

**Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”**

**Mapeamento de QTLs em múltiplos caracteres e ambientes em um cruzamento comercial de cana-de-açúcar usando modelos mistos**

**Gabriel Rodrigues Alves Margarido**

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

**Piracicaba  
2011**

**Gabriel Rodrigues Alves Margarido**  
**Engenheiro Agrônomo**

**Mapeamento de QTLs em múltiplos caracteres e ambientes em um cruzamento comercial  
de cana-de-açúcar usando modelos mistos**

**Orientador:**  
**Prof. Dr. ANTONIO AUGUSTO FRANCO GARCIA**

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**

Margarido, Gabriel Rodrigues Alves

Mapeamento de QTLs em múltiplos caracteres e ambientes em um cruzamento comercial de cana-de-açúcar usando modelos mistos / Gabriel Rodrigues Alves  
Margarido. - - Piracicaba, 2011.

107 p. : il.

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

1. Cana-de-açúcar 2. Correlação genética e ambiental 3. Mapeamento genético  
4. Marcador molecular 5. Melhoramento genético vegetal 6. Variação genética em plantas I. Título

CDD 633.61  
M327m

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

À minha mãe, **Lélia**, exemplo de paciência, perseverança e carinho. Ao meu pai, **Edson**, por me ensinar o valor da integridade e do esforço. A ambos pelo amor incondicional, por acreditarem sempre em meus sonhos e me apoiarem em todas as decisões.

**Dedico.**

“The highest reward for a man's toil is not what he gets for it but what he becomes by it.”

**John Ruskin**

“Success is not final, failure is not fatal: it is the courage to continue that counts.”

**Winston Churchill**

## Agradecimentos

A Deus, pela vida.

Ao Prof. Dr. Antonio Augusto Franco Garcia, por me mostrar a beleza da pesquisa e da ciência, pelos ensinamentos inestimáveis e grande amizade. Sua orientação nesses mais de nove anos mudou minha vida.

Ao Prof. Dr. Roland Vencovsky e à Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Lúcia Carneiro Vieira, por terem me acolhido de braços abertos quando precisei e pelo valioso complemento à minha formação. Aos demais professores do Departamento de Genética, pelas lições e estímulos à pesquisa.

Em especial aos meus irmãos Fábio e Paulo, por ouvirem pacientemente e sempre me auxiliarem com minhas dúvidas técnicas e pessoais. À minha irmã Maria Rita, por me acompanhar ao longo desse caminho.

Em especial à minha namorada Maily Rehder Wimmers Ferreira, por seu exemplo de vida e dedicação. Sua companhia constante e seu amor me dão forças para continuar.

À Universidade de São Paulo, à Escola Superior de Agricultura “Luiz de Queiroz” e ao Departamento de Genética, pela oportunidade e excelência de ensino.

À Prof<sup>a</sup>. Dr<sup>a</sup>. Anete Pereira de Souza, coordenadora do projeto PITE/FAPESP a partir do qual os dados aqui utilizados foram obtidos, pela valiosa parceria e incondicional apoio. À Dr<sup>a</sup>. Karine Miranda Oliveira, do Centro de Tecnologia Canavieira, por trabalhar nos dados moleculares e auxiliar na construção do mapa de ligação.

Ao CNPq, pela bolsa de Iniciação Científica e parte da bolsa de Doutorado, e principalmente à FAPESP, pelos auxílios financeiros durante a graduação e o Doutorado.

A todo o pessoal do laboratório de Genética-Estatística: Priscilla, Millor, Rodrigo, Maria Marta, Marcelo, Graciela, Renato, Edjane, João Ricardo, Carina, Luciano, Guilherme e Adriana, pela convivência, amizade e diversão em todos os momentos, além do suporte nos momentos de dificuldade.

Ao Marcelo, Rodrigo e Maria Marta, por terem desbravado o  $\text{\LaTeX}$  e colocado a tese dentro das normas.

À Maria Marta, por toda sua ajuda no emprego dos modelos mistos e no uso do GenStat.

Ao Marcelo, pela convivência em Los Angeles, pelo apoio e suporte.

Ao Luciano, que em pouco tempo tanto contribuiu para meu entendimento do MIM.

Ao Prof. Dr. Fred van Eeuwijk, ao Prof. Dr. Marcos Malosetti e ao Prof. Dr. Martin Boer, pela valiosa ajuda na análise dos dados.

Aos colegas de pós-graduação, em particular Bruno Marco, Fernando Piotto, Thiago Falda, Inocência, Paulinha, Diego, Gustavo, Luciana, Flavinha, Carla, Alessandra, Larissa e todos os demais, pelos estudos, companhia nas disciplinas e pela diversão nas festas.

Aos funcionários do Departamento, em especial o Seu Zé, Berdan, Valdir, Seu Antônio, Léia, Fernandinho, Macedônio, Márcia, Maídia e Glória, pelo caloroso acolhimento e socorro nos apuros.

Aos moradores da Rep.  $1/2$ -D-Kultura: Chuck, Tãπko, K-ju, Skρ, K-μña, Remédio,  $3/4$  e C-bá, e também ao amigo Zé-Ag<sup>+3</sup>, pelas conversas, festas, risadas e provocações sem fim.

Enfim, a todos os amigos e familiares que, direta ou indiretamente, contribuíram para minha formação e para a construção desta tese, meus sinceros agradecimentos.

## SUMÁRIO

RESUMO . . . . .	9
ABSTRACT . . . . .	11
1 INTRODUÇÃO . . . . .	13
2 REVISÃO . . . . .	17
2.1 Cana-de-açúcar . . . . .	17
2.2 Mapeamento Genético . . . . .	18
2.3 Mapeamento de QTLs . . . . .	24
2.3.1 Métodos de Mapeamento . . . . .	24
2.3.2 QTLs em Cana-de-açúcar . . . . .	28
2.4 Modelos Mistos . . . . .	34
2.4.1 Análise de dados de múltiplos caracteres e múltiplos ambientes através de Modelos Mistos . . . . .	40
2.4.2 Mapeamento de QTLs . . . . .	42
3 MATERIAL E MÉTODOS . . . . .	47
3.1 Material . . . . .	47
3.1.1 Material vegetal . . . . .	47
3.1.2 Dados Moleculares . . . . .	47
3.1.3 Dados Fenotípicos . . . . .	47
3.2 Métodos . . . . .	48
3.2.1 Mapa de Ligação . . . . .	48
3.2.2 Preditores Genéticos . . . . .	50
3.2.3 Análise Fenotípica . . . . .	52
3.2.4 Análise de QTLs . . . . .	54
3.2.4.1 Estratégia de Busca por QTLs . . . . .	55
3.2.4.2 Etapas Finais . . . . .	56
4 RESULTADOS . . . . .	59
4.1 Mapa de Ligação . . . . .	59
4.2 Análise Fenotípica . . . . .	59
4.3 Análise de QTLs . . . . .	66
5 DISCUSSÃO . . . . .	75



6 CONCLUSÃO . . . . .	85
REFERÊNCIAS . . . . .	87

## RESUMO

### **Mapeamento de QTLs em múltiplos caracteres e ambientes em um cruzamento comercial de cana-de-açúcar usando modelos mistos**

Dados coletados de experimentos de melhoramento geralmente contêm observações fenotípicas para vários caracteres avaliados em múltiplos ambientes. Especificamente para cana-de-açúcar, medidas repetidas são obtidas para cana-planta e uma ou mais socas. Tal cenário presta-se naturalmente ao uso de modelos mistos para modelagem de variâncias genéticas heterogêneas e correlações entre caracteres, locais e cortes. Essa abordagem também nos permite incluir informações de marcadores moleculares, auxiliando no entendimento da arquitetura genética dos caracteres quantitativos através do mapeamento de QTLs. Este trabalho teve como objetivo detectar QTLs e interação QTL por ambiente pelo método do Mapeamento de Múltiplos Intervalos, o qual também permite a inclusão de epistasia no processo de busca. Nossa população de mapeamento foi composta por 100 indivíduos oriundos de um cruzamento biparental entre os cultivares brasileiros pré-comerciais SP80-180 e SP80-4966, avaliados em duas localidades (Piracicaba e Jaú, SP, Brasil) e três anos (2004 a 2006) para percentual de fibra, conteúdo de sacarose (POL) e toneladas de cana por hectare (TCH). Um mapa genético com 96 grupos de ligação cobrindo 2468,14 cM já estava disponível para esse cruzamento. O modelo fenotípico selecionado conteve matrizes de covariância separadas para caracteres e ambientes, o que resultou em um modelo mais parcimonioso com bom ajuste aos dados. Foram detectados 13 QTLs com efeitos principais e 8 interações epistáticas, cada um deles exibindo interação QTL por local, QTL por corte ou a interação tripla. Assim, nenhum QTL apresentou efeitos estáveis ao longo de todas as combinações de ambientes. No total, 13 dos 21 efeitos apresentaram algum grau de pleiotropia, afetando pelo menos dois dos três caracteres. Além disso, esses QTLs sempre afetaram os caracteres fibra e TCH na mesma direção, enquanto que POL foi afetado no sentido oposto. Não foi observada evidência em favor da hipótese de QTLs ligados em detrimento da hipótese de QTL pleiotrópico para qualquer das posições genômicas detectadas. Esses resultados fornecem valiosas informações sobre a base genética da variação quantitativa em cana-de-açúcar e sobre a relação genética entre caracteres.

Palavras-chave: Interação QTL  $\times$  E; Correlação genética; Mapeamento de Múltiplos Intervalos; Análise de ligação; Progenie de irmãos completos



## ABSTRACT

**Multi-trait multi-environment QTL mapping in a sugarcane commercial cross using mixed models**

Data collected from breeding trials usually comprise phenotypic observations for various traits evaluated at multiple test environments. Specifically for sugarcane, repeated measures are obtained for plant crop and one or more ratoons. Such scenario naturally lends itself to the use of mixed models for modeling heterogeneous genetic variances and correlations between traits, locations and harvests. This modeling approach also enables us to include molecular marker information, aiding in understanding the genetic architecture of quantitative traits through QTL mapping. Our work was aimed at detecting QTL and QTL by environment interaction by the Multiple Interval Mapping method, which also allows the inclusion of epistasis in the search process. Our mapping population was composed of 100 individuals derived from a biparental cross between the Brazilian pre-commercial cultivars SP80-180 and SP80-4966, evaluated at two locations (Piracicaba and Jaú, SP, Brazil) and three harvest years (2004 through 2006) for fiber content, sugar content (POL) and tonnes of cane per hectare (TCH). A genetic linkage map with 96 linkage groups covering 2468.14 cM was already available for this cross. The selected phenotypic model contained separate covariance matrices for traits and environments, which resulted in a more parsimonious model with good fit to the data. We detected 13 QTL with main effects and 8 epistatic interactions, each exhibiting QTL by location, QTL by harvest or the three-way interaction. Thus, no QTL displayed stable effects across all environment combinations. Overall, 13 of the 21 effects presented some degree of pleiotropy, affecting at least two of the three traits. Furthermore, these QTL always affected fiber and TCH in the same direction, while POL was affected in the opposite way. There was no evidence in favor of the linked QTL over the pleiotropic QTL hypothesis for any of the detected genome positions. These results give valuable insights about the genetic basis of quantitative variation in sugarcane and about the genetic relation between traits.

Keywords: QTL  $\times$  E interaction; Genetic correlation; Multiple Interval Mapping; Linkage analysis; Full-sib progeny



## 1 INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp.) está entre as mais importantes espécies cultivadas, principalmente para os países não-desenvolvidos, por ser a principal fonte de sacarose, do etanol desta derivado e de vários subprodutos obtidos no seu processamento (MATSUOKA; GARCIA; ARIZONO, 1999). Para que a produção consiga acompanhar o contínuo crescimento da demanda mundial, deve haver aumento tanto da área plantada quanto da produtividade agrícola da cultura, o que pode ser conseguido através de aumento na quantidade de cana produzida por unidade de área e no teor de sacarose no caldo (FNP – Consultoria & Comércio, 2006). Claramente, o melhoramento genético tem papel preponderante neste aumento de produtividade, devido à sua importância na seleção de cultivares que apresentem características agrônômicas e industriais favoráveis.

Botanicamente, a cana-de-açúcar é uma planta alógama, da família *Poaceae*, gênero *Saccharum*, o qual faz parte de um complexo de espécies poliplóides e altamente polimórficas. Alguns autores aceitam a ocorrência de seis espécies no gênero *Saccharum*, a saber: *S. officinarum* L. ( $2n = 80$ ), *S. robustum* Brandes e Jeswiet ex Grassl ( $2n = 60 - 205$ ), *S. barberi* Jeswiet ( $2n = 81 - 124$ ), *S. sinense* Roxb. ( $2n = 111 - 120$ ), *S. spontaneum* L. ( $2n = 40 - 128$ ) e *S. edule* Hassk ( $2n = 60-80$ ) (DANIELS; ROACH, 1987 apud MATSUOKA; GARCIA; CALHEIROS, 1999; IRVINE, 1999). Os primeiros trabalhos de melhoramento da cana-de-açúcar envolviam o cruzamento entre *S. officinarum* e *S. spontaneum*, seguido por retrocruzamentos com a primeira destas, prática que resultou em aumento da complexidade do genoma, de modo que os materiais atuais apresentam alto grau de ploidia com frequente aneuploidia (KIDO, 2003; MATSUOKA; GARCIA; ARIZONO, 1999; SOUZA, 2003).

Em linhas gerais, o melhoramento genético desta espécie baseia-se na seleção e clonagem de genótipos superiores de populações segregantes obtidas através de cruzamentos biparentais entre indivíduos heterozigóticos. É aplicada seleção fenotípica para caracteres de importância comercial e agrônômica, os quais geralmente são de natureza quantitativa, pois são controlados por muitos genes e sofrem forte influência ambiental, o que resulta em padrão contínuo de variação fenotípica (FALCONER; MACKAY, 1996). Esses programas de melhoramento podem ser muito beneficiados pelo desenvolvimento de métodos que possibilitem uma seleção mais precoce e eficiente de genótipos superiores, como os estudos baseados em dados de marcadores moleculares, notadamente estudos da arquitetura genética através do mapeamento de QTLs (*Quantitative Trait Loci*), isto é,

regiões cromossômicas contendo locos que controlam caracteres poligênicos (FALCONER; MACKAY, 1996; LIU, 1998; LYNCH; WALSH, 1998; ZENG, 2001).

Para espécies em que é difícil ou impossível obter linhagens endogâmicas, como a cana-de-açúcar, as análises só podem ser realizadas em progênies de irmãos completos oriundas do cruzamento de dois indivíduos não endogâmicos. A população segregante gerada por este procedimento apresenta uma mistura de diferentes padrões de segregação para os locos marcadores (WU et al., 2002a). Além disso, a fase de ligação entre marcadores é desconhecida, o que dificulta a construção de mapas genéticos e o mapeamento de QTLs (LIN et al., 2003). Apesar de haver alguns poucos métodos delineados especificamente para mapear QTLs nessas progênies (SONG; SOLLER; GENIZI, 1999; JOHNSON; JANSEN; ARENDONK, 1999; LIN et al., 2003; GAZAFFI, 2009; PASTINA, 2010), geralmente têm sido empregadas as simples análises de marcas individuais ou os já estabelecidos métodos para retrocruzamentos, como Mapeamento por Intervalo (IM) e Mapeamento por Intervalo Composto (CIM) (LANDER; BOTSTEIN, 1989; ZENG, 1993, 1994; JANSEN; STAM, 1994). Para tanto, é comumente utilizada a abordagem do duplo *pseudo-testcross*, a qual fornece dois mapas de ligação, sendo que em cada um deles há marcadores segregando exclusivamente para um determinado genitor (GRATTAPAGLIA; SEDEROFF, 1994). Até o momento, o sofisticado método do Mapeamento de Múltiplos Intervalos (MIM) (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999) não foi expandido para uso nessas progênies. Tal metodologia é adequada por permitir a inclusão de interações epistáticas durante o processo de busca por QTLs.

Especificamente no caso da cana-de-açúcar, o elevado grau de ploidia representa uma dificuldade adicional, devido ao complexo padrão de segregação dos cromossomos na meiose (HEINZ; TEW, 1987). Para espécies poliplóides, o mapeamento pode ser realizado de maneira equivalente à do duplo *pseudo-testcross*, através de fragmentos em dosagem única no genoma, denominados marcadores SDRF (*Single Dose Restriction Fragment*). Tais marcadores podem apresentar segregação tipo 1:1 ou 3:1, quando presentes em apenas um ou em ambos os genitores, respectivamente, já que o padrão de polimorfismo é representado pela presença ou ausência de bandas individuais (WU et al., 1992). Diversos trabalhos empregaram análises de marcas individuais, IM ou CIM em dados dessa natureza (SILLS et al., 1995; DAUGROIS et al., 1996; MING et al., 2001, 2002a, 2002c; HOARAU et al., 2002; JORDAN et al., 2004; SILVA; BRESSIANI, 2005; McINTYRE et al., 2005a, 2005b, 2006; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2006; RABOIN et al., 2006, 2008; WEI et al., 2006; AL-JANABI et al., 2007; AITKEN et al., 2008; PIPERIDIS

et al., 2008; PINTO et al., 2010). Mais recentemente, tem havido uma tendência no sentido de integrar as informações contidas nos mapas individuais em um único mapa, a partir da presença de marcadores em heterozigose em ambos os genitores (GARCIA et al., 2006). A construção de um mapa genético integrado apresenta inúmeras vantagens, pois permite aumentar a saturação do mapa de ligação e estender a caracterização da variação polimórfica em todo o genoma.

Devido à natureza perene da cana-de-açúcar, os mesmos indivíduos permanecem no campo por vários anos após o plantio e podem ser submetidos a vários cortes, anualmente, dando origem às chamadas “socas” da cana. Estatisticamente, isto dá origem a dados de medidas repetidas ao longo do tempo (SMITH et al., 2007). Além disso, em programas de melhoramento genético é comum a avaliação dos materiais genéticos em diferentes locais e, adicionalmente, esses experimentos comumente envolvem a avaliação de diversos caracteres simultaneamente, visto que materiais elite devem concentrar alelos favoráveis para muitas características (WELHAM et al., 2010). Dados de múltiplos caracteres e múltiplos ambientes, denominados MTME (*multi-trait multi-environment*), permitem o estudo da interação genótipo por ambiente ( $G \times E$ ) e de sua base genética, a interação QTL por ambiente ( $QTL \times E$ ), bem como da correlação entre caracteres e do entendimento de suas causas, a saber, pleiotropia e ligação genética (JIANG; ZENG, 1995; MALOSETTI et al., 2004, 2008). Claramente, a modelagem direta de todas as fontes de variação e covariação em uma única análise MTME aumenta o poder estatístico e fornece informações de grande utilidade para o melhorista, que podem ser canalizadas para programas de seleção assistida por marcadores (DEKKERS; HOSPITAL, 2002; BOER et al., 2007; MALOSETTI et al., 2008).

