, v. 21, n. 9, p. 2128-2129, 2005.

LU, Y.H.; D'HONT, A.; WALKER, D.I.T.; RAO, P.S.; FELDMANN, P.; GLASZAMNN, J.C. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. **Euphytica**, Dordrecht, v. 78, n. 1-2, p. 7-18, 1994a.

LU, Y.H.; D'HONT, A.; PAULET, F.; GRIVET, L.; ARNAUD, M.; GLASZMANN, J.C. Molecular diversity and genome structure in modern sugarcane varieties. **Euphytica**, Dordrecht, v. 78, n. 3, p. 217-226, 1994b.

MACCHERONI, W; JORDÃO, H; DEGASPARI, R; MATSOUKA, S. Development of a dependaple microsatellite-based fingerprinting system for sugarcane. **Proceedings of the International Society of Sugar Cane Technologists**, Durban, v. 26, p. 889-900, 2007.

MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, London, v. 12, n. 2, p. 1360-1385, 2006.

MALOSETTI, M.; LINDEN, C.G.; VOSMAN, B.; EEUWIJK, F.A. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. **Genetics**, Baltimore, v. 175, n. 2, p. 879-889, 2007.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. **Melhoramento de espécies cultivadas**. Viçosa: Editora da UFV, 2005. p. 225-274.

McINTYRE, C.L.; WHAN, V.A; CROFT, B.; MAGAREY, R.; SMITH, G.R. Identification and validation of molecular markers associated with Pachymetra root rot and brown rust resistance in sugarcane using map- and association-based approaches. **Molecular Breeding**, Dordrecht, v. 16, n. 2, p. 151-161, 2005a.

McINTYRE, C.L.; CASU R.E.; DRENTH, J.,KNIGTH, D.; WHAN, V.A.; CROFT, B.J.; JORDAN, D.R.; MANNERS, J.M. Resistance gene analogues in sugarcane and their association with quantitative trait loci for rust resistance. **Genome**, Ottawa, v. 48, n. 3, p. 391-400, 2005b.

MING, R.; LIU, S.C.; BOWERS, J.E.; IRVINE, J.E.; PATERSON, A.H. Construction of a *Saccharum* consensus genetic map from two interspecific crosses. **Crop Science**, Madison, v. 42, n. 2, p. 570-583, 2002a.

MING, R.; SIN-CHIEH, L.; MOORE, P.H.; IRVINE, J.E; PATERSON, A.H. QTL Analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. **Genome Research**, New York, v. 11, n. 12, p. 2075-2084, 2001.

MING, R.; LIU, S-C.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. **Genome Research**, New York, v. 11. n. 12, p. 2075-2084, 2002c.

MING, R.; WANG, Y.W.; DRAYE, X.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. **Theoretical and Applied Genetics**, New York, v. 105, n. 2-3, p. 332-345, 2002b.

MING, R; MOORE, P.H; WU, K-K.; D'HONT, A.; GLASZMANN, J.C.; TEW, T.L.; MIRKOV, T.E.; Da SILVA, J.; JIFON, J.; RAI, M.; SCHNELL, R.J.; BRUMBLEY, S.M.; LAKSHMANAN, P.; COMSTOCK, J.C.; PATERSON, A.H. Sugarcane Improvement through Breeding and Biotechnology. In: JANICK, J. (Ed.). **Plant breeding reviews**. New York: John Wiley & Sons, 2006. chap. 2, p. 15-118.

MING, R.; LIU, SC.; LIN, Y.R.; Da SILVA, J.; WILSON, W.; BRAGA, D.; VAN DEYNZE, A.; WENSLAFF, T.F.; WU, K.K.; MOORE, P.H.; BURNQUIST, W.; SORRELS, M.E.; IRVINE, J.E.; PATERSON, A.H. Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and poliploid genomes. **Genetics**, Bethesda, v. 150, n. 4, p. 1663-1682, 1998.

MISSIAGGIA A.; GRATTAPAGLIA D. Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. **Genetics and Molecular Research**, Ribeirão Preto, v. 5, n. 1, p. 72-78, 2006.

MUDJE, J.; ANDERSON, W.R.; KEHRE, R.L.; FAIRBANKS, D.J. A RAPD genetic map of *Saccharum officinarum*. **Crop Science**, Madison, v. 36, n. 5, p. 1362-1366, 1996.

NAIR, N.V.; SREENIVASAN, T.V.; MOHAN, M. Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 46, n. 1, p. 73-79, 1999.

NEI, M.; LI, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences**, Washington, v. 76, n. 10, p. 5269–5273, 1979.

NORDBORG, M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. **Genetics**, Baltimore, v. 154, n. 2, p. 923-929, 2000.

NORDBORG, M.; TAVARÉ, S. Linkage disequilibrium: what history has to tell us. **Trends in Genetics**, London, v. 18, n. 2, p. 83-90, 2002.

NORDBORG, M.; BOREVITZ, J.O.; BERGELSON, J.; BERRY, C.C.; CHORY, J. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature Genetics**, New York, v. 30, n. 2, p. 190-193, 2002.

OLIVEIRA, K.M.; PINTO, L.R.; MARCONI, T.G.; MARGARIDO, G.R.A.; PASTINA, M.M.; TEIXEIRA, L.H.M.; FIGUEIRA, A.V.; ULIAN, E.C.; GARCIA, A.A.F.; SOUZA, A.P. Functional integrated genetic linkage map based on EST markers for a sugarcane (*Saccharum* spp.) commercial cross, **Molecular Breeding**, Berlin, v. 20, n. 3, p. 189-208, 2007.

OLIVEIRA, K.M.; PINTO, L.R.; MARCONI, T.G.; MOLINARI, M.; ULIAN, E.C.; CHABREGAS, S.M.; FALCO, M.C.; BURNQUIST, W.; GARCIA, A.A.F.; SOUZA A.P. Characterization of new polymorphic functional markers for sugarcane. **Genome**, Ottawa, v. 52, n. 2, p. 191-209, 2009.

ORAGUZIE, N.C.; WILCOX, P.L.; RIKKERINK, E.H.A.; De SILVA, H.N. Linkage Disequilibrium. In: ORAGUZIE, N.C.; RIKKERINK, E.H.A.; GARDINER, S.E.; De SILVA, H.N. (Ed.) **Association mapping in plants**. New York: Springer, 2007. chap. 2, p. 11-39.

PAN, Y.B.; BURNER, D.M.; LEGENDRE, B.L.; GRISHAM, M.P.; WHITE, W.H. An Assessment of the genetic diversity within a collection of *Saccharum spontaneum* L. with RAPD-PCR. **Genetic Resources and Crop Evolution**, Dordrecht, v. 51, n. 8, p. 895-903, 2004.

PERRIER, X.; JACQUEMOUD-COLLET, J.P. **DARwin software 5.0**. Montpellier, France, 2006. Disponível em: <http://darwin.cirad.fr/darwin>. Acesso em: 6 julho 2009.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. chap. 1, p. 3-56.

PINTO, L.R.; OLIVEIRA, K.M.; ULIAN, E.C.; GARCIA, A.A.F.; SOUZA, A.P. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. **Genome**, Ottawa, v. 47, n. 5, p. 795–804, 2004.

PINTO, L.R.; GARCIA, A.A.F.; PASTINA, M.M.; TEIXEIRA, L.H.M.; BRESSIANI, J.A.; ULIAN, E.C.; BIDOIA, M.A.P.; SOUZA, A.P. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. **Euphytica**, Dordrecht, v. 172, n. 3, p. 313-327, 2010.

PIPERIDIS, N.; JACKSON, P.A.; D´HONT, A.; BESSE, P.; HOARAU, J.Y.; COURTOIS, B.; AITKEN, K.S.; McINTYRE, C.L. Comparative genetics in sugarcane enables structured map enhancement and validation of marker-trait association. **Molecular Breeding**, Berlin, v. 21, n. 2, p. 233-247, 2008.

PRICE, A.; PATTERSON, N.; PLENGE, R.; WEINBLATT, M.; SHADICK, N.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, New York, v. 38, n. 8, p. 904-909, 2006.

PRITCHARD, J.K.; ROSENBERG, N.A. Use of unlinked genetic markers to detect population stratification in association studies. **American Journal of Human Genetics**, Chicago, v. 65, n. 1, p. 220-228, 1999.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Baltimore, v.155, n. 2, p. 945-959, 2000b.

PRITCHARD, J.K.; STEPHENS, M.; ROSENBERG, N.A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics**, Chicago, v. 67, n. 1, p. 170-181, 2000a.

R Development Core Team (2010). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2010. Disponível em: <http://www.r-project.org/>. Acesso em: 7 março 2010.

RABOIN, L.M.; PAUQUET, J.; BUTTERFIELD, M.; D'HONT, A.; GLASZMANN, J.C. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. **Theoretical and Applied Genetics**, New York, v. 116, n. 5, p. 701–714, 2008.

RABOIN, L.M.; OLIVEIRA, K.M.; LECUNFF, L.; TELISMART, H.; ROQUES, D.; BUTTERFIELD, M.; HOARAU, J.Y.; D'HONT, A. Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk colour and a new rust resistance gene. **Theoretical and Applied Genetics**, New York, v. 112, n. 7, p. 1382–1391, 2006.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, London, v. 5, n. 2, p. 94-100, 2002.