A ocorrência de correlações genéticas tanto entre caracteres quanto entre ambientes, aliada à heterogeneidade de variâncias, torna os modelos mistos particularmente úteis para a análise de dados MTME, em razão da flexibilidade de modelagem de estruturas de variância-covariância para os termos aleatórios (SMITH; CULLIS; THOMPSON, 2001; SMITH et al., 2007; MALOSETTI et al., 2008). Adicionalmente, é possível fazer suposições mais realistas sobre a estrutura dos erros, o que reduz vieses na análise (BALZARINI, 2001). Assim, essa classe de modelos fornece embasamento adequado para o estudo das interações  $G \times E$  (PIEPHO, 1997; CULLIS et al., 1998; SMITH; CULLIS; THOMPSON, 2001, 2005; SMITH et al., 2007) e  $QTL \times E$  (PIEPHO, 2000; VERBYLA et al., 2003; MALOSETTI et al., 2004, 2006, 2008; BOER et al., 2007). Por estas razões, o presente trabalho teve como objetivo empregar modelos mistos para mapear QTLs em um conjunto de dados MTME de cana-de-açúcar, através de uma expansão da estratégia de busca do método MIM para



progênies segregantes obtidas pelo cruzamento de genitores não homozigóticos.

## 2 REVISÃO

### 2.1 Cana-de-açúcar

A cana-de-açúcar (*Saccharum* spp.) está entre as mais importantes espécies cultivadas, principalmente para os países não-desenvolvidos. É a principal fonte de sacarose e ainda possibilita a extração de vários subprodutos a partir do caldo e do bagaço obtidos no seu processamento (MATSUOKA; GARCIA; ARIZONO, 1999; MEZA, 2002). O Brasil é o principal produtor mundial, sendo que na safra 2010/2011 a área nacional colhida foi de aproximadamente 8,8 milhões de hectares, resultando numa produção próxima a 700 milhões de toneladas. Nos últimos anos têm ocorrido aumentos consistentes da demanda por açúcar e álcool, neste último caso devido ao forte aumento das vendas de carros e motos *flex fuel*. Externamente, a relação estoque sobre consumo de açúcar está abaixo de 20% há três anos e os preços estão em alta, e projeta-se que o Brasil continue batendo recordes de exportação (AgraFNP, 2011).

Segundo projeções, nos próximos dez anos deve haver crescimento de aproximadamente 70% na área plantada com cana-de-açúcar, enquanto que a produção deve crescer 100%. Para que isto seja possível, deve haver um aumento de produtividade agrícola da cultura da ordem de 12%, o que pode ser conseguido através de aumento na quantidade de cana produzida por unidade de área e no teor de sacarose no caldo (FNP – Consultoria & Comércio, 2006). Claramente, o melhoramento genético da cana-de-açúcar tem papel preponderante neste aumento de produtividade, devido à sua importância na seleção de cultivares que apresentem características agronômicas e industriais favoráveis.

Botanicamente, a cana-de-açúcar é uma planta alógama, da família *Poaceae*, tribo *Andropogoneae*, subtribo *Saccharinae*, gênero *Saccharum*. Alguns autores aceitam a ocorrência de seis espécies no gênero *Saccharum*, a saber: *S. officinarum* L. ( $2n = 80$ ), *S. robustum* Brandes e Jeswiet ex Grassl ( $2n = 60 - 205$ ), *S. barberi* Jeswiet ( $2n = 81 - 124$ ), *S. sinense* Roxb. ( $2n = 111 - 120$ ), *S. spontaneum* L. ( $2n = 40 - 128$ ) e *S. edule* Hassk ( $2n = 60-80$ ) (DANIELS; ROACH, 1987 apud MATSUOKA; GARCIA; CALHEIROS, 1999; IRVINE, 1999).

*S. officinarum* L. é um complexo poliplóide e é conhecida como “cana-nobre”, devido aos colmos grossos e suculentos, com alto conteúdo de sacarose e boas características gerais para a industrialização. *S. spontaneum* L. é uma espécie altamente polimórfica e apresenta alta adaptabilidade a diversas condições ambientais (MATSUOKA; GARCIA; ARIZONO, 1999). Os primeiros traba-

lhos de melhoramento da cana-de-açúcar envolviam o cruzamento entre estas duas espécies, seguido por retrocruzamentos com *S. officinarum*. No entanto, devido à maior contribuição citológica e gamética desta espécie, houve um aumento da complexidade do genoma da cana-de-açúcar, de modo que os materiais atuais apresentam alto grau de ploidia com frequente aneuploidia (KIDO, 2003; MATSUOKA; GARCIA; ARIZONO, 1999; SOUZA, 2003). Notadamente, esta alta complexidade genética dificulta o melhoramento da cana-de-açúcar (HEINZ; TEW, 1987).

Em linhas gerais, o melhoramento genético desta espécie baseia-se na seleção e clonagem de genótipos superiores de populações segregantes, obtidas através de cruzamentos sexuais entre indivíduos diferentes. No entanto, as seleções realizadas nas etapas iniciais não podem apresentar alta intensidade, pois há poucos clones de cada genótipo, o que torna as medidas fenotípicas muito sujeitas à influência ambiental (caracteres de mediana e baixa herdabilidade), comprometendo a seleção de materiais promissores. Portanto, a liberação de uma nova variedade somente ocorrerá quando houver muitos resultados experimentais, em diferentes locais e anos de cultivo. Em função disso, um programa de melhoramento demanda aproximadamente 12 anos para a obtenção de um novo cultivar (MATSUOKA; GARCIA; ARIZONO, 1999).

Neste contexto, os programas de melhoramento genético dessa espécie podem ser muito beneficiados pelo desenvolvimento de métodos que possibilitem uma seleção mais precoce e eficiente de genótipos superiores. Estudos baseados em dados de marcadores moleculares, como a seleção assistida por marcadores, têm, portanto, grande potencial de utilização (LU et al., 2003 apud DEKKERS; HOSPITAL, 2002).

## 2.2 Mapeamento Genético

Qualquer genoma possui diversas formas alélicas, ou polimorfismos, que podem ser empregadas como marcadores genéticos. Essas variações podem ser fenotípicas, oriundas de proteínas ou de fragmentos de DNA, sejam estes codantes ou não. Os marcadores fenotípicos, isoenzimáticos ou bioquímicos correspondem ao produto de expressão de um ou mais genes e podem ser influenciados pelo ambiente e por outros genes (SOUZA, 2001). Além disso, seu pequeno número dificultava a detecção de associações entre eles e caracteres de importância agrônômica. Não obstante, esses marcadores contribuíram significativamente para análises de ligação gênica e para a construção dos primeiros mapas genéticos (FERREIRA; GRATTAPAGLIA, 1998).

Já os marcadores moleculares detectam o polimorfismo diretamente no nível de DNA e, por-

tanto, não são influenciados pelo ambiente ou por outros genes. Seu número é virtualmente ilimitado, e usualmente eles são seletivamente neutros e desprovidos de efeito epistático ou pleiotrópico (SOUZA, 2001). De maneira simplificada, os marcadores moleculares podem ser divididos em codominantes, em que os alelos podem ser visualizados simultaneamente e todas as classes genótípicas podem ser distinguidas, e dominantes, nos quais a presença de um alelo mascara a presença de outros alelos, de modo que há mistura de classes genótípicas (SOUZA, 2001). Entre os codominantes, destacam-se o RFLP (*Restriction Fragment Length Polymorphism*), EST-RFLP (*Expressed Sequence Tag RFLP*), microssatélites ou SSR (*Simple Sequence Repeat*), EST-SSR e SNPs (*Single Nucleotide Polymorphism*). Os dominantes englobam principalmente os marcadores do tipo RAPD (*Random Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*) e DAiT (*Diversity Array Technology*) (SCHLÖTTERER, 2004).

Dentre as várias aplicações dos marcadores moleculares, destacam-se estudos de diversidade genética (WOUW et al., 2010), construção de mapas de ligação (SEMAGN; BJØRNSTAD; NDJONDJOP, 2006), mapeamento de QTLs (MACKAY, 2001b) e seleção assistida por marcadores (DEKKERS; HOSPITAL, 2002). Esses estudos fornecem informações detalhadas a respeito da organização genômica e contribuem marcadamente para a compreensão da arquitetura genética de diversos caracteres em várias espécies vegetais (SCHLÖTTERER, 2004).

Em particular, os mapas de ligação prestam-se a diversos tipos de aplicação, tanto para estudos de cunho teórico quanto para aplicações no melhoramento genético de plantas. Vale destacar seu emprego em trabalhos de mapeamento de QTLs (MOHAN et al., 1997; DOERGE, 2002), como fator facilitador da seleção assistida por marcadores (DEKKERS; HOSPITAL, 2002), na clonagem posicional de genes (MOHAN et al., 1997; VUYLSTEKE et al., 1999; FRARY et al., 2000), mapeamento comparativo entre espécies evolutivamente relacionadas (AHN; TANKSLEY, 1993; PATERSON et al., 2000), como arcabouço para ancoragem de mapas físicos (YIM et al., 2002) e também na montagem de genomas (LEWIN et al., 2009), entre outros.

A grande maioria das metodologias estatísticas desenvolvidas para construção de mapas genéticos é voltada para populações segregantes de espécies diplóides e oriundas de linhagens homocigóticas. Para muitas espécies é difícil ou impossível obter linhagens endogâmicas, como é o caso de espécies florestais, frutíferas e da cana-de-açúcar, devido ao longo intervalo entre gerações, à forte depressão por endogamia ou à complexidade do genoma. Neste caso, as análises só podem ser realizadas em populações  $F_1$ , isto é, progênies de irmãos completos oriundas do cruzamento de

dois indivíduos não endogâmicos. A população segregante gerada por este procedimento apresenta número de alelos que varia de um a quatro por loco, além da possibilidade de haver alelos nulos. Assim, ocorre uma mistura de diferentes tipos de marcadores, ou seja, há diferentes padrões de segregação possíveis para cada loco marcador segregante (WU et al., 2002a). Além disso, a fase de ligação entre os marcadores é desconhecida, o que dificulta a construção de mapas genéticos e o mapeamento de QTLs (LIN et al., 2003). Especificamente no caso da cana-de-açúcar, o elevado grau de ploidia representa uma dificuldade adicional, devido ao complexo padrão de segregação dos cromossomos na meiose (HEINZ; TEW, 1987).

Uma alternativa comumente utilizada para construir mapas nessas populações é a de duplo *pseudo-testcross*, na qual a população é genotipada com marcadores que apresentem segregação do tipo 1:1. Essa estratégia fornece dois mapas de ligação, sendo que em cada um deles há marcadores segregando exclusivamente para um determinado genitor, uma vez que o outro genitor apresenta alelos fixados para os mesmos locos (GRATTAPAGLIA; SEDEROFF, 1994). Neste caso, a fase de ligação não é conhecida *a priori*, mas inferida *a posteriori* com base na segregação da progênie.

No caso de poliplóides, o mapeamento pode ser realizado através de fragmentos em dosagem única no genoma, isto é, fragmentos presentes em apenas um cromossomo, denominados marcadores SDRF (*Single Dose Restriction Fragment*). Tais marcadores podem apresentar segregação tipo 1:1 ou 3:1, quando presentes em apenas um ou em ambos os genitores, respectivamente, já que o padrão de polimorfismo é representado pela presença ou ausência de bandas individuais. Neste caso, marcadores RFLP e microssatélites passam a funcionar como dominantes (SILVA et al., 1995). De fato, a construção de mapas genéticos em cana-de-açúcar só teve início depois que Wu et al. (1992) propuseram uma estratégia de mapeamento em poliplóides com pareamento bivalente, a qual faz uso exatamente desses marcadores em dose única. Do ponto de vista teórico, tal abordagem equivale à do duplo *pseudo-testcross*, já que é obtido um mapa de ligação para cada genitor.

Vários mapas genéticos já foram construídos para a cana-de-açúcar utilizando-se marcadores do tipo SDRF (AL-JANABI et al., 1993, 2007; SOBRAL; HONEYCUTT, 1993; SILVA et al., 1993, 1995; D'HONT et al., 1994; GRIVET et al., 1996; MUDGE et al., 1996; MING et al., 1998, 2002b; GUIMARÃES et al., 1999; HOARAU et al., 2001; ROSSI et al., 2003; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2005, 2007; OLIVEIRA et al., 2007; RABOIN et al., 2006; GARCIA et al., 2006; ALWALA et al., 2008). Uma descrição bastante detalhada desses mapas é feita por Pastina et al. (2010), de modo que será apresentado aqui apenas um breve relato de suas

principais características e resultados.

Originalmente, a estratégia de duplo *pseudo-testcross* com marcadores SDRF foi empregada para construção dos primeiros mapas. Silva et al. (1993) e Al-Janabi et al. (1993) utilizaram o cruzamento entre o clone SES208 de *S. spontaneum* e seu duplo-haplóide derivado por cultura de anteras, ADP068, com marcadores do tipo RFLP e RAPD, respectivamente. Em ambos os mapas, o número de marcadores ficou próximo de 180, alocados em pouco mais de 40 grupos de co-segregação, de tal forma que a cobertura do genoma foi de aproximadamente 1.500 cM. Posteriormente, Silva et al. (1995) reuniram os dados desses dois trabalhos e obtiveram um mapa com 64 grupos de co-segregação, cobrindo 3.300 cM no total.

D'Hont et al. (1994) empregaram dados de isoenzimas e RFLP de uma pequena população de 32 indivíduos oriunda por autofecundação da variedade SP 701006 para obter 25 grupos de co-segregação. Uma progênie obtida por autofecundação também foi utilizada por Grivet et al. (1996), neste caso derivada do cultivar R570. Neste trabalho foram alocados 408 marcadores (RFLP e isoenzima) em 96 grupos de co-segregação, quantia que se aproxima bastante do número de cromossomos usualmente observado em materiais comerciais. Esses autores também investigaram a origem dos marcadores, separando-os em marcas derivadas de *S. spontaneum* ou de *S. officinarum*, o que permitiu verificar a cobertura de cada genoma e a ocorrência de recombinação entre eles. Posteriormente, essa progênie foi expandida de 77 para 295 indivíduos e, através de quase 1.000 marcadores do tipo AFLP, foi utilizada para construção de 120 grupos de co-segregação com cobertura de 5.849 cM, o que representa um terço do genoma (HOARAU et al., 2001). Rossi et al. (2003) utilizaram um subconjunto dessa mesma população, com marcadores RFLP e SSR derivados de EST, para obter um mapa com 128 grupos de co-segregação. O cultivar R570 também foi empregado para gerar outras populações segregantes, em um cruzamento com o clone MQ76-53, em que mais de 1.000 marcadores AFLP, RFLP e SSR foram genotipados (RABOIN et al., 2006), e em um cruzamento com a variedade M134/75, em que um mapa cobrindo mais de 6.000 cM foi construído (Al-JANABI et al., 2007).

Devido à origem interespecífica da maioria dos clones comerciais de cana-de-açúcar, diversos cruzamentos entre as espécies genitoras foram realizados para estudar a organização do genoma. Primeiramente, Mudge et al. (1996) cruzaram o material La Purple de *S. officinarum* com Mol 5829 de *S. robustum* para obter 50 grupos de co-segregação com marcadores RAPD. Em seguida, Guimarães et al. (1999) utilizaram o mesmo cruzamento para obter mapas com maior cobertura do

genoma, empregando para tanto marcas do tipo RAPD, RFLP e AFLP. Ming et al. (2002b) realizaram dois cruzamentos, o primeiro entre PIN 84-1 de *S. spontaneum* e Muntok Java de *S. officinarum*, obtendo mais de 70 grupos de co-segregação, e o segundo entre Green German (*S. officinarum*) e IND 81-146 (*S. spontaneum*), obtendo números de grupos de co-segregação semelhantes aos do cruzamento anterior. Marcadores do tipo RFLP foram empregados em ambos os cruzamentos. Mais recentemente, Alwala et al. (2008) cruzaram os materiais La Striped (*S. officinarum*) e SES 147B (*S. spontaneum*). Este trabalho é marcado pelo uso de marcadores AFLP somados a outros obtidos por técnicas mais modernas, como SRAP (*Sequence Related Amplified Polymorphism*) e TRAP (*Target Region Amplified Polymorphism*). Mapas assim obtidos podem ser canalizados para estudos de mapeamento comparativo entre espécies relacionadas, de modo que a evolução e organização dos genomas possam ser melhor compreendidas. Especificamente no caso do gênero *Saccharum*, as espécies de maior interesse são o milho e o sorgo, ambas da tribo Andropogonae, sendo este último o diplóide mais proximalmente relacionado à cana-de-açúcar (MING et al., 2002a).

O cruzamento entre o clone IJ76-514 e a variedade Q165 foi utilizado, em dois trabalhos distintos, para obtenção de um mapa para cada genitor, através de marcadores AFLP, RAF (*Randomly Amplified DNA Fingerprints*) e SSR (AITKEN; JACKSON; McINTYRE, 2005, 2007). O primeiro destes trabalhos, no qual foi construído o mapa da variedade Q165, representa o maior mapa já obtido para cana-de-açúcar, com mais de 1.000 marcadores alocados em 136 grupos de co-segregação e com comprimento total de 9.058,3 cM. O mapa de IJ76-514, apesar de menor, também é bastante longo, cobrindo 4.906,4 cM no total. Outro ponto bastante interessante desses dois trabalhos é que eles fizeram uso de marcadores em dose dupla no genoma, os quais segregam em uma proporção 11:3 e, portanto, não se encaixam nos padrões observados para *testcrosses*. No entanto, vale ressaltar que a estratégia de inclusão dos mesmos não é baseada em análises multiponto, pois eles foram inseridos, a partir de informações obtidas de análises de dois pontos, em um mapa previamente construído para os marcadores em dose única.

Recentemente, Garcia et al. (2006) obtiveram um mapa único, ou integrado, para uma progênie oriunda do cruzamento entre os cultivares pré-comerciais SP80-180 e SP80-4966. Neste trabalho, 357 marcadores RFLP, AFLP e SSR foram alocados em 131 grupos de co-segregação, com comprimento total de 2.602,4 cM. A construção deste mapa foi realizada através do software chamado *OneMap* (MARGARIDO; SOUZA; GARCIA, 2007). Tal abordagem merece destaque por ser o primeiro mapa construído em cana-de-açúcar que não fez uso da estratégia de duplo *pseudo-testcross*.

Posteriormente, Oliveira et al. (2007) incluíram marcadores RFLP e SSR derivados de EST neste mapa integrado, ampliando-o para 192 grupos de co-segregação e 6.261,1 cM. É importante notar que vários desses grupos de co-segregação são compostos por apenas duas marcas ligadas, provavelmente em razão da saturação insuficiente do genoma, o que causa fragmentação dos grupos de ligação em grupos menores. Isto explica o fato de o número de grupos de co-segregação encontrado ser bastante superior ao número esperado de cromossomos.

Finalmente, Andru et al. (2011) empregaram marcadores AFLP, SSR e TRAP, não apenas em dose única mas também presentes em duas cópias, para construir um mapa do cultivar norte-americano LCP 85-384 a partir de uma progênie de 300 indivíduos obtida por autofecundação. O mapa assim obtido foi composto por 784 marcadores alocados em 108 grupos de co-segregação, com comprimento total de 5.617 cM, o que corresponde a aproximadamente 45,6% do genoma.

A integração das informações contidas nos mapas individuais em um único mapa somente pode ser feita com a presença de marcadores em heterozigose em ambos os genitores, que são utilizados para estabelecer relações de ligação entre os marcadores segregando individualmente em cada genitor (GARCIA et al., 2006). A construção de um mapa genético integrado, utilizando-se marcadores com diferentes tipos de segregação, apresenta inúmeras vantagens, pois permite aumentar a saturação do mapa de ligação e estender a caracterização da variação polimórfica em todo o genoma. Especificamente para poliplóides, marcadores codominantes podem ser úteis para reunir os grupos de co-segregação em seus respectivos grupos de homologia (SILVA et al., 1993).

Diversos trabalhos abordaram a construção de mapas genéticos integrados em populações  $F_1$  (RITTER; GEBHARDT; SALAMINI, 1990; RITTER; SALAMINI, 1996; MALIEPAARD; JANSEN; VAN OOIJEN, 1997; RIDOUT et al., 1998), porém um fator limitante dos modelos está na impossibilidade de caracterizar corretamente as fases de ligação entre marcadores. Mais recentemente, Wu et al. (2002b) propuseram um método baseado em máxima verossimilhança que permite estimar simultaneamente a ligação e as fases de ligação entre diferentes tipos de marcadores. Por fim, esta última abordagem foi expandida por Wu et al. (2002a) para o contexto multiponto, a fim de utilizar a informação de todos os marcadores em um mesmo grupo de ligação para estimar com maior precisão as frações de recombinação e as fases de ligação. A disponibilidade de um mapa em que as fases de ligação foram corretamente estimadas é extremamente importante para o mapeamento de QTLs, pois permite a aplicação de uma abordagem robusta para a caracterização das fases de ligação entre marcadores e QTLs. Falhas na correta caracterização das fases de ligação podem



levar a sérios vieses para a estimação de posições e efeitos de QTLs nessas populações (LIN et al., 2003).

### 2.3 Mapeamento de QTLs

O termo *Quantitative Trait Loci*, ou QTL, tem sido comumente utilizado para denominar regiões cromossômicas que contêm genes (ou locos) que controlam caracteres poligênicos, isto é, caracteres com padrão contínuo de variação fenotípica (FALCONER; MACKAY, 1996; LIU, 1998; LYNCH; WALSH, 1998). O recente advento dos marcadores moleculares possibilitou a ampliação dos conhecimentos sobre os QTLs, facilitando os estudos da arquitetura genética de caracteres importantes (FERREIRA; GRATTAPAGLIA, 1998; MACKAY, 2001a; ZENG, 2001; SIBOV et al., 2003a, 2003b). De fato, um dos fatores de maior importância para o mapeamento de QTLs é a disponibilidade de um mapa de ligação saturado e de elevada confiabilidade (LYNCH; WALSH, 1998).

A partir das informações fornecidas pelos marcadores moleculares, é possível mapear individualmente os locos que controlam o caráter quantitativo (QTLs), o que é fundamental para o melhor entendimento da herança dos caracteres poligênicos (FERREIRA; GRATTAPAGLIA, 1998). Mapear QTLs significa fazer inferências em todo o genoma sobre as relações entre o genótipo e o fenótipo de caracteres quantitativos. Essas inferências incluem informações sobre número, posição, efeitos, interações dos alelos dos QTLs dentro dos locos (dominância) e entre os locos (epistasia), efeitos pleiotrópicos dos QTLs e interações entre QTLs e ambientes (JIANG; ZENG, 1995; ZENG; KAO; BASTEN, 1999; ZENG, 2001; MALOSETTI et al., 2004, 2008).

De forma geral, o mapeamento de QTLs fundamenta-se no estabelecimento de associações entre caráter quantitativo (fenótipo) e marcadores moleculares (genótipo). Para tanto, é necessária uma população que apresente variabilidade genética e elevado desequilíbrio de ligação. Logo, populações experimentais do tipo retrocruzamentos,  $F_2$  e linhagens puras recombinantes (*recombinant inbred lines* - RILs) são utilizadas neste tipo de estudo (DOERGE, 2002). Além disso, são necessárias metodologias genético-estatísticas sofisticadas, com forte suporte computacional, devido à complexidade das análises.

#### 2.3.1 Métodos de Mapeamento

Inicialmente, as análises eram feitas para cada marca individualmente. Tal abordagem baseia-

se no teste da hipótese de diferença entre as médias fenotípicas das classes genotípicas de cada marcador, o que pode ser realizado através de testes estatísticos como o teste  $t$ , análise de variância, regressão linear simples (WELLER, 1986; LYNCH; WALSH, 1998; EDWARDS; STUBER; WENDEL, 1987; STUBER; EDWARDS; WENDEL, 1987) e teste da razão de verossimilhanças (WELLER, 1986). As principais vantagens desses métodos são a sua simplicidade e rapidez de execução, além de serem facilmente estendidos para múltiplos locos e não necessitarem de um mapa de ligação. Por outro lado, esses testes estatísticos são viesados, pois há confundimento entre o efeito do QTL e a sua distância em relação ao marcador, resultando em baixo poder de detecção quando os marcadores não saturam completamente o genoma e/ou quando as amostras são pequenas (BROMAN, 2001). Além disso, não é possível inferir sobre a localização dos QTLs.

Com o intuito de contornar algumas dessas limitações, Lander e Botstein (1989) propuseram o método de Mapeamento por Intervalo (*Interval Mapping* - IM). Esta abordagem utiliza as frequências conjuntas de um par de marcadores adjacentes para fazer inferências a respeito de um possível QTL localizado dentro do intervalo entre essas marcas. Esse método combina a análise de marcas individuais com estatísticas baseadas em verossimilhança, através de um modelo de regressão linear simples. Como o genótipo dos QTLs não é observável diretamente, técnicas apropriadas para modelos de misturas são usadas, baseadas em probabilidades condicionais, permitindo a estimação separada dos parâmetros envolvidos na análise (efeito e posição dos QTLs) e resultando no aumento do poder dos testes (LANDER; BOTSTEIN, 1989; LYNCH; WALSH, 1998). Porém, o IM também está sujeito a viés estatístico, pois utiliza a informação de apenas dois marcadores de cada vez, desconsiderando a interferência de possíveis QTLs fora do intervalo em estudo (DOERGE, 2002). Assim, podem surgir os chamados “QTLs fantasmas”, isto é, podem ocorrer falsos positivos devido à presença de QTLs em intervalos adjacentes ao intervalo sendo estudado.

Devido às limitações do IM, Zeng (1993, 1994) e Jansen e Stam (1994), de forma independente, propuseram modelos que, além de considerar os marcadores que flanqueiam o intervalo sendo mapeado, levam em consideração o efeito de marcadores fora desse intervalo, denominados cofatores. Esse método, chamado Mapeamento por Intervalo Composto (*Composite Interval Mapping* - CIM), combina o IM com um modelo de regressão múltipla, de modo a controlar a variação fora do intervalo investigado. A inclusão de marcas com efeitos significativos sobre o caráter quantitativo diminui o resíduo do modelo, resultando em expressivo aumento no poder do teste e grande diminuição dos QTLs fantasmas. No entanto, assim como o IM, o CIM também detecta efeitos de

QTLs em diferentes regiões genômicas separadamente, não sendo capaz de estudar completamente a arquitetura genética dos caracteres quantitativos. Entre outras limitações do CIM, citam-se: *i*) a análise pode ser afetada por uma distribuição irregular de marcadores no genoma; *ii*) é difícil estimar a contribuição conjunta à variância genética de múltiplos QTLs ligados; *iii*) o CIM não pode ser diretamente estendido para analisar epistasia (ZENG; KAO; BASTEN, 1999).

Para contornar tais problemas, foi proposto o método de Mapeamento de Múltiplos Intervalos (*Multiple Interval Mapping* - MIM) (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999), o qual tem como objetivo o mapeamento simultâneo de múltiplos QTLs, e não apenas um único QTL por vez, como fazem o IM e o CIM. Assim, modelos MIM são adequados para a identificação e estimação dos parâmetros da arquitetura genética dos caracteres quantitativos, incluindo o número, posições, efeitos e interações dos QTLs significativos e sua contribuição para a variância genética (ZENG; KAO; BASTEN, 1999). Para tanto, esse modelo baseado em verossimilhança faz uma busca simultânea em todo o genoma, permitindo que seja feita concomitantemente a análise de epistasia entre múltiplos QTLs.

Entre as muitas vantagens do MIM, merece destaque a possibilidade de obtenção da estimativa do valor genotípico dos indivíduos, útil para a realização da seleção assistida por marcadores. Além disso, por incluir diretamente o efeito de todos os QTLs e das interações epistáticas no modelo, o método aumenta a capacidade de identificar QTLs, com maior poder dos testes e maior precisão das estimativas das posições dos QTLs. O MIM auxilia no entendimento da arquitetura genética dos caracteres quantitativos, permitindo avaliar padrões de interação epistática e a obtenção de estimativas integrais dos efeitos dos QTLs, bem como de suas variâncias e covariâncias. As estimativas obtidas por este método são mais verossímeis, tanto biológica quanto estatisticamente, e torna-se possível compreender a importância relativa dos componentes genéticos envolvidos no controle do caráter, o que é essencial para entender sua arquitetura genética (ZENG, 2001). Biologicamente, o método MIM melhor adequa-se à visão do mapeamento de QTLs como um problema de seleção de modelos ou seleção de variáveis, isto é, dos locos e das interações epistáticas mais fortemente evidenciados pelos dados (BROMAN; SEN, 2009).

Sem exceção, os métodos acima descritos foram delineados para mapeamento de QTLs no caso univariado, isto é, para apenas um caráter avaliado em um único ambiente. Jiang e Zeng (1995) propuseram um método de mapeamento de QTLs para dados de múltiplos caracteres, baseado no método da máxima verossimilhança, o qual foi posteriormente denominado mCIM (*multiple-trait*

CIM). O termo *caráter* é utilizado neste contexto em sentido mais amplo, visto que o mesmo caráter avaliado em diferentes condições ambientais pode ser tratado como diferentes caracteres, de modo que o modelo adequa-se também a esta situação. Assim, tal metodologia permite não apenas o estudo de QTLs com efeitos pleiotrópicos, mas também da interação genótipo por ambiente ( $G \times E$ ) e do entendimento das causas da correlação genética entre caracteres, a saber, pleiotropia e ligação genética (JIANG; ZENG, 1995). Em alguns trabalhos anteriores havia sido feito mapeamento de QTLs para diversos caracteres separadamente, como em Paterson et al., 1988, 1991, e Stuber et al., 1992 apud Jiang e Zeng (1995), mas esta abordagem não aproveita a estrutura de correlação dos dados, além de dificultar a interpretação dos resultados e trazer desvantagens como redução de poder e aumento das variâncias das estimativas dos parâmetros. Finalmente, não permite a realização de testes formais de hipóteses biologicamente interessantes. (JIANG; ZENG, 1995).

Posteriormente, Knott e Haley (2000) propuseram uma aproximação por mínimos quadrados para mapeamento de QTLs no caso de múltiplos caracteres, a fim de contornar as principais limitações do método baseado na máxima verossimilhança, já que este é computacionalmente intensivo, requer programas especializados, pode carecer de robustez, os procedimentos de teste são lentos e a expansão para incluir múltiplos QTLs é problemática.

Jiang e Zeng (1995) e Knott e Haley (2000), através de simulações computacionais, mostraram que a análise integrada geralmente é mais informativa do que análises separadas para cada caráter. As vantagens incluem: *i*) aumento do poder de detecção de QTLs para caracteres moderada ou altamente correlacionados; *ii*) melhoria da precisão na estimação de parâmetros; *iii*) possibilidade de realização de procedimentos adequados para testes de hipóteses envolvendo múltiplos caracteres. Dentro deste último ponto, é de grande importância testar se associações significativas com diferentes caracteres são devidas a QTLs pleiotrópicos ou múltiplos QTLs não pleiotrópicos ligados fisicamente. Além de auxiliar na compreensão das causas de correlações genéticas, essas informações podem ser empregadas em programas de melhoramento genético, os quais geralmente têm interesse em quebrar ligações desfavoráveis de genes (JIANG; ZENG, 1995).

Tais constatações motivaram também a expansão do método MIM para o contexto de múltiplos caracteres (SILVA, 2010). Esta abordagem, denominada MTMIM (*multiple traits MIM*), faz uso das mesmas estratégias de busca originalmente propostas para o caso univariado, em um processo de seleção de variáveis que visa obter um modelo parcimonioso e de bom ajuste aos dados. Entretanto, a inclusão de vários caracteres em um único modelo melhora as estimativas e o poder estatístico, em

comparação ao método MIM aplicado a cada carácter individualmente, além de fornecer uma base estatística para o teste de hipóteses de interesse.

Outras abordagens multivariadas para detecção de QTLs foram propostas, mas encontraram menor aceitação no meio científico. Weller et al. (1996) e Mangin et al. (1998) apud Caliński et al. (2000) sugeriram a aplicação de transformações canônicas dos dados originais, seguidas de análises univariadas dos caracteres independentes resultantes. A maior vantagem desse método seria a possibilidade de se usar programas já disponíveis para o caso de um único carácter, mas as desvantagens incluem a eventualidade de não se encontrar uma transformação canônica capaz de garantir que todos os QTLs influenciem apenas um carácter canônico. Korol et al. (1998) propuseram uma expansão do método multivariado da máxima verossimilhança para múltiplos intervalos, mas seu modelo considerava apenas o caso bivariado e dois QTLs ligados, sendo assim demasiadamente restritivo. Caliński et al. (2000), à semelhança de Knott e Haley (2000), defenderam um método baseado em regressão múltipla, porém seus procedimentos de teste eram excessivamente complicados em comparação aos advogados por Knott e Haley (2000).

### 2.3.2 QTLs em Cana-de-açúcar

Os métodos de mapeamento de QTLs apresentados (IM, CIM, mCIM, MIM e MTMIM) foram desenvolvidos apenas para populações experimentais oriundas de linhagens endogâmicas, como retrocruzamentos, populações  $F_2$  e as provenientes de RILs, em que há apenas dois alelos segregantes e a fase de ligação entre marcadores e QTLs é conhecida *a priori*. Essas metodologias estão bem estabelecidas e foram implementadas em vários programas, tais como o MAPMAKER/EXP (LANDER et al., 1987), QTL Cartographer (BASTEN; WEIR; ZENG, 2005; WANG; BASTEN; ZENG, 2007), MapQTL (VAN OOIJEN, 2009) e R/qtl (BROMAN et al., 2003). Mesmo assim, ainda há muitas limitações ao seu emprego nos programas de melhoramento, pois, na prática, outros fatores devem ser considerados, como a interação entre QTLs e ambientes e a correlação entre caracteres.

Em geral, os trabalhos de mapeamento genético em cana-de-açúcar resultaram na geração de dois mapas, um para cada genitor, com base em marcadores segregando na proporção de 1:1. Portanto, diversos trabalhos de mapeamento de QTLs foram desenvolvidos para essa espécie, geralmente através de análises de marcas individuais ou IM, e por intermédio das metodologias desenvolvidas para populações oriundas de retrocruzamento (SILLS et al., 1995; DAUGROIS et al., 1996; GUIMARÃES; SILLS; SOBRAL, 1997; ASNAGHI et al., 2001; MING et al., 2001, 2002a,

2002c; HOARAU et al., 2002; JORDAN et al., 2004; SILVA; BRESSIANI, 2005; McINTYRE et al., 2005a, 2005b, 2006; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2006; RABOIN et al., 2006, 2008; WEI et al., 2006; AL-JANABI et al., 2007; AITKEN et al., 2008; PIPERIDIS et al., 2008; PINTO et al., 2010). Algumas exceções existem, visto que alguns trabalhos incluíram marcadores com segregação 3:1 e até em dosagens maiores (HOARAU et al., 2002; McINTYRE et al., 2005a, 2005b; REFFAY et al., 2005; AITKEN; JACKSON; McINTYRE, 2006; RABOIN et al., 2006; AL-JANABI et al., 2007; AITKEN et al., 2008; PIPERIDIS et al., 2008). Em alguns trabalhos foram também feitos estudos da interação entre QTLs (epistasia), mas através de uma abordagem digênica, claramente inferior aos modelos MIM (HOARAU et al., 2002; AITKEN; JACKSON; McINTYRE, 2006). É importante ressaltar que estes trabalhos utilizaram as metodologias convencionais que consideram mapeamento para os dois genitores separadamente, o que notadamente traz muitas desvantagens: *i*) redução no tamanho amostral; *ii*) menor saturação do mapa e menor cobertura do genoma, devido ao descarte de marcadores com segregação 3:1 e outras; *iii*) diminuição no poder dos testes; *iv*) dificuldade de interpretação dos resultados. Novamente, os resultados desses trabalhos estão mais bem detalhados no capítulo de Pastina et al. (2010), o qual serviu de base para a breve descrição realizada a seguir.

Os primeiros trabalhos de mapeamento de QTLs surgiram concomitantemente à publicação dos primeiros mapas de ligação. Sills et al. (1995) utilizaram análise de marcas individuais para identificar, na mesma progênie utilizada por Mudge et al. (1996) e descrita na seção 2.2, vinte e quatro associações entre marcadores moleculares e os caracteres quantitativos número de colmos, porcentagem de colmos ramificados, resistência ao carvão (*Ustilago scitaminea*), diâmetro de colmos, teor de açúcar (Pol), teor de fibra e produtividade. Esses autores empregaram também análises de regressão múltipla e identificaram a presença de um termo epistático significativo para o caráter diâmetro de colmos. Guimarães, Sills e Sobral (1997) encontraram, para essa mesma progênie, associações entre marcadores moleculares e tempo de florescimento.

Em seguida, Daugrois et al. (1996) serviram-se da progênie utilizada por Grivet et al. (1996) para, também através de uma análise de marcas individuais, encontrar um marcador com grande efeito sobre a resistência à ferrugem (*Puccinia melanocephala*), o qual exibia segregação do tipo 3:1 (resistentes:susceptíveis) na progênie. A mesma progênie foi empregada por Asnaghi et al. (2001) para avaliação da resistência do cultivar R570 a diversos isolados do patógeno, o que forneceu evidência de um único gene com grande efeito sobre a resistência. Posteriormente, Hoarau et al.