RAFALSKI, A.; MORGANTE, M. Corn and Humans: recombination and linkage disequilibrium in two genomes of similar size. **Trends in Genetics**, London, v. 20, n. 2, p. 103-111, 2004.

REFFAY, N.; JACKSON, P.A.; AITKEN, K.S.; HOARAU, J.Y.; D'HONT, A.; BESSE, P.; McINTYRE, C.L. Characterization of genome regions incorporated from an important wild relative into Australian sugarcane. **Molecular Breeding**, Berlin, v. 15, n. 4, p. 367-381, 2005.

REIMER, S.; POZNIAK, F.R.; CLARKE, F.R.; CLARKE, J.M.; SOMMERS, D.J.; KNOX, R.E.; SINGH, A.K. Association analysis of yellow pigment in an elite collection of durum wheat cultivars and breeding lines**. Genome**, Ottawa, v. 51, n. 12, p. 1016-1025, 2008.

REMINGTON, D.L.; THORNSBERRY, J.M.; MATSUOKA, Y.; WILSON, L.M.; WHITT, S.R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M.M.; BUCKLER, E.S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proceedings of the National Academy of Science**, Washington, v. 98, n. 20, p. 11479-11484, 2001.

ROACH, B.T. Origin and improvement of the genetic base of sugarcane. **Proceedings of the Australian Society of Sugar Cane Technologists**, Tweed Heads, v. 10, p. 34-47, 1989.

ROBBINS, M.D.; SIM, S-C.; YANG, W.; DEYNZE, A.V.; KNAAP, E.; JOOBEUR, T.; FRANCIS, D.M. Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. **Journal of experimental Botany**, Berkley, v. 62, n. 6, p. 1831-1845, 2011.

ROSSI, M.; ARAUJO, P.G.; PAULET, F.; GARSMEUR, O.; DIAS, V.M.; CHEN, H.; VAN SLUYS M.A.; D'HONT A. Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. **Molecular Genetics and Genomics**, New York, v. 269, n. 3, p. 406-419, 2003.

ROSTOKS, N.; RAMSAY, L.; MAcKENZIE, K.; CARDLE, L.; BHAT, P.R.; ROOSE, M.L.; SVENSSON, J.T.; STEIN, N.; VARSHNEY, R.K.; MARSHALL, D.F.; GRANER, A.; CLOSE, T.J.; WAUGH, R. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. **Proceedings of the National Academy of Sciences**, Washington, v. 103, n. 49, p. 18656-18661, 2006.

SCHNELL, R.J.; OLANO, C.T.; BROWN, J.S.; MEEROW, A.W.; CERVANTES-MARTINEZ, C. Retrospective determination of the parental population of the superior cacao (*Theobroma cacao* L.) seedlings and association with microsatellites alleles with productivity. **Journal of the American Society for Horticultural Science**, Alexandria, v. 130, n. 2, p. 181-190, 2005.

SELVI, A.; NAIR, N.V.; BALASUNDARAM, N.; MOHAPATRA, T. Evaluation of maize microsatellite markers for genetic diversity analysis and fingerprinting in sugarcane. **Genome**, Ottawa, v. 46, n. 3, p. 394-403, 2003.

SELVI, A.; NAIR, N.V.; NOYER, J.L.; SINGH, N.K.; BALASUNDARAM, N.; BANSAL, K.C.; KOUNDAL, K.R.; MOHAPATRA, T. Genomic constitution and genetic relationship

among the tropical and subtropical Indian sugarcane cultivars revealed by AFLP. **Crop Science,** Madison, v. 45, n. 5, p. 1750-1757, 2005.

SELVI, A.; NAIR, N.V.; NOYER, J.L.; SINGH, N.K.; BALASUNDARAM, N.; BANSAL, K.C.; KOUNDAL, K.R.; MOHAPATRA, T. AFLP analysis of the phenetic organization and genetic diversity in the sugarcane complex *Saccharum* and *Erianthus*. **Genetic Resources and Crop Evolution**, Dordrecht, v. 53, n. 4, p. 831-842, 2006.

SILLS, G.R.; BRIDGES, W.; AL-JANABI, S.M.; SOBRAL, B.W.S. Genetic analysis of agronomic traits in a cross between sugarcame (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes and Jesw. ex Grassl). **Molecular Breeding**, Berlin, v. 1, n. 4, p. 355-363, 1995.

SOBRAL, B.W.S.; BRAGA, D.P.V.; LAHOOD, E.S.; KEIM, P. Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. **Theoretical and Applied Genetics**, New York, v. 87, n. 7, p. 843-853, 1994.

SOUZA, H.M.B. **Mapeamento genético de híbridos comerciais de cana-de-açúcar (*Saccharum* spp.) utilizando marcadores moleculares do tipo microssatélites**. 2003. 85p. Dissertação (Mestrado em Ciências) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2003.

STAUB, J.E.; SERQUEN, F.; GUPTA, M. Genetic markers, map construction and their application in plant breeding. **HortScience**, Alexandria, v. 31, n. 5, p. 729-741, 1996.

STICH, B.; MÖHRING, J.; PIEPHO, H-P.; HECKENBERG, M.; BUCKLER, E.S.; MELCHINGER, A.E. Comparison of Mixed-Model Approaches for Association Mapping. **Genetics**, Baltimore, v. 178, n. 3, p. 1745-1754, 2008.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. **MEGA5**: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Arizona, 2011. Disponível em: <http://www.megasoftware.net/>. Acesso em: 22 fevereiro 2011.

SHEJI, M.; NAIR, N.V.; CHATURVEDI, P.K.; SELVI, A. Analysis of genetic diversity among *Saccharum spontaneum* L. from geographical regions in India, using molecular markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 53, n. 6, p. 1221-1231, 2006.

SZALMA, S.J.; BUCKLER, E.S.; SNOOK, M.E.; McMULLEN, M.D. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks **Theoretical and Applied Genetics**, New York, v. 110, n. 7, p. 1324-1333, 2005.

THORNSBERRY, J.M.; GOODMAN, M.M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E.S. *Dwarf8* polymorphisms associate with variation in flowering time. **Nature Genetics**, New York, v. 28, n. 3, p. 286-289, 2001.

TOMMASINI, L.; SCHNURBUSCH, T.; FOSSATI, D.; MASCHER, F.; KELLER, B. Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. **Theoretical and Applied Genetics**, New York, v. 115, n. 5, p. 697-708, 2007.

TZVELEV, N.N. The system of grasses (Poaceae) and their evolution. **Botanical Review**, New York, v. 55, n. 3, p. 141-204, 1989.

VAN OOIJEN, J.W.; VOORRIPS, R.E. **JoinMap 3.0**: software for the calculation of genetic linkage maps. Plant Research International, Wageningen, 2001.

VEKEMANS, X. **AFLP-SURV v. 1.0**: A program for genetic diversity analysis with AFLP (and RAPD) population data. Belgium, 2002. Disponível em: <http://www.ulb.ac.be/sciences/lagev/aflp-surv.html>. Acesso em: 01 fevereiro 2011.

VUYLSTEKE M.; MANK R.; BRUGMANS B.; STAM P.; KUIPER M. Further characterization of AFLP data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. **Molecular Breeding**, Dordrecht, v. 6, n. 3, p. 265–276, 2000.

ZHANG, Z.; BUCKLER, E.S.; CASSTEVENS, T.M.; BRADBURY, P.J. Software engineering the mixed model for genome-wide association studies on large samples. **Briefings in Bioinformatics**, Oxford, v. 10, n. 6, p. 664-675, 2009.

ZHAO, K.Y.; ARANZANA, M.J.; KIM, S.; LISTER, C.; SHINDO, C.; TANG, C.L.; TOOMAJIAN, C.; ZHENG, H.G.; DEAN, C.; MARJORAM, P.; NORDBORG, M. An Arabidopsis example of association mapping in structured samples. **PLOS Genetics**, San Francisco, v. 3, n. 1, p. 71-82, 2007.

YAO, J.; WANG, L.X.; LIU, L.H.; ZHAO, C.P.; ZHENG, Y.L. Association mapping of agronomic traits on chromosome 2A of wheat. **Genetica**, Dordretcht, v. 137, n. 1, p. 67–75, 2009.

WANG, J.; McCLEAN P.E.; LEE, R.; GOSS, R.J.; HELMS, T. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. **Theoretical and Applied Genetics**, New York, v. 116, n. 6, p. 777-787, 2008.

WEI, X., JACKSON, P.A.; McINTYRE, C.L.; AITKEN, K.S.; CROFT, B. Associations between DNA markers and resistance to disease in sugarcane and effects of population substructure. **Theoretical and Applied Genetics**, New York, v. 114, n. 1, p. 155-164, 2006.

WEI, X.; JACKSON P.A.; HERMANN, S.; KILIAN, A.; HELLER-USZYNSKA, K.; DEOMANO, E. Simultaneously accounting for population structure, genotype by environment

interaction, and spatial variation in marker–trait associations in sugarcane, **Genome**, Ottawa, v. 53, n. 11, p. 973-981, 2010.

WESSEL, J.; SCHORK, N.J. Generalized genomic distance-based regression methodology for multilocus association analysis. **American Journal of Human Genetics**, Chicago, v. 79, n. 5, p. 792-806, 2006.