(2002) utilizaram seu mapa construído para a mesma população (HOARAU et al., 2001), somado a marcadores presentes em múltiplas doses, e identificaram quarenta possíveis QTAs (*Quantitative Trait Alleles*) para os caracteres comprimento de colmos, número de colmos, diâmetro de colmos e teor de sólidos solúveis (brix), dos quais cinco foram consistentes nos dois anos consecutivos de avaliação. Adicionalmente, os autores encontraram quarenta e uma interações epistáticas digênicas.

Ming et al. (2002b) publicaram dois mapas genéticos, um para cada uma de duas progênes oriundas de cruzamentos interespecíficos (ver seção 2.2). No ano anterior, Ming et al. (2001) haviam aplicado análise de marcas individuais e identificado QTLs para teor de açúcar nessas duas populações, tendo sido detectadas vinte e duas associações para a população PIN 84-1 × Muntok Java e quatorze associações para Green German × IND 81-146. Em seguida, Ming et al. (2002a) buscaram, nessas mesmas progênes e em dados de três anos consecutivos, QTLs para os caracteres produtividade de açúcar, Pol, teor de fibras, número de colmos, peso de colmos e conteúdo de cinzas. No total, foram verificadas cento e duas associações significativas, sendo sessenta e uma em regiões do mapa de ligação, detectadas pelo método IM, e quarenta e uma em marcadores não alocados em grupo algum, detectadas por análise de marcas individuais. Esses autores atentaram ainda para a importante possibilidade de que, caso QTLs sejam mapeados em regiões correspondentes de diferentes cromossomos homólogos, eles representem na realidade diferentes alelos de um mesmo loco. Por fim, Ming et al. (2002c) utilizaram essa mesma estratégia e encontraram sessenta e cinco novas associações nessas populações, desta vez para os caracteres altura de planta e florescimento.

Jordan et al. (2004) utilizaram dados de uma progênie oriunda do cruzamento Q117 × 74C42, avaliada em dois locais e anos, à qual aplicaram análise de marcas individuais em marcadores do tipo RFLP e RAF (*Radio Labelled Amplified Fragment*). Foram encontradas sete associações para o caráter número de colmos e seis para o caráter número de brotos. É interessante ressaltar que três desses locos marcadores afetaram ambos os caracteres, o que pode ser indicativo da ocorrência de QTLs pleiotrópicos. Esses dados foram ainda utilizados em dois trabalhos subsequentes, ambos com uso de análise de marcas individuais (McINTYRE et al., 2005a, 2005b). No primeiro, os autores encontraram sete regiões envolvidas no controle da resistência ao patógeno de raiz *Pachymetra chaunorhiza*, tanto em marcadores não ligados quanto nos mapas de ligação de ambos os genitores, além de quatro QTLs controlando a resistência à ferrugem (McINTYRE et al., 2005a). Ressalta-se que alguns desses marcadores mantiveram-se significativamente associados a esses caracteres de resistência em um trabalho posterior de mapeamento por associação, no qual foi empregado um

painel associativo de 154 clones de elite (WEI et al., 2006). Já no segundo trabalho, McIntyre et al. (2005b) encontraram QTLs para resistência à ferrugem através do uso de sequências similares a genes conhecidos envolvidos no controle desse caráter. De fato, existe uma região em um cromossomo de sorgo, na qual já foi mapeado um QTL para resistência à ferrugem, que está associada com alguns dos marcadores identificados no trabalho de McIntyre et al. (2005b).

Uma progênie oriunda dos clones SP80-180 e SP80-4966 já foi utilizada por Silva e Bressiani (2005) para busca de QTLs para os caracteres Pol, produtividade de cana e teor de fibras, através de análise de marcas individuais. Foi encontrado apenas um QTL para Pol, sendo que o alelo favorável estava presente no genitor com baixo teor de sacarose, isto é, SP80-180. Como os marcadores usados nesse estudo eram derivados de sequências expressas (ESTs), o QTL detectado pode ser utilizado para determinação de genes candidatos. Vale também notar que este foi o primeiro trabalho de mapeamento de QTLs em uma progênie oriunda de genótipos brasileiros.

No mesmo ano, Reffay et al. (2005) aplicaram métodos de marcas individuais à população Q117 × MQ77-340, avaliada em três anos agrícolas. Os caracteres estudados foram a produção de açúcar comercial, brix, Pol, teor de fibra, peso de colmos, produtividade de cana e produtividade de açúcar. Piperidis et al. (2008) compararam os QTLs encontrados nos mapas desses dois genitores com os mapeados nos cultivares R570 e Q165A, verificando associações para o caráter brix em grupos de ligação consistentes em dois ou até três desses mapas. Foram também detectados QTLs para Pol e produção de açúcar comercial, em marcadores não ligados (marcas individuais) ou através do IM no mapa genético.

Estudos de epistasia e da correlação entre caracteres também foram feitos em alguns casos, mas com abordagens estatisticamente inferiores. Um exemplo é o trabalho de Aitken, Jackson e McIntyre (2006), em que a progênie do cruzamento IJ76-514 × Q165A foi avaliada em duas épocas distintas e foram detectadas trinta e sete associações para brix e Pol, através de análise de marcas individuais em cada corte separadamente. A abordagem digênica detectou 97 efeitos epistáticos, das quais apenas nove foram consistentes nos dois anos. Além disso, uma simples meta-análise mostrou que oito dessas interações afetaram mais de um caráter simultaneamente. A mesma progênie foi avaliada por McIntyre et al. (2006) através de marcadores do tipo SNP em gene da sacarose-fosfato sintase, em busca de uma associação entre a frequência alélica e a produção de açúcar comercial. Os resultados mostraram não haver efeito algum do SNP, apesar de dois haplótipos terem sido mapeados em regiões próximas a QTLs previamente associados ao teor de sacarose. Ainda mais



recentemente, Aitken et al. (2008) mapearam QTLs nessa mesma progênie para os caracteres peso de colmos, altura de colmos, número de colmos, diâmetro de colmos e produtividade de cana-de-açúcar, em dados de três cortes consecutivos. É interessante notar que este foi um dos poucos trabalhos que fez uso do método CIM. Neste caso, praticamente metade das associações encontradas foram consistentes nos três anos de avaliação, o que pode tornar os resultados interessantes para a seleção assistida por marcadores. Novamente, neste caso as análises foram feitas separadamente para cada corte e os resultados comparados em uma meta-análise. Um grande número de interações epistáticas (195) foi também observado como resultado da aplicação de análises digênicas.

Raboin et al. (2006) localizaram, através da análise de marcas individuais, uma região associada à resistência à ferrugem, com segregação do tipo 1:1 (resistentes:susceptíveis) na progênie, e uma região controlando o caráter cor de colmos, com segregação do tipo 1:1 (vermelho:não-vermelho) na progênie. O cruzamento utilizado para tanto envolveu os genótipos R570 e MQ76-53 e a progênie foi avaliada em dois e três anos agrícolas para cada um dos dois caracteres, respectivamente.

Além do já mencionado estudo com *Pachymetra chaunorhiza* e ferrugem, Wei et al. (2006) utilizaram o mesmo painel de mapeamento associativo para detectar marcadores associados com a resistência à escaldadura das folhas e à doença de Fiji, também por meio de análise de marcas individuais e marcadores AFLP e SSR. No total para as quatro doenças, foram encontradas trinta e nove associações significativas.

Al-Janabi et al. (2007) aplicaram o método CIM a dados de uma progênie obtida pelo cruzamento entre os genótipos M134/75 e R570, sendo que o primeiro é resistente à mancha amarela causada por *Mycovellosiella koepkey* e o último é susceptível, de modo que a progênie segregou na proporção 3:1 (resistentes:susceptíveis). Foi utilizado um mapa obtido por duplo *pseudo-testcross* e aplicou-se também o método de marcas individuais aos marcadores não ligados e marcadores em doses múltiplas. As avaliações foram feitas em dois cortes e dois locais. Os resultados indicaram evidência de apenas dois QTLs, mas um deles teve forte efeito, já que explicou 23,8% da variação fenotípica.

Em resumo, nota-se que os trabalhos de mapeamento de QTLs quase sempre envolveram cruzamentos biparentais, interespecíficos ou não. Em alguns casos, foram utilizadas progênies obtidas por autofecundação do cultivar R570. Os caracteres analisados foram compostos majoritariamente por componentes de produção (número, comprimento, peso e diâmetro de colmos), teor de açúcar (Pol), teor de fibras, florescimento e resistência a doenças (ferrugem, carvão, mancha amarela e pa-

tógenos de raiz). Com raras exceções, os trabalhos utilizaram métodos de marcas individuais e/ou IM, em mapas obtidos separadamente para cada genitor e em marcadores com segregação dos tipos 1:1 e 3:1. Quando interações epistáticas foram estudadas, isto foi feito através de uma abordagem digênica, a qual não inclui a influência de outros QTLs, epistáticos ou não. Quando análises foram feitas para vários cortes e/ou em vários locais, os métodos estatísticos foram aplicados isoladamente a cada combinação corte-local, com uma subsequente inspeção dos resultados em busca de QTLs com efeitos consistentes ou interatuantes com o ambiente. Claramente, esses modelos estatísticos não são apropriados para essas conclusões (PASTINA, 2010).

Há na literatura alguns poucos métodos para mapeamento de QTLs em progênes oriundas do cruzamento de genitores não endogâmicos (SONG; SOLLER; GENIZI, 1999; JOHNSON; JANSEN; ARENDONK, 1999). Mais recentemente, Lin et al. (2003) expandiram os conceitos apresentados por Wu et al. (2002b) para o mapeamento de QTLs em populações  $F_1$ , mais especificamente para espécies diplóides. O objetivo do método é integrar todas as possíveis fases de ligação entre marcadores e QTL, considerando um intervalo formado por dois marcadores adjacentes. Para tanto, este método supõe a existência de um mapa de ligação já construído e com as fases de ligação estimadas. As possíveis fases de ligação entre marcadores e QTL estão associadas a probabilidades de ocorrência, resultando em um procedimento baseado em um modelo de mistura de distribuições. Essa abordagem é bastante poderosa, porém considera apenas o intervalo formado por dois marcadores e, portanto, está sujeita às mesmas limitações observadas para o modelo IM proposto por Lander e Botstein (1989). A principal limitação é que o modelo não controla a variação causada por QTLs fora do intervalo em estudo e, neste caso, a existência destas regiões adjacentes tende a influenciar as análises, levando à detecção de falsos positivos. Para situações em que há QTLs ligados em um cromossomo, os resultados tornam-se viesados, pois os respectivos efeitos não podem ser isolados (ZENG, 1994). Isto resulta em redução do poder do teste estatístico e, além disso, a epistasia não pode ser diretamente incluída no modelo.

Essa principal limitação foi abordada por Gazaffi (2009), o qual expandiu o modelo proposto por Lin et al. (2003) para o contexto do CIM, incluindo cofatores para controlar a variação causada por QTLs localizados fora do intervalo de interesse. De maneira similar à proposta original do método CIM (ZENG, 1993, 1994; JANSEN; STAM, 1994), o método de Gazaffi (2009) baseia-se em um modelo de misturas em que as probabilidades condicionais dos genótipos dos QTLs são variáveis latentes e faz uso do método da máxima verossimilhança para ajuste do modelo. Tal metodologia foi

aplicada com sucesso em dados de uma progênie segregante de cana-de-açúcar e serviu como base para posteriores expansões do modelo, inclusive para o contexto dos modelos mistos, conforme será detalhado posteriormente.

Há ainda outros pontos que devem ser considerados no mapeamento de QTLs para que os resultados sejam úteis ao melhoramento genético. A cana-de-açúcar é uma espécie perene, em que vários cortes são conduzidos, de modo que os indivíduos ficam sujeitos a diferentes condições ambientais ao longo dos anos, e é importante identificar QTLs que se expressem de forma consistente ao longo desses cortes ou que se expressem sob condições específicas de um determinado ambiente. Isto permitiria a obtenção de melhores estimativas dos valores genéticos dos indivíduos, fornecendo uma importante informação que poderia ser canalizada para programas de seleção assistida, de maneira a auxiliar os programas de melhoramento. Conforme será discutido na seção 2.4.2, tais questões foram abordadas por Pastina (2010), mas apenas para a situação de um único caráter quantitativo. De fato, outro ponto essencial é o estudo da correlação entre caracteres, também de importância para o melhoramento genético, visto que geralmente é aplicada seleção para vários caracteres simultaneamente. Claramente, o mapeamento deveria levar isso em consideração, pois informações a respeito da base genética dos caracteres correlacionados (pleiotropia e ligação entre QTLs) podem auxiliar o melhorista no processo de seleção.

## 2.4 Modelos Mistos

Qualquer modelo linear que misture efeitos fixos e aleatórios é chamado um modelo linear misto (McCULLOCH; SEARLE, 2001). Efeitos fixos são parâmetros associados com toda uma população ou com níveis específicos de fatores experimentais, enquanto que efeitos aleatórios estão associados com unidades experimentais individuais amostradas de uma população qualquer, as quais variam entre experimentos (PINHEIRO; BATES, 2000). Devido à incorporação de efeitos aleatórios, os modelos mistos são também caracterizados por suposições extras sobre a distribuição desses efeitos (EEUWIJK et al., 2001). Nos modelos lineares, efeitos fixos são utilizados para modelar a média do vetor de observações  $\mathbf{y}$ , ao passo que efeitos aleatórios governam a estrutura de variância-covariância de  $\mathbf{y}$  (MCCULLOCH; SEARLE, 2001). É interessante ressaltar que, em virtualmente todos os modelos lineares, a média geral (intercepto) é um efeito fixo e o erro é um efeito aleatório.

Segundo Balzarini (2001), a estrutura de erros em experimentos de melhoramento genético é

muito mais complexa que a usualmente considerada em modelos lineares simples, de modo que os modelos lineares tradicionais, os quais consideram todos os efeitos como sendo fixos, são inadequados para análise de dados dessa natureza, já que a suposição de independência dos erros é muito restritiva. Os modelos lineares mistos, por outro lado, levam em consideração o fato de que as observações são correlacionadas, lidando com isto através da inclusão de efeitos aleatórios, que possuem componentes de variância associados (PIEPHO, 2000). Para um vetor  $\mathbf{y}$  com  $N$  observações, haveria  $N(N + 1)/2$  elementos distintos em  $\text{var}(\mathbf{y})$ . Especificar cada um destes elementos seria um trabalho extremamente complexo e, de fato, a simplificação de tal tarefa é uma das principais razões para a inclusão de fatores aleatórios no modelo, já que variâncias e covariâncias atribuíveis a esses fatores permitem uma estruturação adequada da  $\text{var}(\mathbf{y})$  (MCCULLOCH; SEARLE, 2001).

Como os dois tipos de efeitos são distintos e, portanto, tratados distintamente durante a análise de dados, a decisão de considerar cada fator como fixo ou aleatório é crucial (MCCULLOCH; SEARLE, 2001). Classicamente, o argumento utilizado para auxiliar a tomada dessa decisão baseia-se no fato de os níveis do fator serem amostrados de uma população maior (efeito aleatório) ou serem todos os de interesse do pesquisador (efeito fixo) (SEARLE, 1971). No entanto, Smith, Cullis e Thompson (2005) argumentam que a escolha depende do objetivo da análise e de considerações a respeito das propriedades dos dois tipos de procedimentos de estimação, isto é, predição linear não viesada para efeitos aleatórios e estimação linear não viesada para efeitos fixos. Como exemplo, tem-se a situação de seleção de variedades. Neste caso, o objetivo é ranquear os efeitos estimados de variedades o mais próximo possível aos *rankings* verdadeiros dos efeitos de variedade. Em outras palavras, as estimativas dos efeitos de variedade devem predizer os efeitos verdadeiros da melhor maneira possível, o que implica o uso de *predição*, de tal forma que o efeito de variedade deve ser considerado aleatório. Por outro lado, se o objetivo da análise é determinar a diferença entre pares específicos de variedades, o método da predição é inadequado, pois a predição de uma diferença específica é viesada. Portanto, este efeito de variedade deve ser considerado como fixo. É relevante notar que, no método mais usual da análise de variância para modelos fixos, a decisão de utilizar um fator como fixo ou aleatório influi na realização dos testes de hipóteses, bem como nos resultados dos mesmos, fato que não ocorre no contexto dos modelos mistos.

O uso de efeitos aleatórios de variedades, ou de genótipos, tem ainda como vantagem o fato de permitir análises válidas de dados combinados ao longo de vários anos (SMITH; CULLIS; THOMPSON, 2005). Thompson, 1959 apud Smith, Cullis e Thompson (2005) discute a estimação

de efeitos fixos nesse caso, e esses autores afirmam que uma abordagem de verossimilhança com efeitos genotípicos aleatórios reduz o viés na estimação de efeitos fixos de anos.

A forma usual para um modelo linear misto geral é (SMITH; CULLIS; THOMPSON, 2005):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

em que  $\mathbf{y}$  é o vetor de dados fenotípicos;  $\boldsymbol{\tau}$  é o vetor de efeitos fixos com matriz de delineamento associada  $\mathbf{X}$  (de posto completo);  $\mathbf{u}$  é o vetor de efeitos aleatórios com matriz de delineamento associada  $\mathbf{Z}$ ; e  $\mathbf{e}$  é o vetor aleatório de erros. Assume-se que a distribuição conjunta de  $\mathbf{u}$  e  $\mathbf{e}$  é normal com média zero e matriz de variâncias-covariâncias (VCOV)

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}.$$

A distribuição de  $\mathbf{y}$  é, então, normal com vetor de médias  $\mathbf{X}\boldsymbol{\tau}$  e matriz VCOV  $\text{var}(\mathbf{y}) = \mathbf{H} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ .

Através da especificação das estruturas de  $\mathbf{Z}$ ,  $\mathbf{G}$  e  $\mathbf{R}$ , diferentes modelos para a matriz VCOV dos dados ( $\mathbf{H}$ ) podem ser considerados na abordagem de modelos mistos. É interessante notar que, quando  $\mathbf{Z} = \mathbf{0}$  e  $\mathbf{R} = \sigma^2\mathbf{I}$ , o modelo misto acima reduz-se ao modelo linear tradicional. No caso dos experimentos de melhoramento genético, correlações genéticas podem ser incluídas no modelo misto através da matriz  $\mathbf{G}$ , enquanto que correlações experimentais entre observações podem ser ajustadas pelos elementos fora da diagonal da matriz  $\mathbf{R}$  (BALZARINI, 2001). Vários modelos possíveis para as estruturas de variância-covariância de  $\mathbf{G}$  e  $\mathbf{R}$  são encontrados na Tabela 1 (adaptada de Pastina, 2010, onde detalhes de cada modelo podem ser encontrados).

Os efeitos fixos e aleatórios na equação 1 são estimados utilizando melhor estimação linear não viesada e melhor predição linear não viesada, respectivamente (SMITH; CULLIS; THOMPSON, 2005). Isto conduz às estimativas (HENDERSON, 1990):

$$\begin{aligned} \hat{\boldsymbol{\tau}} &= (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} &= \mathbf{G}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\tau}}). \end{aligned}$$

A abordagem mais comumente empregada para a estimação dos parâmetros de variância e co-

Tabela 1 – Alguns modelos para a estrutura de variância-covariância da matriz  $G$ 

Modelo	Descrição	Matriz $G$
ID	Varição homogênea	$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & 0 & \dots & 0 \\ 0 & \sigma_g^2 + \sigma_{ge}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
DIAG	Varição heterogênea	$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
CS	Simetria composta com variação homogênea	$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 & \dots & \sigma_g^2 \\ \sigma_g^2 & \sigma_g^2 + \sigma_{ge}^2 & \dots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \dots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
CS <sub>Het</sub>	Simetria composta com variação heterogênea	$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 & \dots & \sigma_g^2 \\ \sigma_g^2 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \dots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \dots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
AR1	Modelo autoregressivo de primeira-ordem com variação homogênea	$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 \rho_g & \dots & \sigma_g^2 \rho_g^{d(1,J)} \\ \sigma_g^2 \rho_g & \sigma_g^2 + \sigma_{ge}^2 & \dots & \sigma_g^2 \rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(J,1)} & \sigma_g^2 \rho_g^{d(J,2)} & \dots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
AR1 <sub>Het</sub>	Modelo autoregressivo de primeira-ordem com variação heterogênea	$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 \rho_g & \dots & \sigma_g^2 \rho_g^{d(1,J)} \\ \sigma_g^2 \rho_g & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \dots & \sigma_g^2 \rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(J,1)} & \sigma_g^2 \rho_g^{d(J,2)} & \dots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
FA1	Modelo fator analítico de primeira-ordem	$\begin{bmatrix} \lambda_1^2 + \Psi_1 & \lambda_1 \lambda_2 & \dots & \lambda_1 \lambda_J \\ \lambda_2 \lambda_1 & \lambda_2^2 + \Psi_2 & \dots & \lambda_2 \lambda_J \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_J \lambda_1 & \lambda_J \lambda_2 & \dots & \lambda_J^2 + \Psi_J \end{bmatrix}$
US	Modelo não-estruturado	$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_{g_{12}} & \dots & \sigma_{g_{1J}} \\ \sigma_{g_{21}} & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \dots & \sigma_{g_{2J}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g_{J1}} & \sigma_{g_{J2}} & \dots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$

$\sigma_g^2$  e  $\sigma_{ge}^2$ : componentes de variância para o efeito principal de genótipos e para interação genótipo  $\times$  ambiente, respectivamente;  $\rho_g^{d(j,j')}$ : correlação genética entre ambientes, em que  $d(j, j')$  corresponde à distância em tempo entre eles;  $\sigma_{g_j}^2$  e  $\sigma_{ge_j}^2$ : variância genética específica de cada ambiente para o efeito principal de genótipos e para interação genótipo  $\times$  ambiente;  $\sigma_{g_{jj'}}$ : covariância genética entre os ambientes  $j$  e  $j'$ ;  $\Psi_j$ : variância residual específica de cada ambiente;  $\lambda_j$  e  $\lambda_{j'}$ : elementos (*loadings*) do fator nos ambientes  $j$  e  $j'$ . Os acrônimos referentes aos vários modelos foram mantidos na forma derivada do inglês e de acordo com os padrões utilizados pelo software Genstat.

variância de um modelo misto é baseada no método da máxima verossimilhança residual (*Residual Maximum Likelihood* - REML), uma variante que elimina o viés do método da máxima verossimi-

lhança (*Maximum Likelihood* - ML) nas estimativas dos componentes de variância (PATTERSON; THOMPSON, 1971). O método REML envolve a maximização do logaritmo da verossimilhança, o qual pode ser escrito como:

$$l_R = -\frac{1}{2} \{ \log|\mathbf{H}| + \log|\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y} \},$$

em que  $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}$ . A maximização de tal verossimilhança com respeito ao vetor de parâmetros de variância usualmente requer um esquema iterativo, tal como o algoritmo *Fisher Scoring* (PATTERSON; THOMPSON, 1971), o esquema *Expectation-Maximization* (DEMPSTER; LAIRD; RUBIN, 1977) e o algoritmo *Average Information* (GILMOUR; THOMPSON; CULLIS, 1995).

Se os parâmetros em  $\mathbf{H}$  são conhecidos,  $\hat{\tau}$  é o melhor estimador linear não viesado (*Best Linear Unbiased Estimator* - BLUE) de  $\tau$  e  $\tilde{u}$  é o melhor preditor linear não viesado (*Best Linear Unbiased Predictor* - BLUP) de  $u$ . Geralmente, no entanto, os parâmetros em  $\mathbf{H}$  são desconhecidos e, na prática, são substituídos pelas suas estimativas obtidas por REML. As estimativas resultantes dos efeitos fixos e aleatórios são então denominadas BLUEs empíricos (E-BLUEs) e BLUPs empíricos (E-BLUPs) (SMITH; CULLIS; THOMPSON, 2005; PAYNE et al., 2009).

Quando um modelo misto é utilizado para predição de efeitos aleatórios, como por exemplo as médias de genótipos, as estimativas BLUPs dos efeitos aleatórios são menores do que se a estimação fosse feita considerando os efeitos como fixos, devido à ponderação das médias por pesos que são funções de razões de componentes de variância. Por este motivo, estimativas BLUP de efeitos aleatórios são comumente chamadas de estimativas “encolhidas” de parâmetros (BALZARINI, 2001). É também possível observar, pelos estimadores acima, que as estimativas dos efeitos fixos incluem uma correção para a existência de variâncias heterogêneas e de covariâncias, já que o inverso da matriz  $\mathbf{H}$  aparece na expressão do estimador  $\hat{\tau}$ . Assim, mesmo a parte fixa do modelo é afetada pela inclusão de termos aleatórios, fato que torna as estimativas mais realistas do que as obtidas pelos mínimos quadrados tradicionalmente aplicados em modelos fixos (PINHEIRO; BATES, 2000).

Conforme foi mencionado anteriormente, diferentes estruturas de VCOV podem ser ajustadas para as matrizes  $\mathbf{G}$  e/ou  $\mathbf{R}$ . É necessário, então, dispor de algum critério que possibilite a escolha do modelo mais adequado, levando em consideração a qualidade do ajuste e a parsimônia de cada modelo. Para tanto, geralmente são utilizados métodos baseados na verossimilhança (WOLFINGER, 1993). Dois modelos são aninhados se ambos contêm os mesmos termos e um deles possui

ao menos um termo adicional. Nesta situação, o teste entre eles pode ser feito através do teste da razão de verossimilhanças (*Likelihood Ratio* - LR), como segue:

$$LR = -2 \times \log \left( \frac{L_R}{L_C} \right),$$

em que  $L_R$  e  $L_C$  são as verossimilhanças residuais dos modelos reduzido e completo, respectivamente. Tal teste avalia a hipótese de nulidade ( $H_0$ ) de que os modelos não diferem entre si e tem distribuição  $\chi^2$  com número de graus de liberdade igual à diferença no número de parâmetros entre os dois modelos (EEUWIJK et al., 2001). Em outras palavras, a hipótese  $H_0$  é de que todos os parâmetros presentes no modelo completo, mas não no reduzido, são iguais a zero. É essencial notar que esse teste considera apenas o número de parâmetros nas estruturas de variância-covariância, de tal modo que a parte fixa de ambos os modelos deve ser a mesma.

Nos casos em que os modelos examinados não são aninhados entre si, é comum empregar os critérios de seleção de modelos denominados Critério de Informação de Akaike (*Akaike Information Criterion* - AIC - AKAIKE, 1974) e Critério de Informação Bayesiano (*Bayesian Information Criterion* - BIC - SCHWARZ, 1978). Os valores destes testes são (PIEPHO, 2000):

$$\begin{aligned} \text{AIC} &= -2\log(L) + 2 \times n_{\text{PAR}} \\ \text{BIC} &= -2\log(L) + \log(N) \times n_{\text{PAR}}, \end{aligned}$$

em que  $\log(L)$  é o logaritmo da verossimilhança residual,  $n_{\text{PAR}}$  é o número total de parâmetros de variância e  $N$  é o número total de observações. É possível notar que o BIC, o qual é baseado no conceito de fatores de Bayes, penaliza mais fortemente o número de parâmetros no modelo, tendendo a selecionar modelos mais parsimoniosos. Já o AIC, que é baseado na teoria da informação de Kullback-Leibler, tende a selecionar modelos que se ajustam melhor aos dados e, portanto, permite a inclusão de um número maior de parâmetros (BURNHAM; ANDERSON, 2004). Como o logaritmo da verossimilhança apresenta um coeficiente negativo nas expressões acima, e a inclusão de parâmetros eleva os valores de AIC e BIC, nota-se que valores mais baixos desses critérios correspondem a modelos de preferência.

Visto que os parâmetros referentes à parte fixa dos modelos não estão envolvidos no cálculo desses critérios, sua significância não pode ser por eles avaliada. Assim, geralmente utiliza-se para tanto o teste de Wald (VERBEKE; MOLENBERGHS, 2000).



### 2.4.1 Análise de dados de múltiplos caracteres e múltiplos ambientes através de Modelos Mistos

Em programas de melhoramento genético, é comum a avaliação de diversos materiais genéticos em diferentes ambientes, nos chamados METs (*Multi-Environment Trials*). Com o intuito de analisar este tipo de dados, os modelos mistos exibem vantagens que os tornam superiores aos modelos lineares tradicionais. Entre elas, destacam-se: facilidade de lidar com dados incompletos, habilidade de se utilizar modelos mais realistas para a variação residual dentro de cada experimento e a possibilidade de considerar alguns efeitos como aleatórios ao invés de fixos (SMITH; CULLIS; THOMPSON, 2005). Por esses motivos, o uso de modelos mistos para análises de dados de METs tem se popularizado.

Ensaio do tipo MET permitem o estudo da interação  $G \times E$ , a qual complica a análise de dados univariados (PIEPHO, 2000; EEUWIJK et al., 2001; SMITH; CULLIS; THOMPSON, 2001; VERBYLA et al., 2003; BOER et al., 2007; WELHAM et al., 2010). Neste caso de apenas um caráter, há abundantes modelos para análise da interação  $G \times E$ , como o modelo de regressão de Finlay e Wilkinson (1963), o método AMMI (*Additive Main effects and Multiplicative Interaction effects*) (GABRIEL, 1978; DUARTE; VENCOSKY, 1999) e os modelos mistos (PIEPHO, 1997; CULLIS et al., 1998; SMITH; CULLIS; THOMPSON, 2001, 2005; SMITH et al., 2007). Considerando esta última classe de modelos, a modelagem de variâncias e covariâncias genéticas entre ambientes, aliada à modelagem de resíduos de maneira adequada para cada ambiente, é uma forma de se obter conclusões confiáveis a respeito de diferenças genotípicas (GILMOUR; CULLIS; VERBYLA, 1997; CULLIS et al., 1998; SMITH; CULLIS; THOMPSON, 2001; MALOSETTI et al., 2008).

Boer et al. (2007) afirmam que devem ser feitos testes explícitos para a presença, magnitude e forma da interação  $G \times E$ , com o intuito de auxiliar no entendimento da arquitetura genética de caracteres quantitativos. Ressaltam ainda que alguns trabalhos recentes evidenciam a importância dessa interação, enquanto outros demonstram o oposto.

Em um modelo misto com efeitos aleatórios para o produto cruzado de genótipos e ambientes, a matriz  $G$  contém as variâncias genéticas para ambientes individuais na diagonal e as covariâncias genéticas entre pares de ambientes fora da diagonal. Diversas formas de estruturação dessa matriz são encontradas na Tabela 1. Smith et al. (2007), trabalhando com dados do tipo MET de cana-de-açúcar, ressaltaram a existência de dois fatores distintos de variação: locais e anos de corte, por se tratar de uma cultura perene e, portanto, haver medidas repetidas ao longo do tempo para os

mesmos genótipos. Logo, propuseram uma maneira alternativa de modelar a estrutura da matriz de variâncias e covariâncias, através de uma análise combinada de todos os locais e cortes. Ao invés de fazer uso de uma única matriz VCOV para a combinação fatorial de locais e cortes, em que cada combinação local  $\times$  corte corresponde a um ambiente, esses autores modelaram uma matriz de variâncias e covariâncias para corte (denotada em seu trabalho por  $\mathbf{G}_h$ ) e outra para locais ( $\mathbf{G}_l$ ), de modo que a matriz  $\mathbf{G}$  global pode ser obtida por  $\mathbf{G} = \mathbf{G}_h \otimes \mathbf{G}_l$ , em que  $\otimes$  denota o produto de Kronecker. Neste caso, cada uma dessas duas matrizes componentes pode assumir uma das formas exibidas na Tabela 1, e o modelo final pode ter número de parâmetros sensivelmente menor que o obtido pela abordagem convencional. Apenas como exemplo, considere-se um cenário hipotético em que genótipos foram avaliados em 2 cortes e 2 locais. Uma matriz de variâncias e covariâncias não estruturada para os 4 ambientes ( $a$ ) pode ser escrita como:

$$\mathbf{G} = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a1a2} & \sigma_{a1a3} & \sigma_{a1a4} \\ \sigma_{a2a1} & \sigma_{a2}^2 & \sigma_{a2a3} & \sigma_{a2a4} \\ \sigma_{a3a1} & \sigma_{a3a2} & \sigma_{a3}^2 & \sigma_{a3a4} \\ \sigma_{a4a1} & \sigma_{a4a2} & \sigma_{a4a3} & \sigma_{a4}^2 \end{bmatrix}.$$

Como tal matriz é simétrica, ou seja,  $\sigma_{ij} = \sigma_{ji} \forall i, j \in \{a1, a2, a3, a4\}$ , há 10 parâmetros no total. Já quando são modeladas duas matrizes não estruturadas independentes, uma para cortes e outra para locais, tem-se:

$$\mathbf{G}_h = \begin{bmatrix} \sigma_{h1}^2 & \sigma_{h1h2} \\ \sigma_{h2h1} & \sigma_{h2}^2 \end{bmatrix} \quad \text{e} \quad \mathbf{G}_l = \begin{bmatrix} \sigma_{l1}^2 & \sigma_{l1l2} \\ \sigma_{l2l1} & \sigma_{l2}^2 \end{bmatrix}.$$

Há 3 parâmetros em cada matriz, de modo que há 6 no total. A redução no número de parâmetros torna-se mais marcante à medida que o número de cortes e/ou locais aumenta. Por exemplo, com 3 cortes e 10 locais, é necessário estimar 465 parâmetros quando é utilizada uma única matriz para todas as combinações, e apenas 61 quando o produto de Kronecker é empregado.

Além de considerar diversas condições ambientais, os experimentos em programas de melhoramento genético comumente envolvem a avaliação de diversos caracteres simultaneamente, visto que materiais elite devem concentrar alelos favoráveis para produção, resistência a doenças, pragas e estresses abióticos, caracteres agrônômicos e de qualidade dos produtos, entre outros (WELHAM et al., 2010). Além disso, caracteres de baixa herdabilidade ou de difícil mensuração podem ser

melhorados através de seleção indireta realizada em caracteres correlacionados e com maior herdabilidade (MALOSETTI et al., 2008).

Como os experimentos de melhoramento genético geralmente possuem dados de múltiplos caracteres e múltiplos ambientes (em inglês, *multi-trait multi-environment* - MTME), a ocorrência de interação  $G \times E$  em dados de vários caracteres é ainda mais desafiadora do que no caso univariado, no que diz respeito à análise e interpretação dos resultados (MALOSETTI et al., 2008). Tal fato é notado pela escassez de literatura e metodologia estatística para o tratamento adequado de dados dessa natureza.

Tradicionalmente, esses dados têm sido avaliados separadamente, ajustando-se um modelo para cada caráter e cada local, através de metodologias convencionais (PIEPHO, 2000). Em seguida, são realizadas comparações não-formais, geralmente gráficas, ou alguma forma de análise conjunta é empregada para tirar conclusões a respeito da interação  $G \times E$  e da correlação entre caracteres. Tinker e Mather, 1995 apud Piepho (2000) afirmam que a realização de análises separadas elimina a complicação de lidar com a interação  $G \times E$  e com a heterogeneidade ambiental, o que torna a abordagem útil caso haja interesse apenas no conjunto de ambientes considerado. Por outro lado, a interpretação de tais resultados é complicada e não se faz uso da replicação embutida fornecida por múltiplos ambientes.

Segundo Malosetti et al. (2008), a modelagem direta da interação  $G \times E$  e das correlações genéticas entre caracteres, em uma única análise MTME, aumenta o poder estatístico e torna a análise mais realista. Ainda segundo esses autores, as suposições de variância genética homogênea em todos os ambientes e de ausência de covariâncias genéticas entre ambientes e caracteres são irrealistas, de modo que os modelos mistos representam uma abordagem natural e apropriada para a análise conjunta desses dados. Claramente, para tanto, modelos mais sofisticados são necessários.

#### **2.4.2 Mapeamento de QTLs**

Piepho (2000) contrasta duas abordagens para o ajuste de modelos de busca por QTLs: os baseados no método exato da máxima verossimilhança sob modelos de misturas (LANDER; BOTSTEIN, 1989; ZENG, 1993, 1994; JANSEN; STAM, 1994; KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999), discutidos anteriormente, e os baseados na aproximação por regressão linear múltipla (HALLEY; KNOTT, 1992; MARTÍNEZ; CURNOW, 1992). Broman (2001) e Broman e Sen (2009) afirmam que este último método fornece uma aproximação notavelmente satisfatória para o mape-

amento por intervalo tradicional, com as vantagens de exigir muito menos tempo computacional e poder ser realizado com programas convencionais de estatística. É feita uma ressalva para o fato de que a aproximação pode não ser adequada em regiões com pouca informação genotípica, como em intervalos excessivamente grandes entre marcadores adjacentes e na proximidade de marcadores com muitos dados perdidos. Piepho (2000) argumenta que as vantagens dessa aproximação são ainda mais consideráveis em situações muito complexas, como por exemplo no caso de o modelo incluir vários efeitos aleatórios. Devido a esses motivos, o mapeamento de QTLs no contexto dos modelos mistos é feito através da aproximação por regressão linear.