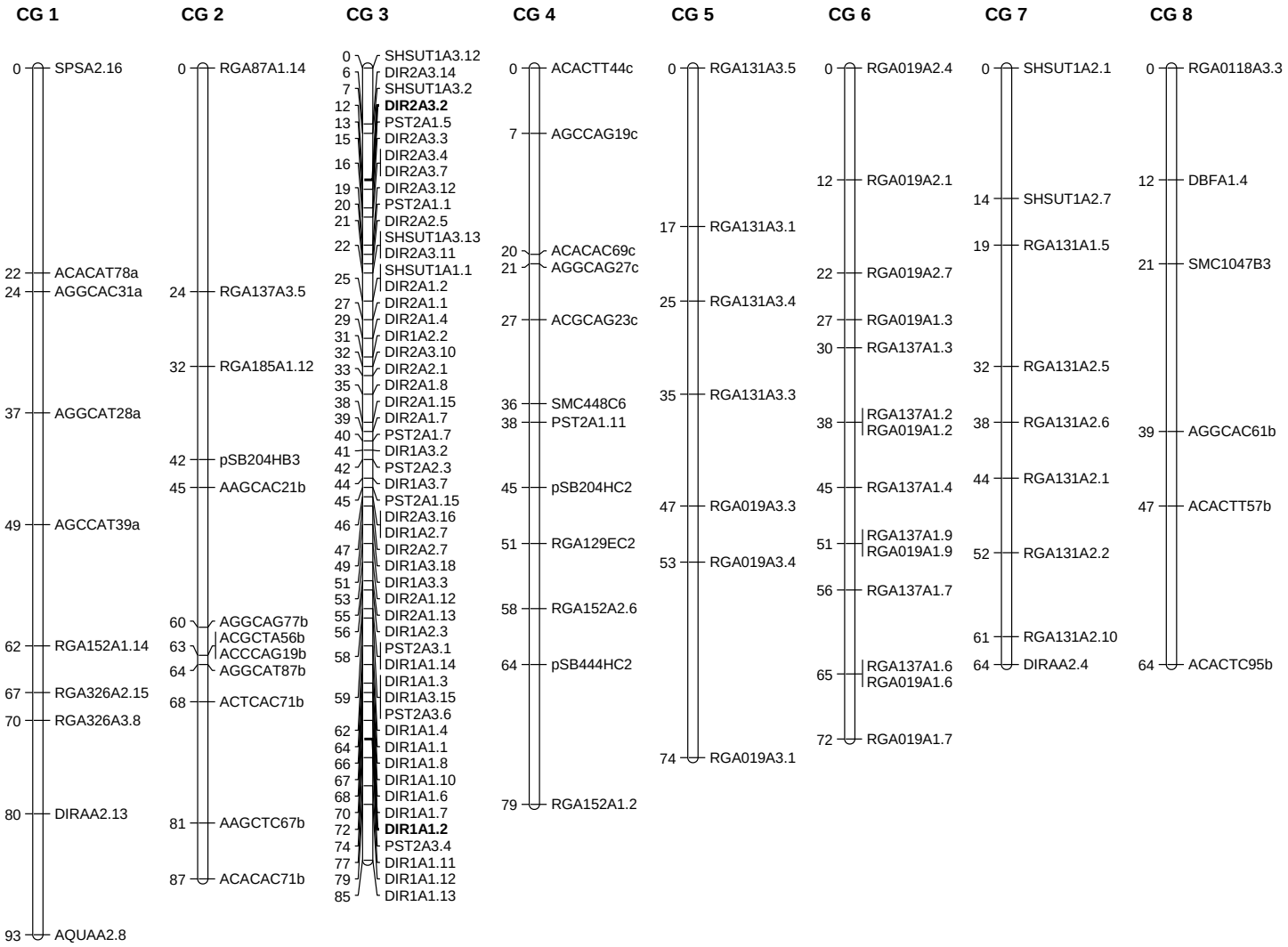
WHITT, S.R.; WILSON, L.M.; TENAILLON, M.I.; GAUT, B.S.; BUCKLER, E.S. Genetic diversity and selection in the maize starch pathway. **Proceedings of the National Academy of Science**, Washington, v. 99, n. 20, p. 12959-12962, 2002.

WILSON, L.M.; WHITT, S.R.; IBANEZ, A.M.; ROCHEFORD, T.R.; GOODMAN, M.M.; BUCKLER, E.S. Dissection of maize kernel composition and starch production by candidate gene association. **Plant Cell**, Rockville, v. 16, n. 10, p. 2719-2733, 2004.

WU, K.K.; BURNQUIST, W.; SORRELS, M.E.; TEW, T.L.; MOORE, P.H.; TANKSLEY, S.D. The detection and estimation of linkage in polyploids using single does restriction fragments. **Theoretical and Applied Genetics**, New York, v. 83, n. 3, p. 294-300, 1992.

YU, J.M.; PRESSOIR, G.; BRIGGS, W.H.; VROH, Bi, IV; YAMASAKI, M.; DOEBLEY, J.F.; McMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, n. 2, p. 203-208, 2006.

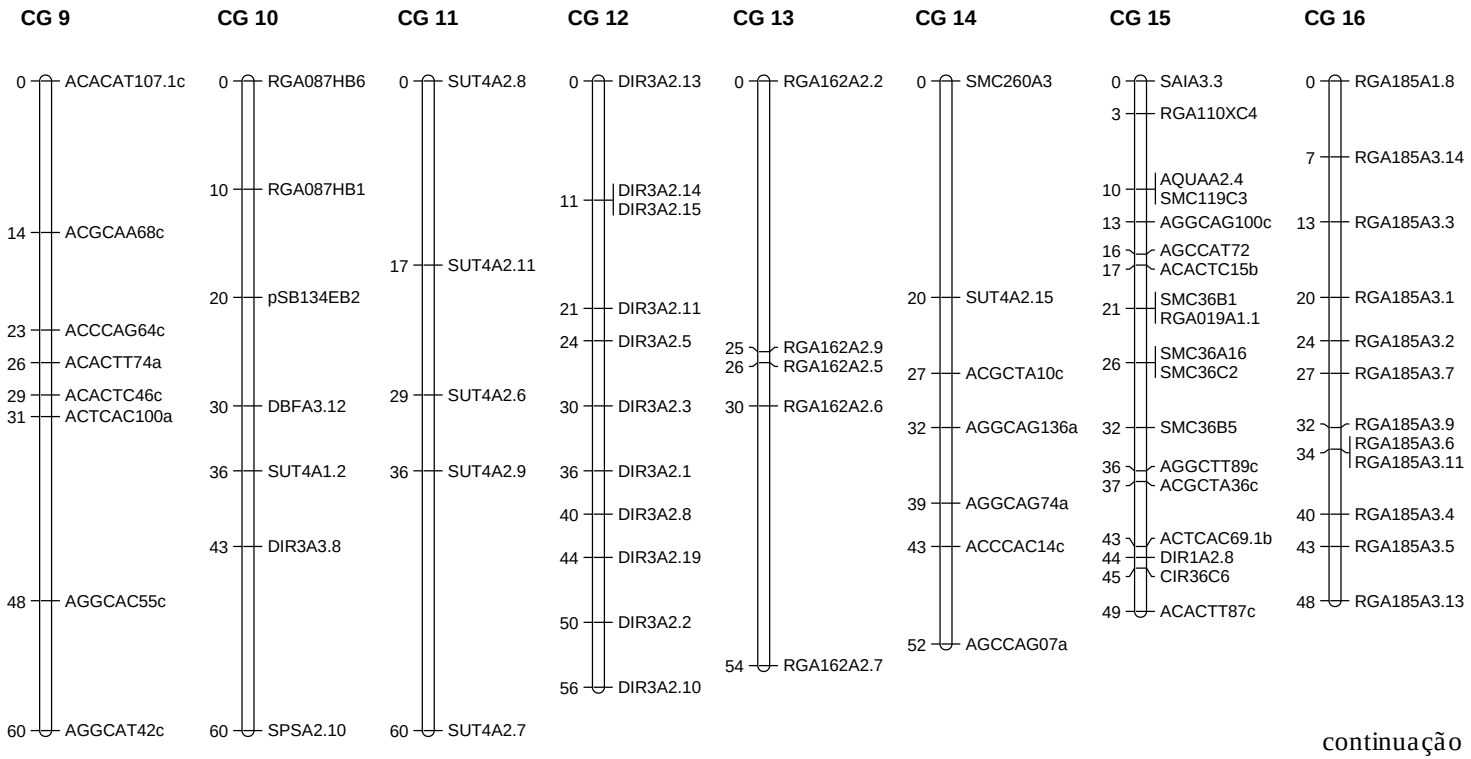
ANEXOS



continua

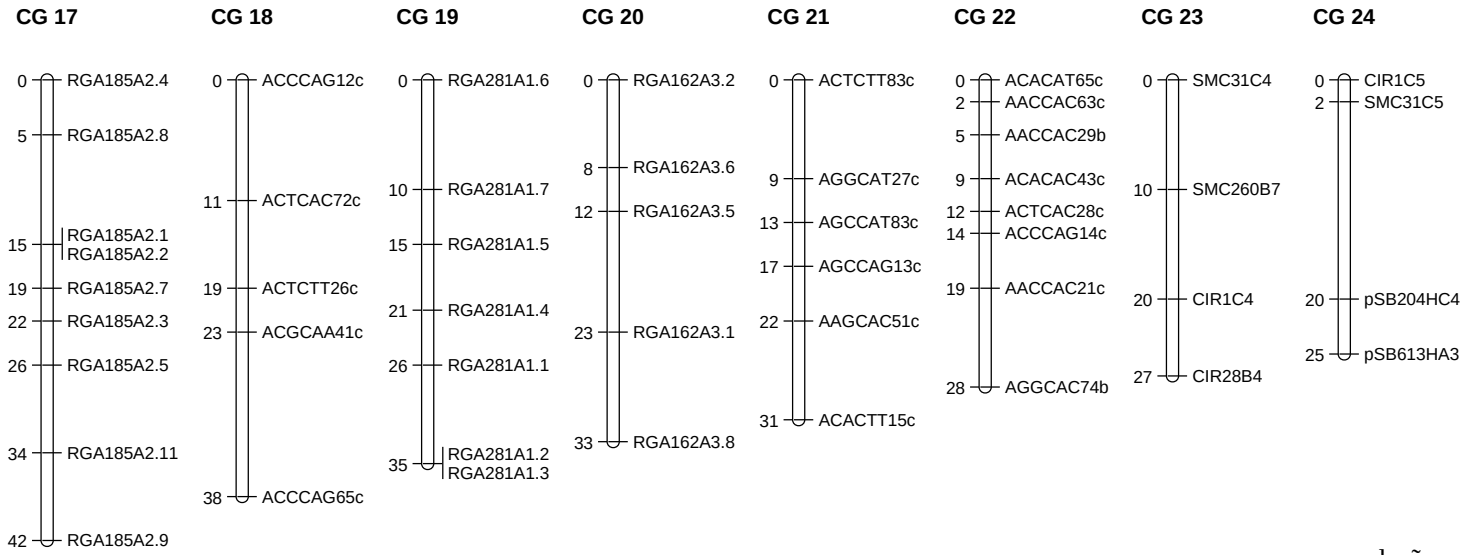
Anexo A - Grupos de co-segregação do mapa genético entre as cultivares SP80-180 x SP80-4966, cujas marcas forma utilizadas para o estudo dos níveis de desequilíbrio de ligação no genoma de cana-de-açúcar. No total são apresentados 24 grupos de co-sergregação e 243 marcas. As marcas em negrito

no CG 3 (DIR2A3.2 e DIR1A1.2), separadas por uma distancia de 60 cM, apresentaram associação significativa ($p < 1{,}19 \times 10^{-6}$). Os grupos de co-segregação foram construídos por meio do programa JoinMap 3.0 (VAN OOIJEN; VOORRIPS, 2001).



continuação

Anexo A - Grupos de co-segregação do mapa genético entre as cultivares SP80-180 x SP80-4966, cujas marcas forma utilizadas para o estudo dos níveis de desequilíbrio de ligação no genoma de cana-de-açúcar. No total são apresentados 24 grupos de co-sergregação e 243 marcas. As marcas em negrito no CG 3 (DIR2A3.2 e DIR1A1.2), separadas por uma distancia de 60 cM, apresentaram associação significativa ($p < 1{,}19 \times 10^{-6}$). Os grupos de co-segregação foram construídos por meio do programa JoinMap 3.0 (VAN OOIJEN; VOORRIPS, 2001).



conclusão

Anexo A - Grupos de co-segregação do mapa genético entre as cultivares SP80-180 x SP80-4966, cujas marcas forma utilizadas para o estudo dos níveis de desequilíbrio de ligação no genoma de cana-de-açúcar. No total são apresentados 24 grupos de co-sergregação e 243 marcas. As marcas em negrito no CG 3 (DIR2A3.2 e DIR1A1.2), separadas por uma distancia de 60 cM, apresentaram associação significativa ($p < 1{,}19 \times 10^{-6}$). Os grupos de co-segregação foram construídos por meio do programa JoinMap 3.0 (VAN OOIJEN; VOORRIPS, 2001).

Anexo B - Estimativas dos BLUPs para a característica Altura de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib. Preto
CTC1	243,09	249,98	262,75	253,52	263,78
CTC2	265,63	247,47	267,12	252,62	265,91
CTC3	239,87	247,24	nd	253,79	266,15
CTC7	249,07	253,64	nd	252,84	261,96
CTC8	266,21	253,91	nd	nd	264,60
CTC11	241,45	250,09	nd	253,03	nd
CTC12	238,27	251,31	nd	nd	263,38
IAC48-65	240,53	250,72	nd	nd	nd
IAC64-257	252,71	nd	nd	nd	264,34
IAC82-2045	248,38	252,30	nd	252,71	263,33
IAC82-3092	278,89	251,39	nd	251,60	269,52
IAC83-4157	251,44	nd	nd	253,47	262,97
IAC86-2210	252,51	251,97	262,66	251,02	265,67
IAC86-2480	229,26	249,26	nd	nd	264,20
IAC87-3396	272,31	251,95	263,36	248,69	270,03
IAC91-1099	242,85	248,48	262,51	254,49	264,52
IAC91-2195	257,31	250,12	263,63	252,34	265,89
IAC91-2205	255,64	250,39	262,45	253,64	265,28
IAC91-2218	235,14	248,63	261,95	252,79	265,58
IAC91-3093	232,15	250,00	nd	nd	263,85
IAC91-4168	265,26	254,13	262,12	253,16	263,66
IAC91-5155	279,12	253,84	261,08	nd	267,77
IACSP93-1042	250,50	253,80	264,50	252,87	259,88
IACSP93-2060	259,60	251,19	nd	nd	266,42
IACSP93-3046	245,92	254,15	257,74	253,78	264,75
IACSP93-3050	246,51	249,70	263,42	nd	265,11
IACSP93-6006	256,40	252,56	261,96	250,70	266,64
IACSP94-2094	240,01	254,42	261,62	nd	261,30
IACSP94-2101	237,25	249,51	263,02	250,42	266,29
IACSP94-2180	239,42	251,80	nd	nd	263,04
IACSP94-4002	246,26	252,46	262,30	nd	263,43
IACSP94-4004	247,80	251,37	263,13	251,89	264,29
IACSP95-2048	236,26	254,79	261,86	251,89	260,57
IACSP95-2078	235,74	252,22	261,85	252,84	262,12
IACSP95-2213	256,50	253,86	262,12	nd	263,62

Anexo B - Estimativas dos BLUPs para a característica Altura de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib. Preto
IACSP95-2289	237,40	255,34	259,71	250,98	263,22
IACSP95-3018	256,28	252,46	261,81	nd	265,29
IACSP95-3028	258,20	256,82	262,32	nd	260,68
IACSP95-5000	256,78	252,84	260,14	255,02	263,90
IACSP96-2019	242,41	255,21	263,19	nd	259,27
IACSP96-2022	235,72	nd	nd	nd	262,01
IACSP96-2037	246,16	251,38	265,42	253,10	260,54
IACSP96-2041	249,50	258,16	262,76	251,27	258,73
IACSP96-2042	247,88	256,99	262,36	252,27	259,06
IACSP96-2086	236,65	nd	nd	nd	262,14
IACSP96-2100	239,66	251,59	262,58	252,69	262,70
IACSP96-3057	243,57	252,01	262,96	251,89	263,24
IACSP96-7586	264,05	252,89	262,09	nd	265,65
IACSP97-2000	249,52	249,53	261,29	nd	267,82
IACSP97-2025	256,28	nd	nd	nd	264,83
IACSP97-2053	241,01	250,91	261,59	nd	264,98
IACSP97-2055	249,64	246,62	263,60	nd	268,44
IACSP97-2109	252,41	249,61	262,29	nd	267,13
IACSP97-3354	243,32	nd	261,18	nd	nd
IACSP97-3402	261,10	254,51	260,37	nd	265,34
IACSP97-3406	246,15	247,99	264,25	nd	265,94
IACSP97-6682	244,14	251,97	262,59	253,04	262,57
IACSP97-7065	248,58	251,34	263,93	nd	263,24
IACSP97-7077	249,35	250,98	262,65	nd	264,98
IACSP98-2071	263,04	nd	nd	nd	265,75
IACSP99-2033	265,06	nd	nd	nd	266,02
IACSP99-2058	293,68	nd	nd	nd	269,94
NA5679	240,05	250,65	nd	nd	nd
POJ2878	234,81	249,93	nd	nd	nd
RB72-454	260,86	256,00	262,88	251,91	261,68
RB83-5054	240,59	248,94	nd	nd	266,07
RB83-5486	257,83	252,55	265,77	250,62	263,12
RB84-5210	249,69	251,09	263,29	nd	nd
RB84-5257	262,82	256,17	nd	nd	261,88
RB85-5035	247,48	251,67	nd	nd	nd