Apesar da abundância de metodologias para modelagem da interação  $G \times E$  em dados de METs e do grande número de trabalhos sobre mapeamento de QTLs, pouco tem sido feito para a incorporação da base genética da interação  $G \times E$ , ou seja, da expressão diferencial de QTLs em função de variações ambientais - a chamada interação QTL por ambiente, ou  $QTL \times E$  (JIANG; ZENG, 1995; MALOSETTI et al., 2004). Merecem destaque os trabalhos de Jiang e Zeng (1995), no caso de modelos fixos de mistura, e de Piepho (2000) e Verbyla et al. (2003), no caso de modelos mistos. Há na literatura alguns outros estudos de mapeamento de QTLs em dados de múltiplos ambientes ou múltiplos caracteres, os quais empregaram modelos mistos para tanto (CROSSA et al., 1999; MALOSETTI et al., 2004; EEUWIJK et al., 2005, 2010; VARGAS et al., 2006; MATHEWS et al., 2008).

Recentemente, Pastina (2010) adaptou as ideias de Gazaffi (2009) e as empregou sob o prisma dos modelos mistos, de modo a combinar a flexibilidade desta classe de modelos em ajustar adequadamente a estrutura de variância-covariância intrínseca dos dados com os conceitos envolvidos na busca por QTLs em progênies de irmãos-completos. Novamente, tal abordagem foi empregada para análise dos mesmos dados de cana-de-açúcar utilizados por Gazaffi (2009), o que resultou na detecção de diversos QTLs para caracteres de produção, ao mesmo tempo em que a variação residual foi devidamente controlada para efeitos de local, corte, controle local dos experimentos e presença de testemunhas comerciais, bem como suas diversas interações. Além disso, foi realizada a modelagem apropriada das matrizes de covariâncias genéticas dentro de locais e de cortes (PASTINA, 2010), fato que por sua vez torna não viesadas as estimativas dos efeitos dos QTLs (PIEPHO, 2000; MALOSETTI et al., 2004).

Para um conjunto de dados com estrutura do tipo MTME, é possível identificar QTLs com efeitos em múltiplos caracteres em múltiplos ambientes, através da inclusão de informação dos

marcadores moleculares (MALOSETTI et al., 2008). Tanto no caso dos modelos fixos baseados em verossimilhança vistos anteriormente, quanto para os trabalhos de mapeamento de QTLs no contexto dos modelos mistos para dados deste tipo, as análises foram restritas a múltiplos caracteres ou múltiplos ambientes, sem integrar essas duas camadas de informação (PIEPHO, 2000; VERBYLA et al., 2003; MALOSETTI et al., 2004; BOER et al., 2007). Sob o prisma dos modelos mistos, a busca por QTLs em uma análise integrada de dados MTME foi realizada apenas por Malosetti et al. (2006, 2008), o que torna bastante evidente a necessidade de desenvolvimento de maior número de estudos dessa natureza, em especial para espécies perenes como a cana-de-açúcar.

A maior vantagem de mapear QTLs sob a abordagem dos modelos mistos, em uma única análise MTME, é que os efeitos dos QTLs são testados levando em consideração a estrutura de correlação genética presente nos dados. Isto garante a obtenção de estimativas não viesadas e, por conseguinte, torna as conclusões mais confiáveis. A não inclusão de correlações genéticas em dados dessa natureza aumenta a ocorrência de falsos positivos, ou seja, leva à detecção de QTLs espúrios, de modo que os resultados tornam-se artificial e excessivamente otimistas (PIEPHO, 2005; MALOSETTI et al., 2008). Assim como visto para os modelos fixos, uma análise de QTLs conjunta para múltiplos caracteres e múltiplos ambientes aumenta o poder estatístico de detecção de QTLs (MALOSETTI et al., 2006; GUO; NELSON, 2008).

Os QTLs podem ser incluídos no modelo como efeitos fixos ou aleatórios, em função dos objetivos do trabalho (WANG et al., 1999; PIEPHO, 2000; VERBYLA et al., 2003; MALOSETTI et al., 2004; BOER et al., 2007). Isto é aliado à flexibilidade de modelagem da matriz de variâncias e covariâncias, a qual inclui a correlação devida a efeitos poligênicos, isto é, a QTLs não incluídos explicitamente no modelo (PIEPHO, 2000).

Além disso, esta análise pode fornecer informações valiosas a respeito das bases genéticas de como os caracteres se relacionam. Em outras palavras, é possível determinar se há predominância de QTLs com efeito em caracteres isolados ou de QTLs com efeitos pleiotrópicos (GUO; NELSON, 2008; MALOSETTI et al., 2008). É ainda possível verificar se cada QTL pleiotrópico afeta os caracteres correlacionados no mesmo sentido da correlação fenotípica, ou se há variação desses sinais, de modo que possa ser feita seleção no sentido de quebrar ou amenizar essa correlação.

Outro aspecto interessante é o estudo da base genética da interação  $G \times E$ . Através do estudo das interações  $QTL \times E$ , pode-se detectar quais QTLs têm efeitos constantes em todos os ambientes e quais são por eles influenciados, com alterações de magnitude ou até dos sinais dos efeitos. Assim,

ao passo que alguns QTLs não contribuem em absoluto para a interação  $G \times E$ , os efeitos dos demais são influenciados pelas condições ambientais com maior ou menor intensidade, e isto evidencia a importância individual de cada um em explicar a ocorrência de interação  $G \times E$  (MALOSETTI et al., 2008). A compreensão do impacto e do padrão de interação  $QTL \times E$  é essencial para o desenvolvimento de trabalhos de seleção assistida por marcadores (BOER et al., 2007).

Do ponto de vista do melhoramento genético, enfim, os resultados de uma análise de QTLs em dados MTME são muito interessantes, pois fornecem informações a respeito da interação  $QTL \times E$ , das causas de correlação genética entre caracteres (pleiotropia e/ou ligação gênica) e da intersecção desses dois níveis, isto é, da ocorrência de QTLs ligados e/ou pleiotrópicos exibindo interação  $QTL \times E$ . Este último fato pode fazer, por exemplo, com que as correlações genéticas entre caracteres variem ao longo dos ambientes (MALOSETTI et al., 2008). Conforme mencionado anteriormente, o uso de modelos mistos também é interessante por melhorar a parte experimental da análise, visto que fontes de variação adicionais, como o efeito de blocos e de covariáveis, são incluídas diretamente na análise, em contraste ao simples uso de médias aritméticas como é feito usualmente. Além disso, podem ser feitas suposições mais realistas com relação à estrutura de correlação entre os resíduos, levando-se em consideração fontes sistemáticas e aleatórias de erro no campo experimental.

Boer et al. (2007) salientam que os efeitos principais dos QTLs e os efeitos de interação  $QTL \times E$  são observados em um dado *background* genético. Neste caso, a combinação da epistasia entre QTLs e o *background* com a interação  $QTL \times E$  deve afetar os resultados da seleção assistida por marcadores. A esse respeito, Wang et al. (1999) propuseram o mapeamento de QTLs com efeitos epistáticos e interação  $QTL \times E$  através da abordagem de modelos mistos, sendo este o único trabalho encontrado na literatura que fez uso da abordagem de modelos mistos para estudar, explicitamente, as interações entre QTLs. Na realidade, esses autores ressaltam que a questão da epistasia ainda é um problema em grande parte não resolvido. À exceção do método MIM, as análises de QTLs geralmente assumem ausência de epistasia para simplificação dos modelos estatísticos. No entanto, a violação desta forte suposição resulta em estimativas viesadas dos efeitos e das posições dos QTLs, além de reduzir a precisão e o poder de detecção.

Wang et al. (1999) ainda destacam que os poucos trabalhos que utilizam mapas genéticos para analisar QTLs com efeitos epistáticos não permitem a quantificação e caracterização acurada da epistasia na presença de interação  $G \times E$  e, portanto, de interação  $QTL \times E$ . Em contrapartida, os trabalhos sobre mapeamento de QTLs que incluem a interação  $QTL \times E$  não consideram a ocor-

rência de epistasia, de modo que a interação  $G \times E$  não pode ser adequadamente dissecada em seus componentes  $QTL \times E$ . De fato, Jiang e Zeng (1995) propuseram que seu modelo mCIM deveria ser expandido para incluir efeitos epistáticos, além de ser possivelmente transportado para o contexto dos modelos mistos, com o intuito de levar em conta fatores experimentais como efeito de blocos, por exemplo. Analogamente ao método CIM, o qual inclui cofatores com efeitos principais para absorver a variação genética do *background*, Wang et al. (1999) advocam a inclusão de marcadores interatuantes ligados a QTLs epistáticos com o objetivo de aumentar o poder e a acurácia do mapeamento de QTLs. No entanto, neste caso os cofatores são incluídos como efeitos aleatórios apenas como forma de controlar a variação genética residual, e não há interesse particular em estimar esses efeitos epistáticos. Este último caso seria, com efeito, o objetivo do método MIM, o qual ainda não foi abordado no contexto dos modelos mistos.

## 3 MATERIAL E MÉTODOS

### 3.1 Material

#### 3.1.1 Material Vegetal

Dados moleculares e fenotípicos foram obtidos para uma progênie oriunda de um cruzamento biparental entre os clones de elite pré-comerciais SP80-180 (B3337 x policruzamento) e SP80-4966 (SP71-1406 x policruzamento), realizado em 1996 na Estação Experimental do Centro de Tecnologia Canavieira (CTC), localizada no município de Camamu, BA. O clone SP80-180 (genitor feminino) apresenta baixo teor de sacarose e alta produtividade de colmos, enquanto que SP80-4966 (genitor masculino) tem alto teor de sacarose e baixa produção de colmos.

Desse cruzamento resultou uma progênie composta por 500 indivíduos. A população de mapeamento foi formada por uma amostra aleatória de 100 indivíduos representativos da progênie original, os quais forneceram material vegetal para genotipagem e para os ensaios de produção.

#### 3.1.2 Dados Moleculares

Marcadores do tipo *Restriction Fragment Length Polymorphism* (RFLP) e *Expressed Sequence Tag* (EST) derivados de RFLP e *Simple Sequence Repeat* (SSR) foram utilizados para genotipagem dos genitores e da progênie. Todos estes marcadores já haviam sido gerados e codificados, conforme detalhado em Garcia et al. (2006) e Oliveira et al. (2007). Para atribuição de valores aos marcadores, cada alelo segregante foi avaliado com base na presença ou ausência em cada indivíduo da progênie, comportando-se, portanto, como um marcador dominante. Apenas marcadores em dose única foram efetivamente considerados. Os padrões de segregação dos marcadores foram testados contra as razões esperadas através do teste de qui-quadrado ( $\chi^2$ ), para as situações de marcadores em dose única em apenas um dos genitores (padrão 1:1) e em ambos os genitores (padrão 3:1). Locos com fortes desvios em relação às proporções esperadas foram descartados, após o controle de erro tipo I para múltiplos testes através da correção de Bonferroni.

#### 3.1.3 Dados Fenotípicos

A progênie segregante foi plantada em dois locais no ano de 2003, nas cidades de Piracicaba



e Jaú, Estado de São Paulo. Foram realizadas avaliações nos três primeiros anos de colheita, ou seja, na cana-planta e nos dois primeiros anos de cana-soca. Os caracteres fenotípicos avaliados foram a produção de cana (toneladas de cana por hectare, ou TCH - *tonnes of cane per hectare*), percentual de fibra (*Fiber*) e conteúdo de açúcar no caldo de cana (*Pol*). Os acrônimos e os nomes dos caracteres foram aqui mantidos em inglês para facilitar a comparação dos resultados com os de Pastina (2010).

Em cada localidade, o delineamento experimental consistiu de um delineamento em blocos ao acaso com duas repetições. No entanto, os 100 clones não foram completamente casualizados. Em primeiro lugar, os clones foram aleatoriamente divididos em três grupos de 36, 38 e 26 clones, respectivamente. Foi praticada casualização apenas dos clones dentro desses grupos, enquanto que os grupos não foram casualizados. Em ambos os experimentos, cada um dos grupos foi aumentado com quatro testemunhas (os cultivares comerciais SP80-1842, SP81-3250, SP80-1816 e RB72454). Ambos os genitores foram também incluídos no primeiro grupo, mas não considerados nas análises estatísticas. Um *layout* dos experimentos de campo é apresentado na Figura 1.

## 3.2 Métodos

### 3.2.1 Mapa de Ligação

No total, foram utilizados 741 marcadores moleculares, com 459 locos apresentando padrão de segregação 1:1 (100 do tipo RFLP, 27 EST-RFLP e 332 EST-SSR) e 282 locos com segregação na proporção 3:1 (88 RFLP, 10 EST-RFLP e 184 EST-SSR). De acordo com a notação proposta por Wu et al. (2002b), marcadores segregando para o genitor SP80-180 foram denotados por  $D_1$ , apresentando a configuração alélica ' $ao \times oo$ '; locos informativos para o genitor SP80-4966 foram denotados por  $D_2$ , com a configuração ' $oo \times ao$ '; e marcadores segregando para ambos os genitores foram denotados por  $C$ , com configuração ' $ao \times ao$ '. Em todos esses casos, o alelo  $a$  é dominante sobre o alelo (nulo)  $o$ .

A construção do mapa foi executada através do pacote *OneMap* (MARGARIDO; SOUZA; GARCIA, 2007), com base em uma abordagem multiponto (WU et al., 2002a, 2002b). Inicialmente, análises de dois-pontos foram empregadas para a determinação dos grupos de ligação, considerando-se um limiar mínimo de 6 para o LOD Score e um limiar máximo de 0.5 para a fra-

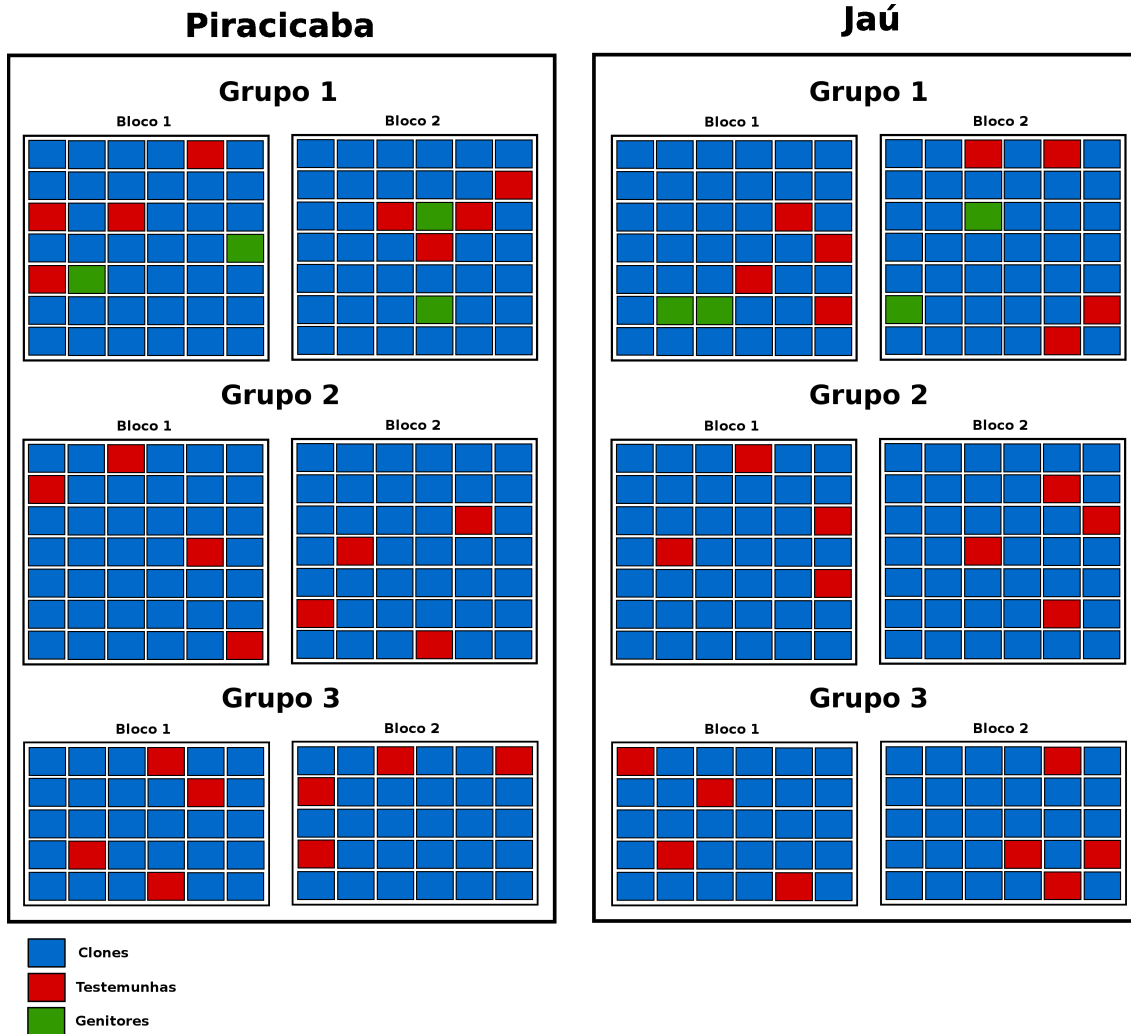


Figura 1 – Local 1: Piracicaba, SP; e Local 2: Jaú, SP

ção de recombinação. Para os menores grupos de ligação, os quais eram compostos por cinco ou menos marcadores, todas as possíveis ordens foram comparadas e a mais provável foi selecionada, utilizando-se como critério de seleção o logaritmo da verossimilhança. Tal procedimento é análogo ao comando `compare` implementado no software MAPMAKER/EXP (LANDER et al., 1987). Já no caso dos grupos de ligação com mais de cinco marcadores, o algoritmo `order` foi iniciado com cinco marcadores, os quais foram ordenados através da comparação de todas as possíveis ordens; em seguida, os marcadores restantes foram sequencialmente incluídos no mapa de ligação parcial, sempre na posição de maior verossimilhança. Tal procedimento é similar ao realizado pelo comando `try` do software MAPMAKER/EXP. Maiores detalhes a respeito desses algoritmos e a comprovação de sua eficácia podem ser encontrados em Mollinari et al. (2009). Finalmente, o comando

$$\begin{array}{c}
 P \qquad \qquad \qquad Q \\
 \begin{array}{c}
 P_m^1 \quad \left| \quad \left| \quad P_m^2 \quad \quad Q_m^1 \quad \left| \quad \left| \quad Q_m^2 \\
 P^1 \quad \quad \quad \quad P^2 \quad \quad \times \quad Q^1 \quad \quad \quad \quad Q^2 \\
 P_{m+1}^1 \quad \left| \quad \left| \quad P_{m+1}^2 \quad \quad Q_{m+1}^1 \quad \left| \quad \left| \quad Q_{m+1}^2
 \end{array}
 \end{array}$$

Figura 2 – Representação de um cruzamento biparental entre genitores heterozigóticos  $P$  e  $Q$ .  $P_m^{\{1,2\}}$ ,  $Q_m^{\{1,2\}}$ ,  $P_{m+1}^{\{1,2\}}$  e  $Q_{m+1}^{\{1,2\}}$  são os alelos dos marcadores nos locos  $m$  e  $m + 1$ ;  $P^1$ ,  $P^2$ ,  $Q^1$  e  $Q^2$  são os alelos do QTL

ripple foi aplicado a fim de se verificar se inversões locais haviam ocorrido. Para apresentação final dos mapas obtidos, as frações de recombinação foram convertidas em distâncias genéticas e expressas em centiMorgans (cM), através da função de mapeamento de Kosambi (KOSAMBI, 1944).

Por fim, os grupos de ligação foram reunidos em grupos de homologia putativos. Para que dois grupos de ligação fossem associados ao mesmo grupo de homologia, pelo menos dois pares de alelos derivados do mesmo loco RFLP, EST-RFLP ou EST-SSR deveriam ser compartilhados entre eles (OKADA et al., 2010).

### 3.2.2 Preditores Genéticos

Seguindo a notação proposta por Lin et al. (2003), considera-se uma progênie de irmãos-completos obtida do cruzamento entre dois indivíduos parentais heterozigóticos, denotados por  $P$  e  $Q$  (Figura 2).

Tal modelo foi originalmente proposto para espécies diplóides, de modo que a notação é mais abrangente e engloba diversos outros padrões de segregação, os quais não são observados para cana-de-açúcar. De fato, no presente trabalho apenas marcadores em dose única foram considerados, de modo que aqui os contrastes representam apenas categorias do tipos presença e ausência dos alelos do QTL. Para um intervalo definido por dois marcadores,  $m$  e  $m + 1$ , cada um com alelos denotados por 1 e 2, os genótipos para estes locos podem ser representados por  $P_m^{\{1,2\}}$ ,  $Q_m^{\{1,2\}}$ ,  $P_{m+1}^{\{1,2\}}$  e  $Q_{m+1}^{\{1,2\}}$ , em que  $\{1, 2\}$  indica as possibilidades alélicas para cada loco. Supõe-se ainda a existência de um QTL entre esses dois marcadores, com alelos  $P^1$  e  $P^2$  para o genitor  $P$ ,  $Q^1$  e  $Q^2$  para o genitor  $Q$ . Assim, a segregação do QTL na progênie resultará na ocorrência de quatro classes genotípicas ( $P^1Q^1$ ,  $P^1Q^2$ ,  $P^2Q^1$  e  $P^2Q^2$ ), com proporção de segregação igual a 1:1:1:1. Deste modo, havendo

quatro classes genóticas com a mesma frequência esperada, podem ser definidos três contrastes ortogonais:

$$\begin{aligned}
 &+ P^1Q^1 + P^1Q^2 - P^2Q^1 - P^2Q^2 \\
 &+ P^1Q^1 - P^1Q^2 + P^2Q^1 - P^2Q^2 \\
 &+ P^1Q^1 - P^1Q^2 - P^2Q^1 + P^2Q^2
 \end{aligned}$$

Os dois primeiros contrastes correspondem a efeitos aditivos do QTL nos genitores  $P$  e  $Q$ , denotados por  $\alpha_p$  e  $\alpha_q$ , respectivamente; o terceiro e último contraste corresponde ao efeito de dominância ou interação intra-alélica do QTL, isto é, à interação entre os efeitos aditivos de cada genitor, denotado por  $\delta_{pq}$ . Preditores genéticos foram obtidos para uma série de posições  $w$  ao longo do genoma ( $w = 1, \dots, W$ ). Estes preditores genéticos foram posteriormente incluídos como variáveis explanatórias nos modelos mistos para busca por QTLs. Para o indivíduo  $i$  e no ponto  $w$  do genoma, os preditores genéticos foram obtidos por:

$$\begin{aligned}
 x_{p_{iw}} &= p(P^1Q^1|\mathbf{M}_i) + p(P^1Q^2|\mathbf{M}_i) - p(P^2Q^1|\mathbf{M}_i) - p(P^2Q^2|\mathbf{M}_i) \\
 x_{q_{iw}} &= p(P^1Q^1|\mathbf{M}_i) - p(P^1Q^2|\mathbf{M}_i) + p(P^2Q^1|\mathbf{M}_i) - p(P^2Q^2|\mathbf{M}_i) \\
 x_{pq_{iw}} &= p(P^1Q^1|\mathbf{M}_i) - p(P^1Q^2|\mathbf{M}_i) - p(P^2Q^1|\mathbf{M}_i) + p(P^2Q^2|\mathbf{M}_i)
 \end{aligned}$$

em que  $x_{p_{iw}}$ ,  $x_{q_{iw}}$  e  $x_{pq_{iw}}$  são os valores esperados das variáveis explanatórias para os efeitos aditivos do QTL nos genitores  $P$  e  $Q$ , e para o efeito de dominância, respectivamente, na posição  $w$ , dada a informação de todos os marcadores moleculares ( $\mathbf{M}_i$ ) no grupo de ligação correspondente ao intervalo em questão, para o indivíduo  $i$  (HALEY; KNOTT, 1992; MARTÍNEZ; CURNOW, 1992; LYNCH; WALSH, 1998). As probabilidades condicionais  $p(P^1Q^1|\mathbf{M}_i)$ ,  $p(P^1Q^2|\mathbf{M}_i)$ ,  $p(P^2Q^1|\mathbf{M}_i)$  e  $p(P^2Q^2|\mathbf{M}_i)$  foram calculadas de acordo com a abordagem multiponto, através de um modelo de cadeia de Markov oculta implementado na versão 1.0-1 do pacote *OneMap* (MARGARIDO; SOUZA; GARCIA, 2007). Esta metodologia multiponto é bastante superior à tradicional abordagem de dois pontos, pois contorna a propagação de erros devida à falta de informatividade de algumas combinações de marcadores, fato particularmente importante quando há marcadores dominantes ligados em repulsão (MOLLINARI et al., 2009). Esses valores de probabilidade foram obtidos para todas as posições com marcadores genéticos e em uma série de pseudo-marcadores ao longo de todo o genoma, espaçados em no máximo 1 cM.

Especificamente no caso do mapa de ligação utilizado no presente trabalho, dada a complexidade genômica da cana-de-açúcar e visto que apenas marcadores com padrão de segregação 1:1 e

3:1 foram empregados, na grande maioria das posições do genoma os marcadores não forneceram informação suficiente para separação das quatro classes genótípicas. Como consequência, nessas regiões a matriz de preditores genéticos pode ser singular, ou seja, alguns preditores podem ser obtidos a partir de combinações lineares dos outros. É bastante conhecido o fato de que a multicolinearidade pode causar problemas graves na estimação dos parâmetros em um modelo de regressão, devido a problemas numéricos computacionais resultantes de uma sensibilidade excessiva da solução de uma equação linear a mudanças nos elementos do sistema. Consequentemente, as estimativas obtidas têm sua precisão reduzida, ou seja, variâncias altas (BELSLEY; KUH; WELSCH, 1980). Portanto, a ocorrência de multicolinearidade foi investigada para cada posição genômica através do cálculo dos índices de condição, oriundos da decomposição de valor singular da matriz de preditores (BELSLEY; KUH; WELSCH, 1980). Apenas os contrastes informativos foram incluídos na análise de QTLs. Como exemplo, é natural que grupos de ligação compostos por marcadores apenas do tipo  $D_1$  forneçam informação suficiente apenas para a estimação do contraste referente ao efeito aditivo no genitor  $P$  ( $x_{piw}$ ), uma situação análoga ao que ocorre no duplo *pseudo-testcross*. O mesmo ocorre com marcadores do tipo  $D_2$ . Mesmo no caso de grupos de ligação com os três tipos de marcadores, colinearidades podem surgir em função da distribuição desses locos.

### 3.2.3 Análise Fenotípica

O primeiro passo da análise de dados consistiu em ajustar um modelo misto apropriado para os dados fenotípicos. Para tanto, foram ajustadas e comparadas diferentes estruturas de variância-covariância (VCOV) para o efeito genético dentro de caracteres, locais e cortes (Tabela 2). Devido a limitações do software Genstat (PAYNE et al., 2009), o qual foi utilizado para a realização das análises com modelos mistos, as seis possíveis fatorações de local e corte foram combinadas para gerar um único fator, denominado aqui por ambiente. Assim, por exemplo, o primeiro corte em Piracicaba é considerado um ambiente; analogamente, o primeiro corte em Jaú é um outro ambiente. Os segundo e terceiro cortes em cada um dos dois locais completam os possíveis ambientes.

O modelo aqui empregado é uma versão multi-variada do apresentado por Pastina (2010), de modo que a descrição matemática do modelo seguirá a mesma notação. Por sua vez, esta notação é similar à apresentada por Eckermann et al. (2001), Verbyla et al. (2003) e Boer et al. (2007). O

modelo estatístico empregado foi (variáveis sublinhadas correspondem a efeitos aleatórios):

$$\underline{y}_{iscjkr} = \mu_{scjkr} + \underline{G}_{iscjk} + \underline{\varepsilon}_{iscjkr} \quad (2)$$

em que  $\underline{y}_{iscjkr}$  é o fenótipo da  $r$ -ésima repetição do  $i$ -ésimo genótipo no grupo  $s$ , para o caráter  $c$  no local  $j$  e corte  $k$ ;  $\mu_{scjkr}$  é a média do grupo  $s$ , para o caráter  $c$  no local  $j$  e corte  $k$ ;  $\underline{G}_{iscjk}$  é o efeito genético do genótipo  $i$  no grupo  $s$ , para o caráter  $c$  no local  $j$  e corte  $k$ ; e  $\underline{\varepsilon}_{iscjkr}$  é um efeito não genético. De acordo com sua origem, os genótipos podem ser separados em dois grupos,  $n = n_g + n_c$ , em que  $n_g$  é o número de genótipos (clones) na progênie ( $i = 1, \dots, n_g$ ), e  $n_c$  é o número de testemunhas ( $i = n_g + 1, \dots, n_g + n_c$ ). O modelo para  $\underline{G}_{iscjk}$  é dado por:

$$\underline{G}_{iscjk} = \begin{cases} \underline{g}_{ica} & i = 1, \dots, n_g \\ t_{iscjk} & i = n_g + 1, \dots, n_g + n_c \end{cases} \quad (3)$$

em que  $\underline{g}_{ica}$  é uma variável aleatória para o efeito genético do genótipo  $i$  para o caráter  $c$  no ambiente  $a$ , e  $t_{iscjk}$  representa um efeito fixo para a testemunha  $i$  no grupo  $s$ , para o caráter  $c$  no local  $j$  e corte  $k$ .

É importante mencionar que o software Genstat permite o ajuste de efeitos fixos separados para local e corte, no caso das testemunhas, mas não permite que o mesmo seja feito para a variável aleatória referente aos genótipos da progênie, de modo que o local e corte foram combinados para formar os ambientes, indexados por  $a$ . Apesar de as testemunhas não serem relevantes para a detecção de QTLs, elas são importantes em fornecer informação sobre a variação não-genética que pode estar presente (VERBYLA et al., 2003; BOER et al., 2007). Assumiu-se que o vetor  $\underline{\mathbf{g}} = (g_{1111}, \dots, g_{ICJK})$  tem uma distribuição normal multivariada com vetor de médias zero e matriz VCOV  $\mathbf{G}$  (Tabela 2),  $\underline{\mathbf{g}} \sim N(0, \mathbf{G})$ . No caso do termo referente ao erro residual, assumiu-se a distribuição  $\underline{\varepsilon}_{iscjkr} \sim N(0, \mathbf{R})$ .

A análise do mesmo conjunto de dados para cada caráter isoladamente, realizada por Pastina (2010), evidenciou que um modelo com  $\mathbf{R} = \sigma^2 \mathbf{I}$  mostrou-se adequado, ou seja, as variâncias genéticas foram similares para os vários ambientes, e não houve correlação entre as diferentes combinações de local e corte. No presente estudo, por se tratar de um modelo multivariado, a suposição de variâncias homogêneas não faz sentido, pois a escala de medida dos caracteres é diferente. Assim, espera-se uma variância muito maior para o caráter produção de colmos, o qual

foi medido em toneladas por hectare, do que para os valores percentuais de fibra e açúcar no caldo de cana. Portanto, ajustou-se aqui um modelo com variâncias residuais heterogêneas, uma para cada combinação caráter  $\times$  local  $\times$  corte, com covariâncias iguais a zero. Quatorze diferentes modelos para a matriz VCOV foram examinados e comparados com base no valor de BIC (Tabela 2).

### 3.2.4 Análise de QTLs

O modelo utilizado para o mapeamento de QTLs foi análogo ao Mapeamento de Múltiplos Intervalos (MIM) (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999), o qual permite o ajuste de múltiplos QTLs simultaneamente, tanto com efeitos principais quanto epistáticos. No entanto, no presente caso o modelo MIM não foi tratado como um modelo de misturas, mas sim foi empregada a aproximação de mínimos quadrados. Assim, o modelo misto apresentado no item anterior foi expandido para incluir informação dos marcadores moleculares:

$$\begin{aligned} \underline{y}_{iscjkr} = & \mu_{scjkr} + \sum_{w=1}^m (x_{p_{iw}} \alpha_{p_{cjkw}} + x_{q_{iw}} \alpha_{q_{cjkw}} + x_{pq_{iw}} \delta_{pq_{cjkw}}) \\ & + \sum_{w \neq z \in (1, \dots, m)}^n \sum_{\tau_1} \sum_{\tau_2} I_{(\tau_1 \tau_2)_{wz}} \left( x_{(\tau_1 \tau_2)_{iwz}} \beta_{(\tau_1 \tau_2)_{cjkwz}} \right) + \underline{G}_{iscjk}^* + \underline{\varepsilon}_{iscjkr} \end{aligned} \quad (4)$$

em que  $\alpha_{p_{cjkw}}$ ,  $\alpha_{q_{cjkw}}$  e  $\delta_{pq_{cjkw}}$  são os efeitos dos preditores genéticos aditivos para o genitor  $P$  e  $Q$ , e do preditor genético de dominância, respectivamente, específicos para cada combinação de caráter  $\times$  local  $\times$  corte, para o QTL da posição  $w$ . Similarmente,  $\beta_{(\tau_1 \tau_2)_{cjkwz}}$  é o efeito do preditor genético epistático  $\tau_1 \tau_2$  entre os QTLs das posições  $w$  e  $z$ , específico para cada caráter-local-corte. Os índices  $\tau_1$  e  $\tau_2$  consideram os possíveis contrastes para os eventuais QTLs presentes nas posições  $w$  e  $z$ , respectivamente, e  $I_{(\tau_1 \tau_2)_{wz}}$  é uma variável indicadora que assume o valor 1 se o efeito epistático correspondente está presente no modelo, ou 0 caso contrário. Os preditores genéticos para os efeitos de epistasia ( $x_{(\tau_1 \tau_2)_{iwz}}$ ) são obtidos simplesmente pela multiplicação dos preditores genéticos para os efeitos principais correspondentes, devido à propriedade distributiva da multiplicação. O termo  $\underline{G}_{iscjk}^*$  refere-se à variação genética residual, não explicada pelos QTLs, sendo assim marcado por um asterisco para diferenciá-lo do termo genético descrito na análise fenotípica. A matriz VCOV utilizada para  $\underline{G}_{iscjk}^*$  foi a selecionada na análise fenotípica descrita no item anterior.

Assumindo que um QTL putativo sob teste não tenha efeito ao longo das possíveis combinações de caracteres, locais e cortes, a hipótese de nulidade testada através do teste de Wald (VERBEKE;

MOLENBERGHS, 2000) foi:

$$H_0 : \begin{cases} \alpha_{p_{111w}} = \alpha_{p_{112w}} = \dots = \alpha_{p_{CJKw}} = 0 \\ \alpha_{q_{111w}} = \alpha_{q_{112w}} = \dots = \alpha_{q_{CJKw}} = 0 \\ \delta_{pq_{111w}} = \delta_{pq_{112w}} = \dots = \delta_{pq_{CJKw}} = 0 \end{cases}$$

Analogamente, a hipótese de nulidade para um efeito epistático putativo foi:

$$H_0 : \beta_{(\tau_1 \tau_2)_{111wz}} = \beta_{(\tau_1 \tau_2)_{112wz}} = \dots = \beta_{(\tau_1 \tau_2)_{CJKwz}} = 0$$

Posições genômicas com  $p$ -valores significativos no perfil produzido pelo modelo (4), de acordo com critérios a serem descritos no item 3.2.4.1, foram sequencialmente incluídas no modelo.

O mesmo modelo (4) também foi aplicado à análise dos 424 marcadores não posicionados no maps, levando em consideração o fato de que, nestes casos, o preditor genético podia assumir apenas os valores  $-1$  (alelo nulo  $o$ ) ou  $1$  (alelo  $a$ ). Novamente, o teste de Wald foi empregado para identificar possíveis efeitos de QTLs associados a marcadores individuais.

### 3.2.4.1 Estratégia de Busca por QTLs

Após o ajuste de um modelo misto com estrutura de variâncias-covariâncias adequada aos dados fenotípicos, a busca por QTLs foi realizada em diversas etapas, com a inclusão sucessiva de efeitos significativos no modelo.

Inicialmente, foram feitas buscas por efeitos principais nos 96 grupos de ligação. Em cada busca, se o menor  $p$ -valor encontrado em todo o genoma era  $\leq 0,001$ , a posição correspondente era incluída no modelo. Este procedimento foi repetido até que efeitos principais significativos não pudessem mais ser encontrados. Em seguida, a mesma estratégia foi empregada para os marcadores individuais, não mapeados em grupos de ligação. Neste ponto, novamente foi utilizado o limiar  $0,001$  para o  $p$ -valor e marcas individuais foram sequencialmente incluídas no modelo enquanto significativas.

Finalizada esta primeira etapa de buscas, foram testados todos os possíveis efeitos epistáticos envolvendo os QTLs previamente encontrados, combinados dois a dois. Como o número de testes realizados neste passo é muito inferior ao que ocorre na busca por efeitos principais em todo o genoma, e com o intuito de favorecer a inclusão de termos epistáticos no modelo, foram incluídas



interações com  $p$ -valor  $\leq 0,05$ , mesmo sabendo que eventuais falso positivos poderiam ser incluídos no modelo.