Anexo B - Estimativas dos BLUPs para a característica Altura de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	264,62	258,26	258,77	252,13	263,81
RB85-5156	241,22	250,16	260,60	255,49	263,53
RB85-5453	230,03	260,85	257,62	251,95	257,83
RB85-5536	274,86	256,90	261,26	nd	263,95
RB86-7515	281,93	261,81	259,87	248,64	265,04
RB92-5211	242,21	250,17	262,82	252,52	264,40
RB92-5268	261,83	247,08	261,77	254,59	269,17
RB92-5345	256,73	250,89	262,97	251,88	266,16
RB92-579	262,84	248,63	270,33	246,25	267,53
RB92-8064	283,20	254,76	nd	250,50	267,85
RB93-5744	285,39	254,03	258,18	257,09	266,52
SP70-1143	246,40	249,98	nd	nd	265,83
SP71-1406	240,49	251,58	nd	nd	263,41
SP79-1011	241,74	257,00	nd	251,84	258,60
SP79-2233	252,14	nd	nd	nd	264,26
SP80-1816	264,96	256,07	264,02	251,89	261,05
SP80-1842	280,61	253,11	264,11	250,54	267,40
SP80-3280	285,27	260,29	263,22	250,14	262,16
SP81-3250	261,76	253,26	260,98	252,63	265,72
SP83-2847	287,02	255,68	263,64	249,79	266,94
SP84-1201	234,49	249,89	nd	nd	nd
SP84-2025	262,93	nd	263,86	nd	nd
SP85-3877	250,79	nd	nd	252,06	nd
SP85-5077	268,86	254,59	nd	nd	nd
SP86-155	252,90	255,82	262,82	251,14	261,59
SP86-42	254,97	249,78	262,30	nd	267,31
SP87-344	232,18	249,24	262,74	nd	nd
SP87-365	248,27	249,42	nd	nd	266,63
SP89-1115	243,54	250,62	261,86	253,66	263,96
SP90-1638	241,02	250,78	nd	nd	nd
SP90-3414	250,13	250,58	260,59	255,54	nd
SP91-1049	249,54	254,20	263,03	253,35	260,35
SP91-3011	248,03	251,74	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região

Anexo C - Estimativas dos BLUPs para a característica Diâmetro de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
CTC1	2,44	2,28	2,74	2,77	2,77
CTC2	2,22	2,25	2,74	2,75	2,80
CTC3	2,34	2,28	nd	2,76	2,80
CTC7	2,23	2,35	nd	2,73	2,74
CTC8	2,27	2,27	nd	nd	2,81
CTC11	2,14	2,29	nd	2,73	nd
CTC12	2,27	2,28	nd	nd	2,80
IAC48-65	2,20	2,29	nd	nd	nd
IAC64-257	2,51	nd	nd	nd	2,82
IAC82-2045	2,27	2,31	nd	2,74	2,77
IAC82-3092	2,28	2,28	nd	2,73	2,82
IAC83-4157	2,33	nd	nd	2,77	2,76
IAC86-2210	2,32	2,26	2,71	2,79	2,80
IAC86-2480	2,29	2,30	nd	nd	2,78
IAC87-3396	2,37	2,34	2,73	2,72	2,77
IAC91-1099	2,38	2,20	2,76	2,79	2,81
IAC91-2195	2,14	2,28	2,71	2,76	2,76
IAC91-2205	2,29	2,30	2,73	2,73	2,79
IAC91-2218	2,08	2,33	2,73	2,73	2,73
IAC91-3093	2,29	2,28	nd	nd	2,80
IAC91-4168	2,11	2,35	2,71	2,72	2,74
IAC91-5155	2,19	2,30	2,72	nd	2,76
IACSP93-1042	2,36	2,30	2,74	2,74	2,78
IACSP93-2060	2,28	2,31	nd	nd	2,77
IACSP93-3046	2,30	2,35	2,73	2,69	2,78
IACSP93-3050	2,22	2,29	2,71	nd	2,79
IACSP93-6006	2,37	2,36	2,72	2,69	2,79
IACSP94-2094	1,97	2,31	2,66	nd	2,79
IACSP94-2101	2,27	2,25	2,72	2,75	2,82
IACSP94-2180	2,26	2,31	nd	nd	2,78
IACSP94-4002	2,36	2,35	2,73	nd	2,74
IACSP94-4004	2,43	2,30	2,73	2,74	2,80
IACSP95-2048	2,29	2,33	2,70	2,74	2,77
IACSP95-2078	2,14	2,28	2,72	2,74	2,78
IACSP95-2213	2,17	2,31	2,70	nd	2,78

Anexo C - Estimativas dos BLUPs para a característica Diâmetro de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
IACSP95-2289	2,48	2,33	2,72	2,74	2,78
IACSP95-3018	2,39	2,31	2,72	nd	2,79
IACSP95-3028	2,45	2,34	2,70	nd	2,79
IACSP95-5000	2,42	2,23	2,75	2,76	2,83
IACSP96-2019	2,27	2,34	2,72	nd	2,74
IACSP96-2022	2,23	nd	nd	nd	2,78
IACSP96-2037	2,40	2,30	2,74	2,75	2,78
IACSP96-2041	2,25	2,34	2,70	2,73	2,77
IACSP96-2042	2,39	2,33	2,72	2,76	2,75
IACSP96-2086	2,30	nd	nd	nd	2,79
IACSP96-2100	2,33	2,31	2,72	2,74	2,79
IACSP96-3057	2,19	2,30	2,72	2,73	2,79
IACSP96-7586	2,33	2,32	2,68	nd	2,81
IACSP97-2000	2,08	2,31	2,72	nd	2,75
IACSP97-2025	2,45	nd	nd	nd	2,81
IACSP97-2053	2,67	2,35	2,73	nd	2,78
IACSP97-2055	2,41	2,27	2,74	nd	2,81
IACSP97-2109	2,45	2,26	2,72	nd	2,84
IACSP97-3354	2,18	nd	2,70	nd	nd
IACSP97-3402	2,21	2,30	2,71	nd	2,78
IACSP97-3406	2,80	2,25	2,76	nd	2,87
IACSP97-6682	2,44	2,25	2,72	2,76	2,83
IACSP97-7065	2,51	2,27	2,70	nd	2,86
IACSP97-7077	2,18	2,30	2,72	nd	2,76
IACSP98-2071	2,35	nd	nd	nd	2,79
IACSP99-2033	2,17	nd	nd	nd	2,77
IACSP99-2058	2,28	nd	nd	nd	2,78
NA5679	2,30	2,30	nd	nd	nd
POJ2878	2,35	2,31	nd	nd	nd
RB72-454	2,38	2,31	2,73	2,73	2,80
RB83-5054	2,20	2,27	nd	nd	2,80
RB83-5486	2,38	2,23	2,75	2,76	2,82
RB84-5210	2,25	2,30	2,71	nd	nd
RB84-5257	2,16	2,32	nd	nd	2,75
RB85-5035	2,16	2,28	nd	nd	nd

Anexo C - Estimativas dos BLUPs para a característica Diâmetro de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	2,31	2,31	2,73	2,72	2,79
RB85-5156	2,19	2,27	2,71	2,75	2,81
RB85-5453	2,44	2,30	2,71	2,78	2,78
RB85-5536	2,25	2,34	2,72	nd	2,74
RB86-7515	2,45	2,34	2,70	2,78	2,76
RB92-5211	2,19	2,27	2,71	2,76	2,80
RB92-5268	2,31	2,30	2,70	2,73	2,82
RB92-5345	2,16	2,32	2,70	2,73	2,78
RB92-579	2,27	2,23	2,73	2,75	2,85
RB92-8064	2,39	2,32	nd	2,72	2,80
RB93-5744	2,38	2,33	2,69	2,73	2,81
SP70-1143	2,38	2,26	nd	nd	2,83
SP71-1406	2,23	2,29	nd	nd	2,79
SP79-1011	2,27	2,30	nd	2,75	2,78
SP79-2233	2,31	nd	nd	nd	2,79
SP80-1816	2,23	2,32	2,72	2,74	2,76
SP80-1842	2,20	2,25	2,72	2,72	2,84
SP80-3280	2,45	2,36	2,71	2,74	2,75
SP81-3250	2,18	2,33	2,76	2,69	2,76
SP83-2847	2,20	2,26	2,66	2,78	2,83
SP84-1201	2,26	2,29	nd	nd	nd
SP84-2025	2,32	nd	2,72	nd	nd
SP85-3877	2,40	nd	nd	2,76	nd
SP85-5077	2,51	2,33	nd	nd	nd
SP86-155	2,11	2,28	2,70	2,76	2,78
SP86-42	2,12	2,32	2,72	nd	2,75
SP87-344	2,29	2,28	2,73	nd	nd
SP87-365	2,34	2,31	nd	nd	2,79
SP89-1115	2,41	2,32	2,74	2,73	2,77
SP90-1638	2,28	2,30	nd	nd	nd
SP90-3414	2,36	2,28	2,73	2,76	nd
SP91-1049	2,31	2,33	2,73	2,75	2,75
SP91-3011	2,18	2,28	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região

Anexo D - Estimativas dos BLUPs para a característica Número de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
CTC1	11,04	12,95	12,18	12,74	11,84
CTC2	13,39	12,90	12,27	13,00	11,84
CTC3	13,87	12,32	nd	12,90	12,66
CTC7	12,45	13,21	nd	12,33	12,16
CTC8	12,98	12,94	nd	nd	12,06
CTC11	17,83	13,77	nd	12,40	nd
CTC12	13,61	13,08	nd	nd	12,00
IAC48-65	14,90	13,03	nd	nd	nd
IAC64-257	13,52	nd	nd	nd	12,32
IAC82-2045	13,21	12,60	nd	12,55	12,65
IAC82-3092	11,99	12,90	nd	12,47	12,28
IAC83-4157	12,32	nd	nd	12,66	12,27
IAC86-2210	11,97	12,46	12,04	13,28	12,04
IAC86-2480	13,24	12,65	nd	nd	12,39
IAC87-3396	11,72	12,34	12,50	12,88	12,07
IAC91-1099	13,49	12,96	12,16	12,62	12,27
IAC91-2195	13,34	12,84	12,18	12,62	12,37
IAC91-2205	12,72	12,84	12,21	12,86	12,02
IAC91-2218	14,69	13,18	12,32	12,17	12,50
IAC91-3093	11,79	12,74	nd	nd	12,11
IAC91-4168	14,15	12,60	12,50	12,48	12,52
IAC91-5155	14,01	12,87	12,06	nd	12,38
IACSP93-1042	11,97	12,91	12,11	12,70	12,10
IACSP93-2060	12,54	12,51	nd	nd	12,43
IACSP93-3046	13,45	12,99	11,77	13,35	11,90
IACSP93-3050	13,29	13,02	11,92	nd	12,28
IACSP93-6006	11,73	12,12	12,15	12,77	12,75
IACSP94-2094	14,94	13,12	12,43	nd	11,89
IACSP94-2101	13,07	13,19	12,03	12,64	12,10
IACSP94-2180	12,16	12,42	nd	nd	12,48
IACSP94-4002	11,58	12,24	12,20	nd	12,56
IACSP94-4004	12,36	13,11	12,11	12,65	12,02
IACSP95-2048	11,24	12,37	12,14	12,86	12,36
IACSP95-2078	14,36	13,11	12,38	12,65	12,00
IACSP95-2213	13,32	12,94	12,31	nd	11,97

Anexo D - Estimativas dos BLUPs para a característica Número de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
IACSP95-2289	10,39	12,60	12,08	12,96	11,99
IACSP95-3018	12,08	12,39	12,32	nd	12,36
IACSP95-3028	11,98	12,74	12,23	nd	12,09
IACSP95-5000	12,88	12,89	12,16	12,65	12,24
IACSP96-2019	12,58	12,67	12,43	nd	12,03
IACSP96-2022	11,97	nd	nd	nd	12,12
IACSP96-2037	11,44	12,74	12,04	13,02	11,96
IACSP96-2041	12,27	13,03	12,46	12,68	11,69
IACSP96-2042	12,50	13,19	12,28	12,67	11,76
IACSP96-2086	10,52	nd	nd	nd	11,93
IACSP96-2100	11,91	12,90	12,44	12,96	11,51
IACSP96-3057	11,55	12,43	12,25	12,88	12,21
IACSP96-7586	11,79	12,60	12,18	nd	12,25
IACSP97-2000	12,86	12,62	12,15	nd	12,40
IACSP97-2025	13,67	nd	nd	nd	12,34
IACSP97-2053	11,39	12,36	12,08	nd	12,54
IACSP97-2055	12,67	12,26	12,11	nd	12,78
IACSP97-2109	11,41	12,37	12,14	nd	12,48
IACSP97-3354	14,92	nd	12,46	nd	nd
IACSP97-3402	13,26	12,60	12,53	nd	12,08
IACSP97-3406	10,61	12,43	11,93	nd	12,52
IACSP97-6682	12,66	13,20	12,05	12,59	12,07
IACSP97-7065	11,81	12,54	12,16	nd	12,34
IACSP97-7077	12,46	12,95	11,94	nd	12,22
IACSP98-2071	12,80	nd	nd	nd	12,22
IACSP99-2033	12,72	nd	nd	nd	12,21
IACSP99-2058	13,88	nd	nd	nd	12,36
NA5679	13,72	12,88	nd	nd	nd
POJ2878	12,07	12,66	nd	nd	nd
RB72-454	12,02	12,11	12,00	13,32	12,40
RB83-5054	10,63	12,25	nd	nd	12,44
RB83-5486	11,42	12,91	12,19	12,95	11,71
RB84-5210	12,79	12,48	12,46	nd	nd
RB84-5257	13,67	12,92	nd	nd	12,17
RB85-5035	11,20	12,55	nd	nd	nd

Anexo D - Estimativas dos BLUPs para a característica Número de Colmos (cm) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	12,18	12,60	12,04	12,86	12,35
RB85-5156	13,37	13,13	12,44	12,58	11,85
RB85-5453	11,73	12,40	12,31	12,93	12,15
RB85-5536	12,36	12,77	11,94	nd	12,40
RB86-7515	11,29	12,57	11,95	12,79	12,43
RB92-5211	13,24	12,81	12,18	12,85	12,15
RB92-5268	12,97	12,86	12,28	12,62	12,19
RB92-5345	12,12	12,96	12,09	12,54	12,25
RB92-579	11,96	12,34	12,52	13,03	11,94
RB92-8064	12,74	12,95	nd	12,66	12,13
RB93-5744	12,10	12,91	12,03	12,94	11,96
SP70-1143	14,17	13,06	nd	nd	12,10
SP71-1406	11,06	12,47	nd	nd	12,29
SP79-1011	12,94	13,07	nd	12,51	12,19
SP79-2233	14,58	nd	nd	nd	12,45
SP80-1816	12,18	12,55	12,06	12,61	12,64
SP80-1842	11,96	13,13	12,19	12,73	11,76
SP80-3280	11,68	12,27	12,23	12,66	12,63
SP81-3250	12,81	12,51	12,41	12,82	12,19
SP83-2847	12,39	12,62	12,29	13,04	11,93
SP84-1201	15,30	13,08	nd	nd	nd
SP84-2025	12,86	nd	12,20	nd	nd
SP85-3877	12,97	nd	nd	12,80	nd
SP85-5077	13,70	12,87	nd	nd	nd
SP86-155	14,40	12,65	12,37	12,68	12,43
SP86-42	13,69	13,08	11,55	nd	12,65
SP87-344	12,81	12,91	12,03	nd	nd
SP87-365	13,82	12,43	nd	nd	12,68
SP89-1115	11,02	12,21	12,10	13,22	12,18
SP90-1638	17,03	13,30	nd	nd	nd
SP90-3414	15,17	13,24	12,24	12,54	nd
SP91-1049	11,75	12,55	11,78	13,13	12,32
SP91-3011	12,42	12,71	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região

Anexo E - Estimativas dos BLUPs para a característica Fibra % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
CTC1	13,05	13,15	11,61	11,02	11,75
CTC2	13,21	13,49	11,45	10,90	11,74
CTC3	12,71	12,79	nd	11,21	11,81
CTC7	12,52	12,86	nd	11,15	11,74
CTC8	13,45	13,08	nd	nd	11,78
CTC11	12,70	12,97	nd	11,05	nd
CTC12	13,54	13,06	nd	nd	11,82
IAC48-65	13,06	12,97	nd	nd	nd
IAC64-257	12,54	nd	nd	nd	11,69
IAC82-2045	12,57	12,82	nd	11,10	11,85
IAC82-3092	12,31	13,13	nd	11,07	11,50
IAC83-4157	12,32	nd	nd	11,10	11,67
IAC86-2210	12,74	12,76	11,73	11,09	11,88
IAC86-2480	12,60	12,94	nd	nd	11,70
IAC87-3396	13,54	12,79	11,64	11,19	12,04
IAC91-1099	13,29	13,44	11,51	11,04	11,61
IAC91-2195	12,46	12,99	11,66	11,18	11,55
IAC91-2205	12,55	13,17	11,62	11,11	11,51
IAC91-2218	13,07	13,19	11,68	11,07	11,61
IAC91-3093	13,27	12,87	nd	nd	11,94
IAC91-4168	13,36	12,59	11,74	11,36	11,93
IAC91-5155	12,63	12,81	11,72	nd	11,76
IACSP93-1042	12,81	12,86	11,64	11,12	11,85
IACSP93-2060	12,54	12,83	nd	nd	11,79
IACSP93-3046	12,93	12,89	11,68	11,12	11,82
IACSP93-3050	12,78	12,74	11,68	nd	11,91
IACSP93-6006	12,47	12,98	11,67	11,14	11,59
IACSP94-2094	13,19	12,73	11,91	nd	11,79
IACSP94-2101	13,12	13,03	11,60	11,14	11,79
IACSP94-2180	12,63	12,88	nd	nd	11,76
IACSP94-4002	12,44	12,90	11,70	nd	11,64
IACSP94-4004	12,41	13,06	11,65	11,11	11,55
IACSP95-2048	13,24	13,01	11,57	11,09	11,91
IACSP95-2078	12,69	12,82	11,75	11,15	11,73
IACSP95-2213	13,08	12,85	11,72	nd	11,83

Anexo E - Estimativas dos BLUPs para a característica Fibra % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
IACSP95-2289	13,74	12,91	11,49	11,26	12,06
IACSP95-3018	12,65	12,93	11,64	nd	11,72
IACSP95-3028	12,62	12,89	11,70	nd	11,70
IACSP95-5000	12,91	13,08	11,72	11,10	11,60
IACSP96-2019	12,80	12,89	11,60	nd	11,84
IACSP96-2022	13,14	nd	nd	nd	11,84
IACSP96-2037	12,55	13,02	11,72	11,06	11,60
IACSP96-2041	13,13	12,79	11,62	11,16	11,99
IACSP96-2042	12,23	12,87	11,52	11,09	11,85
IACSP96-2086	13,19	nd	nd	nd	11,86
IACSP96-2100	13,70	12,73	11,63	11,14	12,21
IACSP96-3057	13,67	12,92	11,69	11,24	11,85
IACSP96-7586	12,93	12,93	11,73	nd	11,70
IACSP97-2000	13,55	12,93	11,66	nd	11,94
IACSP97-2025	13,17	nd	nd	nd	11,85
IACSP97-2053	12,25	12,91	11,60	nd	11,67
IACSP97-2055	12,52	12,90	11,62	nd	11,73
IACSP97-2109	13,21	12,99	11,55	nd	11,90
IACSP97-3354	13,00	nd	11,66	nd	nd
IACSP97-3402	12,66	12,90	11,61	nd	11,77
IACSP97-3406	12,52	12,87	11,66	nd	11,73
IACSP97-6682	12,71	13,24	11,59	11,09	11,53
IACSP97-7065	12,42	12,93	11,58	nd	11,72
IACSP97-7077	12,83	12,86	11,72	nd	11,75
IACSP98-2071	13,28	nd	nd	nd	11,88
IACSP99-2033	13,27	nd	nd	nd	11,88
IACSP99-2058	13,26	nd	nd	nd	11,87
NA5679	12,77	12,89	nd	nd	nd
POJ2878	13,03	12,96	nd	nd	nd
RB72-454	12,52	12,92	11,77	11,01	11,70
RB83-5054	12,90	13,03	nd	nd	11,69
RB83-5486	13,00	12,84	11,68	11,25	11,77
RB84-5210	12,92	13,04	11,53	nd	nd
RB84-5257	13,48	13,13	nd	nd	11,73
RB85-5035	13,07	12,97	nd	nd	nd

Anexo E - Estimativas dos BLUPs para a característica Fibra % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	13,55	13,18	11,48	11,25	11,76
RB85-5156	12,46	12,98	11,54	11,28	11,59
RB85-5453	12,75	13,06	11,46	11,12	11,83
RB85-5536	13,69	13,08	11,61	nd	11,88
RB86-7515	13,24	12,98	11,72	11,07	11,83
RB92-5211	12,36	12,81	11,63	11,22	11,71
RB92-5268	13,06	12,67	11,76	11,14	11,97
RB92-5345	13,48	12,88	11,59	11,26	11,92
RB92-579	12,94	13,12	11,51	11,13	11,75
RB92-8064	12,84	12,80	nd	11,22	11,81
RB93-5744	13,05	12,67	11,89	11,20	11,77
SP70-1143	13,31	12,86	nd	nd	11,96
SP71-1406	13,19	12,92	nd	nd	11,87
SP79-1011	13,03	12,76	nd	11,38	11,75
SP79-2233	12,40	nd	nd	nd	11,65
SP80-1816	13,43	13,07	11,68	11,14	11,74
SP80-1842	13,39	12,83	11,60	11,16	12,04
SP80-3280	13,40	12,95	11,66	11,14	11,88
SP81-3250	13,18	13,14	11,42	11,12	11,90
SP83-2847	13,93	13,04	11,74	11,12	11,86
SP84-1201	12,59	12,84	nd	nd	nd
SP84-2025	13,26	nd	11,73	nd	nd
SP85-3877	12,11	nd	nd	10,93	nd
SP85-5077	12,42	12,80	nd	nd	nd
SP86-155	13,10	12,65	11,65	11,29	11,96
SP86-42	13,39	12,82	11,64	nd	12,03
SP87-344	12,56	12,78	11,70	nd	nd
SP87-365	13,03	12,98	nd	nd	11,77
SP89-1115	11,89	12,92	11,52	11,19	11,61
SP90-1638	13,02	12,96	nd	nd	nd
SP90-3414	13,20	12,89	11,63	11,26	nd
SP91-1049	12,63	12,83	11,61	11,11	11,87
SP91-3011	12,76	12,89	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região

Anexo F - Estimativas dos BLUPs para a característica Pol % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
CTC1	14,70	15,04	14,98	14,70	14,65
CTC2	16,09	15,20	14,88	14,48	15,13
CTC3	16,09	15,11	nd	14,50	15,02
CTC7	14,37	14,87	nd	14,64	14,73
CTC8	15,39	15,13	nd	nd	14,83
CTC11	15,30	15,01	nd	14,56	nd
CTC12	14,61	15,01	nd	nd	14,78
IAC48-65	15,31	15,07	nd	nd	nd
IAC64-257	14,39	nd	nd	nd	14,74
IAC82-2045	14,29	15,04	nd	14,38	14,81
IAC82-3092	13,99	15,16	nd	14,50	14,50
IAC83-4157	14,82	nd	nd	14,49	14,86
IAC86-2210	14,87	14,94	15,12	14,52	14,84
IAC86-2480	15,19	15,11	nd	nd	14,81
IAC87-3396	15,34	15,11	15,13	14,41	14,87
IAC91-1099	15,35	15,44	15,05	14,48	14,55
IAC91-2195	15,06	14,81	15,28	14,60	14,76
IAC91-2205	15,11	15,04	15,15	14,75	14,53
IAC91-2218	15,16	15,20	15,05	14,43	14,80
IAC91-3093	13,33	14,98	nd	nd	14,53
IAC91-4168	14,08	14,62	15,12	14,22	15,27
IAC91-5155	15,01	14,95	15,08	nd	14,91
IACSP93-1042	14,67	14,92	15,09	14,56	14,80
IACSP93-2060	15,18	14,79	nd	nd	15,12
IACSP93-3046	15,68	15,02	15,23	14,29	15,05
IACSP93-3050	14,34	14,84	15,12	nd	14,83
IACSP93-6006	14,71	14,99	15,05	14,42	14,91
IACSP94-2094	15,39	15,30	15,04	nd	14,68
IACSP94-2101	15,70	15,35	15,05	14,25	14,94
IACSP94-2180	15,21	15,01	nd	nd	14,92
IACSP94-4002	14,99	14,85	15,19	nd	14,90
IACSP94-4004	15,07	15,09	15,10	14,38	14,88
IACSP95-2048	14,73	15,18	15,03	14,47	14,70
IACSP95-2078	14,91	15,07	15,13	14,49	14,73
IACSP95-2213	15,07	15,09	15,00	nd	14,86

Anexo F - Estimativas dos BLUPs para a característica Pol % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
IACSP95-2289	15,57	15,23	15,09	14,55	14,70
IACSP95-3018	15,46	14,86	15,17	nd	15,01
IACSP95-3028	15,57	14,96	15,08	nd	15,03
IACSP95-5000	15,79	15,01	15,15	14,48	14,97
IACSP96-2019	14,45	14,92	15,17	nd	14,73
IACSP96-2022	14,49	nd	nd	nd	14,77
IACSP96-2037	14,84	14,96	14,90	14,54	15,00
IACSP96-2041	15,21	15,00	15,22	14,57	14,71
IACSP96-2042	15,02	14,90	15,22	14,53	14,80
IACSP96-2086	14,75	nd	nd	nd	14,83
IACSP96-2100	14,26	15,15	14,87	14,49	14,76
IACSP96-3057	14,61	15,15	14,97	14,47	14,77
IACSP96-7586	15,28	14,88	15,06	nd	15,06
IACSP97-2000	14,95	14,96	15,05	nd	14,92
IACSP97-2025	15,23	nd	nd	nd	14,93
IACSP97-2053	14,29	14,71	15,18	nd	14,89
IACSP97-2055	14,40	14,95	15,19	nd	14,67
IACSP97-2109	14,36	15,02	14,96	nd	14,81
IACSP97-3354	14,56	nd	14,96	nd	nd
IACSP97-3402	13,94	14,88	15,15	nd	14,67
IACSP97-3406	14,23	14,98	15,05	nd	14,74
IACSP97-6682	14,98	15,21	15,11	14,49	14,63
IACSP97-7065	14,68	14,96	15,00	nd	14,91
IACSP97-7077	15,26	14,77	15,13	nd	15,10
IACSP98-2071	15,64	nd	nd	nd	15,02
IACSP99-2033	15,96	nd	nd	nd	15,09
IACSP99-2058	15,17	nd	nd	nd	14,92
NA5679	14,91	14,98	nd	nd	nd
POJ2878	14,86	14,97	nd	nd	nd
RB72-454	15,24	15,33	15,32	14,09	14,75
RB83-5054	14,03	14,81	nd	nd	14,85
RB83-5486	14,95	14,87	15,20	14,41	14,95
RB84-5210	15,29	15,17	14,96	nd	nd
RB84-5257	15,75	14,96	nd	nd	15,08
RB85-5035	13,38	14,64	nd	nd	nd

Anexo F - Estimativas dos BLUPs para a característica Pol % Cana (%) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	16,27	15,12	15,23	14,33	15,04
RB85-5156	15,01	15,39	14,88	14,79	14,39
RB85-5453	14,19	14,81	14,69	14,77	14,99
RB85-5536	16,04	15,05	15,19	nd	14,93
RB86-7515	14,99	15,28	15,04	14,22	14,90
RB92-5211	14,24	14,56	14,87	14,83	15,01
RB92-5268	16,20	15,21	15,10	14,50	14,91
RB92-5345	14,16	14,43	15,04	14,93	14,86
RB92-579	15,81	15,56	14,86	14,29	14,92
RB92-8064	15,34	14,92	nd	14,40	15,13
RB93-5744	15,52	15,04	15,02	14,36	15,14
SP70-1143	14,99	14,99	nd	nd	14,89
SP71-1406	14,88	15,01	nd	nd	14,84
SP79-1011	15,20	15,12	nd	14,56	14,74
SP79-2233	14,66	nd	nd	nd	14,81
SP80-1816	15,87	14,83	15,16	14,44	15,21
SP80-1842	15,04	14,99	14,84	14,67	14,95
SP80-3280	15,74	14,84	15,23	14,52	15,02
SP81-3250	15,28	15,00	14,81	14,45	15,24
SP83-2847	15,79	15,06	14,94	14,26	15,35
SP84-1201	15,11	15,02	nd	nd	nd
SP84-2025	15,21	nd	15,11	nd	nd
SP85-3877	15,51	nd	nd	14,62	nd
SP85-5077	14,71	14,93	nd	nd	nd
SP86-155	14,84	15,21	14,94	14,46	14,80
SP86-42	16,07	15,12	15,01	nd	15,04
SP87-344	14,27	14,80	15,10	nd	nd
SP87-365	15,68	15,03	nd	nd	15,00
SP89-1115	14,27	14,55	14,95	14,71	15,06
SP90-1638	14,90	14,97	nd	nd	nd
SP90-3414	15,25	14,97	15,00	14,65	nd
SP91-1049	14,32	14,54	15,03	14,77	14,95
SP91-3011	14,42	14,87	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região

Anexo G - Estimativas dos BLUPs para a característica TCH (ton) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continua)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
CTC1	67,55	76,83	105,89	106,36	96,88
CTC2	88,27	73,45	108,74	113,43	105,23
CTC3	85,92	71,60	nd	119,62	107,12
CTC7	87,63	90,82	nd	112,56	96,18
CTC8	85,39	85,86	nd	nd	99,09
CTC11	93,66	84,48	nd	118,48	nd
CTC12	81,31	83,06	nd	nd	98,96
IAC48-65	76,11	77,34	nd	nd	nd
IAC64-257	82,32	nd	nd	nd	102,27
IAC82-2045	83,70	76,82	nd	118,26	101,65
IAC82-3092	84,11	82,16	nd	113,44	101,43
IAC83-4157	77,84	nd	nd	110,48	101,58
IAC86-2210	79,16	83,68	94,92	119,01	96,69
IAC86-2480	80,40	80,16	nd	nd	101,20
IAC87-3396	82,85	83,68	104,14	107,02	102,11
IAC91-1099	88,00	78,64	101,11	117,06	103,85
IAC91-2195	76,18	78,50	100,48	109,48	103,71
IAC91-2205	77,19	80,56	99,94	106,34	106,05
IAC91-2218	71,76	81,93	101,49	101,27	104,29
IAC91-3093	71,98	75,58	nd	nd	99,72
IAC91-4168	79,38	79,90	102,08	109,31	103,18
IAC91-5155	86,87	83,64	100,64	nd	102,57
IACSP93-1042	77,55	81,90	102,35	113,93	94,96
IACSP93-2060	81,50	80,77	nd	nd	101,38
IACSP93-3046	89,48	85,99	91,13	125,51	99,10
IACSP93-3050	81,65	81,80	98,08	nd	103,21
IACSP93-6006	82,07	79,08	101,62	105,56	110,13
IACSP94-2094	82,02	84,47	97,82	nd	101,07
IACSP94-2101	79,92	84,28	97,05	107,65	105,88
IACSP94-2180	71,40	76,18	nd	nd	98,70
IACSP94-4002	72,00	80,55	95,13	nd	100,48
IACSP94-4004	87,94	79,29	94,86	119,39	107,07
IACSP95-2048	67,61	80,68	98,69	114,27	92,36
IACSP95-2078	77,40	81,02	101,38	111,48	99,17
IACSP95-2213	82,06	81,47	96,37	nd	105,55

Anexo G - Estimativas dos BLUPs para a característica TCH (ton) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(continuação)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
IACSP95-2289	65,19	86,62	98,67	104,14	94,83
IACSP95-3018	76,06	79,65	100,88	nd	98,54
IACSP95-3028	79,02	83,78	99,61	nd	97,81
IACSP95-5000	84,51	81,20	97,09	116,14	103,72
IACSP96-2019	79,91	80,09	106,14	nd	95,61
IACSP96-2022	70,64	nd	nd	nd	93,88
IACSP96-2037	74,05	76,67	104,52	119,82	89,62
IACSP96-2041	75,58	89,78	101,87	108,57	91,51
IACSP96-2042	85,10	91,26	103,74	112,70	90,88
IACSP96-2086	63,37	nd	nd	nd	88,65
IACSP96-2100	64,86	78,74	101,44	111,88	91,95
IACSP96-3057	63,81	76,76	99,86	113,14	93,51
IACSP96-7586	79,84	78,09	100,34	nd	103,37
IACSP97-2000	75,42	76,92	97,15	nd	104,54
IACSP97-2025	82,67	nd	nd	nd	102,53
IACSP97-2053	85,81	79,55	102,03	nd	104,50
IACSP97-2055	81,69	69,82	105,46	nd	107,84
IACSP97-2109	77,27	73,66	100,05	nd	106,23
IACSP97-3354	92,60	nd	109,56	nd	nd
IACSP97-3402	85,28	76,28	106,14	nd	103,28
IACSP97-3406	80,68	75,55	106,73	nd	100,12
IACSP97-6682	82,66	86,37	100,94	111,62	97,88
IACSP97-7065	85,05	73,79	106,00	nd	105,74
IACSP97-7077	77,51	79,90	99,51	nd	100,70
IACSP98-2071	80,25	nd	nd	nd	100,79
IACSP99-2033	84,37	nd	nd	nd	103,75
IACSP99-2058	85,89	nd	nd	nd	104,84
NA5679	83,64	82,75	nd	nd	nd
POJ2878	72,82	74,97	nd	nd	nd
RB72-454	82,89	79,53	99,11	116,96	101,39
RB83-5054	64,16	70,68	nd	nd	99,00
RB83-5486	77,33	83,17	99,69	116,50	93,63
RB84-5210	81,93	73,69	108,66	nd	nd
RB84-5257	83,21	84,79	nd	nd	98,59
RB85-5035	69,69	72,72	nd	nd	nd

Anexo G - Estimativas dos BLUPs para a característica TCH (ton) para os 103 genótipos do painel de associação. As estimativas foram obtidas de experimentos conduzidos nas quatro regiões indicadas no período de 1993 a 2009. Foi utilizada uma abordagem de modelos mistos para a obtenção das estimativas

(conclusão)

Cultivar	Região				
	Geral	Goias	Jaú	Piracicaba	Rib, Preto
RB85-5113	81,08	81,14	95,74	115,52	103,30
RB85-5156	77,48	83,17	108,25	105,40	96,28
RB85-5453	76,22	84,23	101,56	109,90	96,50
RB85-5536	83,17	88,29	96,03	nd	99,87
RB86-7515	85,96	86,61	98,91	111,76	101,90
RB92-5211	76,36	78,73	102,45	106,18	104,93
RB92-5268	87,29	74,05	99,11	118,67	108,31
RB92-5345	73,30	81,03	100,67	106,88	101,52
RB92-579	80,50	71,89	109,47	106,56	107,35
RB92-8064	92,68	92,05	nd	106,35	104,79
RB93-5744	95,49	75,89	95,77	131,16	103,22
SP70-1143	75,27	76,23	nd	nd	101,45
SP71-1406	73,85	77,74	nd	nd	98,91
SP79-1011	79,61	88,00	nd	108,95	96,84
SP79-2233	92,40	nd	nd	nd	109,52
SP80-1816	77,24	84,57	98,23	111,35	98,77
SP80-1842	76,43	83,68	96,68	113,11	98,88
SP80-3280	84,75	85,01	101,07	110,10	102,15
SP81-3250	83,67	83,04	107,05	107,32	100,12
SP83-2847	88,34	76,58	97,39	120,40	106,53
SP84-1201	85,46	84,05	nd	nd	nd
SP84-2025	76,32	nd	97,86	nd	nd
SP85-3877	88,76	nd	nd	118,97	nd
SP85-5077	88,54	86,27	nd	nd	nd
SP86-155	86,78	87,63	101,13	112,54	98,48
SP86-42	82,44	80,32	95,70	nd	107,64
SP87-344	77,66	81,02	98,27	nd	nd
SP87-365	86,34	72,42	nd	nd	113,20
SP89-1115	76,61	67,61	102,43	119,42	103,00
SP90-1638	86,26	84,63	nd	nd	nd
SP90-3414	97,78	81,68	103,39	121,68	nd
SP91-1049	79,20	81,92	99,58	115,57	97,27
SP91-3011	79,68	79,90	nd	nd	nd

Nota:
nd: dado não disponível porque a cultivar não foi avaliada na respectiva região