A inclusão de cada novo efeito no modelo faz com que parte da variância fenotípica total seja explicada pelo QTL, o que resulta em redução da variância residual. Por sua vez, tal redução aumenta o poder estatístico dos testes de hipóteses (ZENG, 1993, 1994). Portanto, após a inclusão de termos epistáticos no modelo, foram realizadas novas etapas de buscas por efeitos principais, tanto nos grupos de ligação quanto nas marcas não ligadas. Em seguida, foram feitos novos testes para efeitos epistáticos, e este processo iterativo continuou até que novos efeitos não pudessem mais ser detectados.

### 3.2.4.2 Etapas Finais

Ao final do processo de busca por QTLs, dispunha-se de um modelo com múltiplos efeitos de QTLs, para cada caráter em cada local e corte. Para se determinar quais QTLs permaneciam significativos neste modelo, o teste de Wald foi aplicado após a remoção isolada de cada efeito individual de QTL do modelo completo. QTLs não significativos, apresentando  $p$ -valor superior a 0,05, foram excluídos. A seguir, cada um dos QTLs remanescentes foi testado para determinação da significância das interações QTL  $\times$  Local (QTL  $\times$  L), QTL  $\times$  Corte (QTL  $\times$  C) e QTL  $\times$  Local  $\times$  Corte (QTL  $\times$  L  $\times$  C), novamente através do teste de Wald. Apenas efeitos significativos de interação foram mantidos no modelo.

Em seguida, a posição de cada QTL foi refinada, segundo a proposta original do método MIM (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999; ZENG; KAO; BASTEN, 1999). Para tanto, um QTL por vez teve sua posição variada ao longo de todo o grupo de ligação correspondente, enquanto todos os QTLs foram mantidos fixos na posição em que foram encontrados. Tal procedimento gerou um perfil de  $p$ -valores para cada QTL, e a posição com o menor  $p$ -valor foi selecionada como a nova posição do QTL em questão. Esse refinamento foi realizado de maneira iterativa, até que a posição de todos os QTLs não mais se alterasse.

Após a obtenção deste modelo final com múltiplos QTLs, foi feita uma etapa auxiliar com o objetivo de verificar a existência de evidência em favor das hipóteses de pleiotropia ou ligação gênica. Esta etapa consistiu em remover um QTL por vez, bem como os efeitos epistáticos nos quais esse QTL estava envolvido, e subsequentemente gerar um perfil para cada um dos três caracteres separadamente. Isto foi feito incluindo um novo efeito para um único caráter, em todas as posições

do grupo de ligação, levando em consideração tanto o efeito principal quanto os termos epistáticos que haviam sido removidos. Apesar de não haver, no contexto dos modelos mistos, um teste formal para comparar as hipóteses de pleiotropia *versus* ligação, já que os efeitos de QTLs foram incluídos como efeitos fixos e os dois modelos não são aninhados, os perfis assim gerados indicaram se os picos para cada caráter foram semelhantes ou não, isto é, se as posições correspondentes aos menores *p*-valores eram as mesmas para todos os caracteres. Esta informação serviu como um procedimento *ad hoc* em prol ou em detrimento da hipótese de QTL com efeito pleiotrópico.

A partir do modelo final, foi também estimado o efeito de cada QTL para cada caráter, em cada local e corte, assim como seu respectivo desvio-padrão. Para determinação da significância individual de cada efeito, adotou-se o critério  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ , conforme proposto por Malosetti et al. (2008).

Todas as análises estatísticas envolvendo modelos mistos foram realizadas no software Genstat (PAYNE et al., 2009).

Tabela 2 – Modelos examinados para a matriz de variâncias-covariâncias genéticas ( $G$ )

Matriz $G$	Modelo	$n_{PAR}$ em $G^a$	Descrição
$G = G_{M \times M}^{caráter-amb}$	a) DIAG	$M$	Variância genética heterogênea
	b) $CS_{Het}$	$M + 1$	Simetria composta com variância genética heterogênea
	c) FA1	$2M$	Modelo fator-analítico de primeira ordem
	d) US	$\frac{M(M+1)}{2}$	Modelo não-estruturado
$G = G_{C \times C}^{caráter} \otimes G_{A \times A}^{amb}$	e) DIAG $\otimes$ DIAG	$(C + A) - 1$	Variância genética heterogênea dentro de caracteres e ambientes
	f) DIAG $\otimes$ $CS_{Het}$	$(C + A + 1) - 1$	Variância genética heterogênea dentro de caracteres e simetria composta com variância genética heterogênea dentro de ambientes
	g) DIAG $\otimes$ FA1	$(C + 2A) - 1$	Variância genética heterogênea dentro de caracteres e modelo fator analítico de primeira ordem dentro de ambientes
	h) DIAG $\otimes$ US	$\left[ C + \frac{A(A+1)}{2} \right] - 1$	Variância genética heterogênea dentro de caracteres e modelo não-estruturado dentro de ambientes
	i) $CS_{Het} \otimes CS_{Het}$	$(C + 1 + A + 1) - 1$	Simetria composta com variância genética heterogênea dentro de caracteres e ambientes
	j) $CS_{Het} \otimes$ FA1	$(C + 1 + 2A) - 1$	Simetria composta com variância genética heterogênea dentro de caracteres e modelo fator analítico de primeira ordem dentro de ambientes
	k) $CS_{Het} \otimes$ US	$\left[ C + 1 + \frac{A(A+1)}{2} \right] - 1$	Simetria composta com variância genética heterogênea dentro de caracteres e modelo não-estruturado dentro de ambientes
	l) US $\otimes$ $CS_{Het}$	$\left[ \frac{C(C+1)}{2} + A + 1 \right] - 1$	Modelo não-estruturado dentro de caracteres e simetria composta com variância genética heterogênea dentro de ambientes
	m) US $\otimes$ FA1	$\left[ \frac{C(C+1)}{2} + 2A \right] - 1$	Modelo não-estruturado dentro de caracteres e modelo fator analítico de primeira ordem dentro de ambientes
	n) US $\otimes$ US	$\left[ \frac{C(C+1)+A(A+1)}{2} \right] - 1$	Modelos não-estruturados para a variância genética dentro de caracteres e ambientes

Modelos (a-d) utilizam a combinação fatorial de caracteres e ambientes como diferentes “caracteres”. Modelos (e-n) utilizam o produto direto de matrizes de variância-covariância para caracteres e ambientes. <sup>a</sup> O número de parâmetros para os modelos (e-n) correspondem à soma dos números de parâmetros para as matrizes componentes menos o número de restrições para assegurar a identificabilidade do modelo. Note que o número total de parâmetros de variância-covariância no modelo é igual ao valor mostrado para a matriz  $G$  mais 18 do número de parâmetros na matriz  $R$ .  $M = C \times A$ , em que  $C$  é o número de caracteres e  $A$  é o número de ambientes; por sua vez,  $A = J \times K$ , em que  $J$  é o número de locais e  $K$  o número de cortes.

## 4 RESULTADOS

### 4.1 Mapa de Ligação

O mapa de ligação final consistiu de 96 grupos de ligação (ou *linkage groups* - LGs), que compreenderam 317 (42,8%) do total de 741 marcadores moleculares. Em conjunto, os grupos de ligação cobriram 2.468,14 cM, com uma distância média entre marcadores adjacentes de 7,5 cM. Do total de 96 grupos de ligação, 91 (94,8%) puderam ser organizados em onze grupos de homologia, sendo que os cinco grupos de ligação restantes não possuíam alelos em comum o suficiente para serem associados a grupo algum. O maior grupo de homologia (HGI) foi composto por vinte e dois grupos de ligação, enquanto que os menores (HGX e HGXI) foram compostos por apenas dois marcadores cada (Figura 3).

A observação desta figura revela que grande parte dos grupos de ligação (42 ou 43,7% do total) foi composta por apenas duas marcas ligadas, muitas vezes com distâncias inferiores a 10 cM. Além disso, apenas nove grupos apresentaram mais de cinco marcadores moleculares, sendo que os três maiores apresentaram 14, 11 e 10 marcas. Foram observados alguns pequenos agrupamentos mais adensados de marcadores (*clusters*), como por exemplo nos grupos 1 (HGI) e 3 (HGV). É também possível notar algumas distâncias superiores a 20 cM (*gaps*), o que ocorreu em 11,8% dos intervalos. De forma geral, os mapas genéticos de cana-de-açúcar realmente não são saturados, devido à utilização apenas de marcadores em dose única, o que resulta em ausência de agrupamento entre grupos de ligação homeólogos e, por conseguinte, em menor cobertura do genoma (ALWALA; KIMBENG, 2010).

### 4.2 Análise Fenotípica

Os termos incluídos na parte fixa do modelo, com o intuito de ajustar para os efeitos de blocos e testemunhas, foram ambos significativos e, portanto, mantidos em todas as análises seguintes. Diferentes estruturas de variância-covariância foram comparadas para o efeito de genótipos entre caracteres e ambientes e os resultados são apresentados na Tabela 3. Há dois grupos de modelos: no primeiro grupo (modelos a-d), cada combinação caráter-ambiente é tratada como um diferente “caráter”; já no segundo grupo (modelos e-n), é modelada uma matriz de variâncias-covariâncias

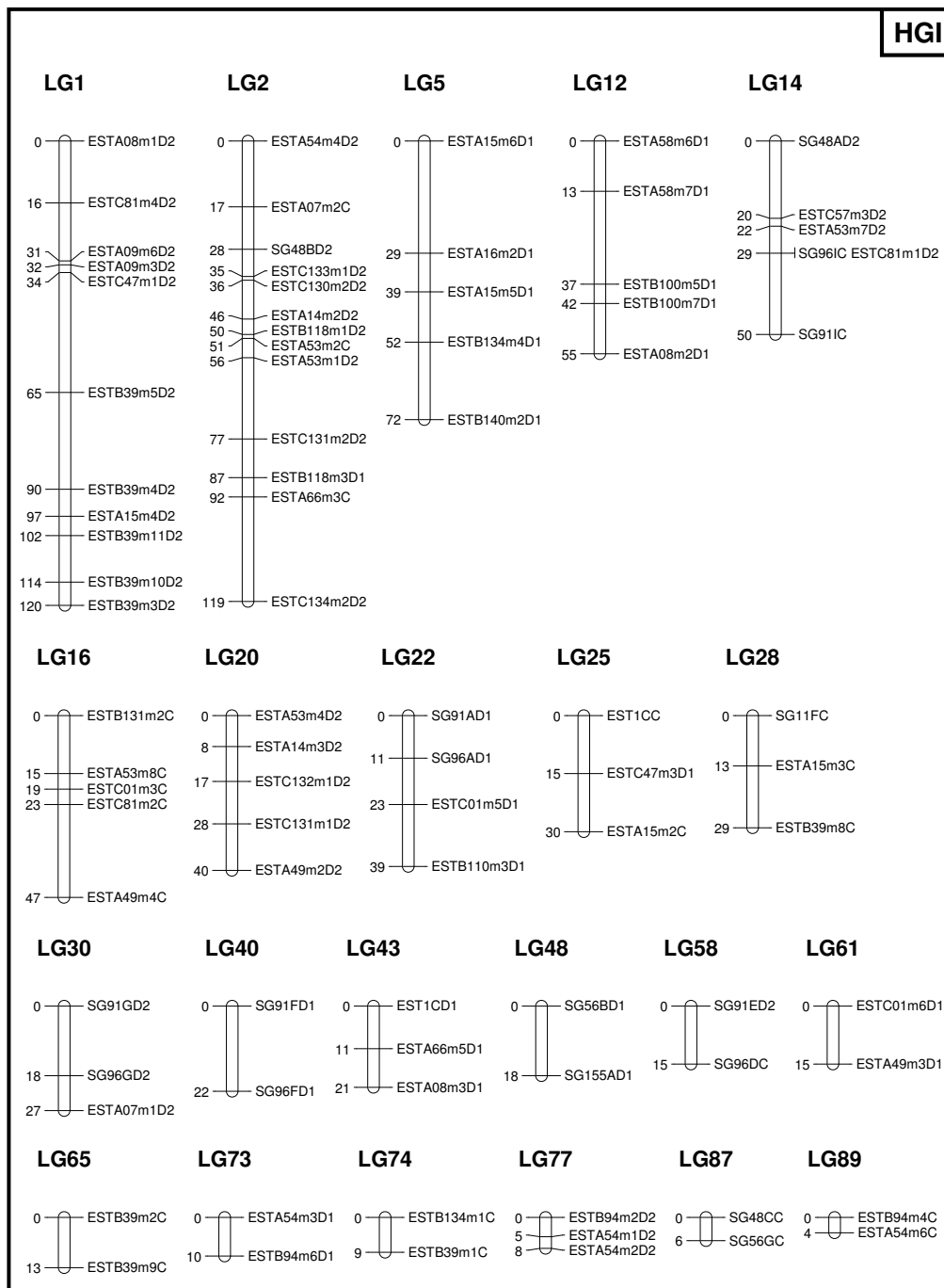


Figura 3 – Mapa genético integrado de um cruzamento comercial de cana-de-açúcar (SP80-180 × SP80-4966) baseado em 100 indivíduos. Distâncias no mapa são apresentadas em centimorgans (Kosambi). LG: do inglês *linkage group*. HG: do inglês *homology group*

(Continua)

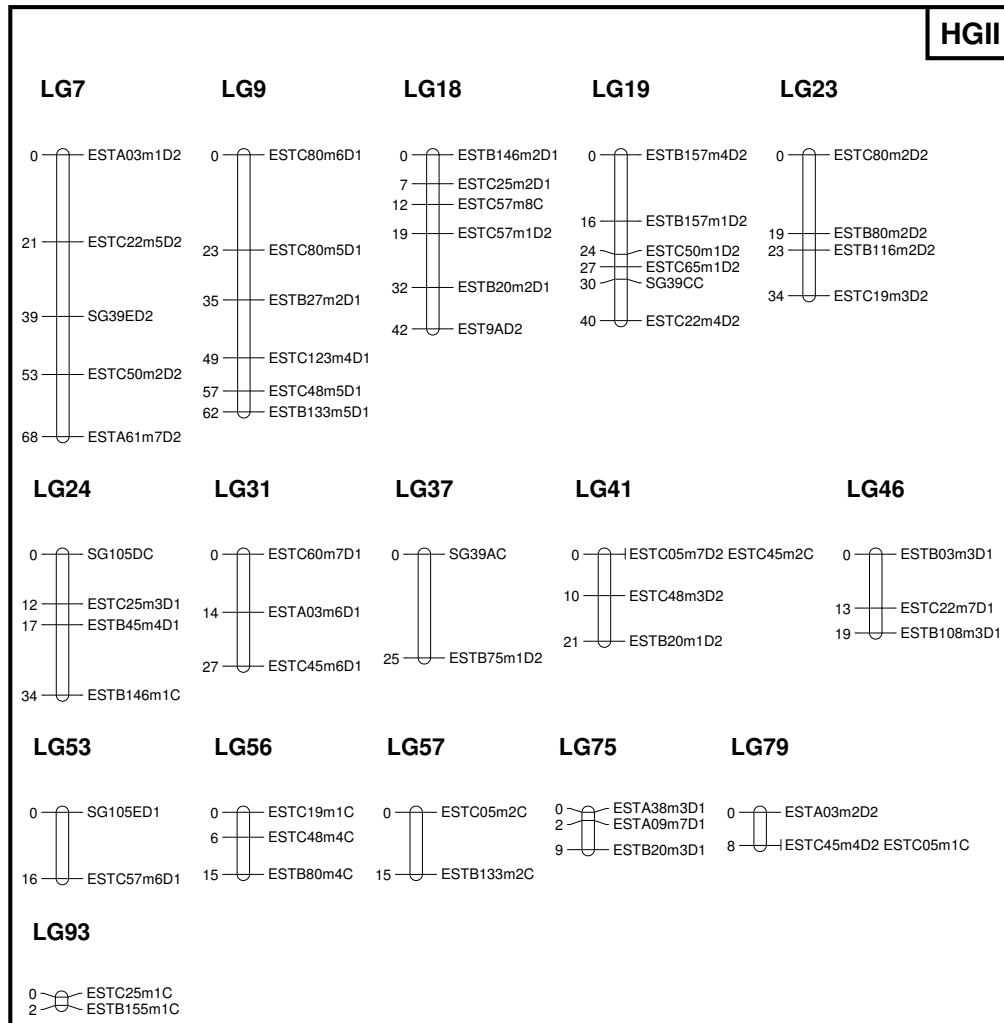


Figura 3 – Mapa genético integrado de um cruzamento comercial de cana-de-açúcar (SP80-180 × SP80-4966) baseado em 100 indivíduos. Distâncias no mapa são apresentadas em centimorgans (Kosambi). LG: do inglês *linkage group*. HG: do inglês *homology group*

(Continuação)

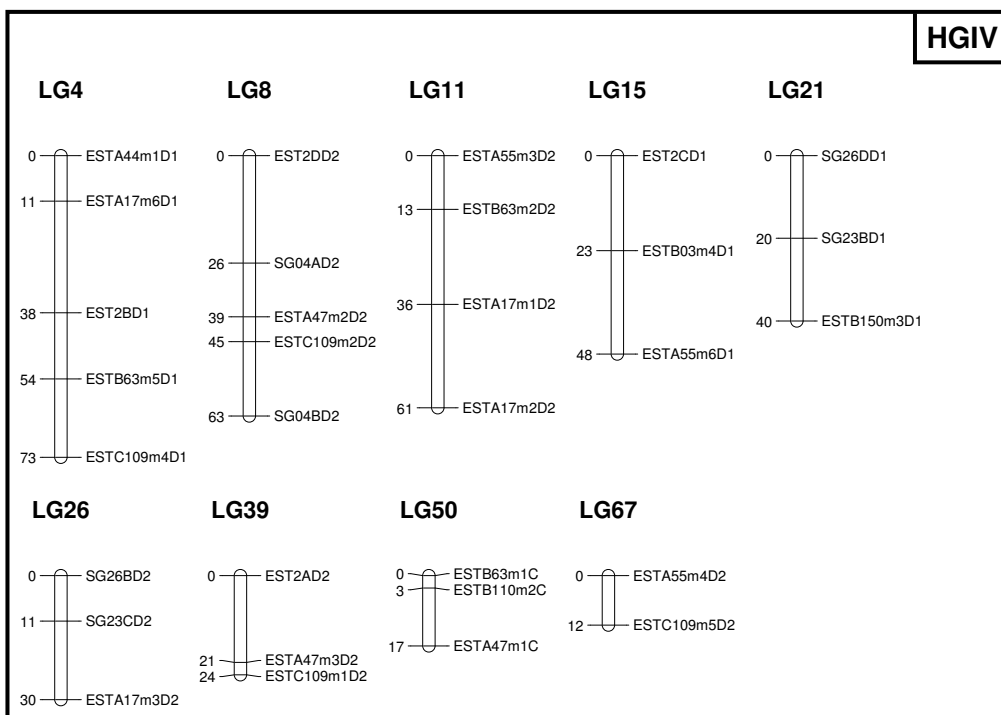
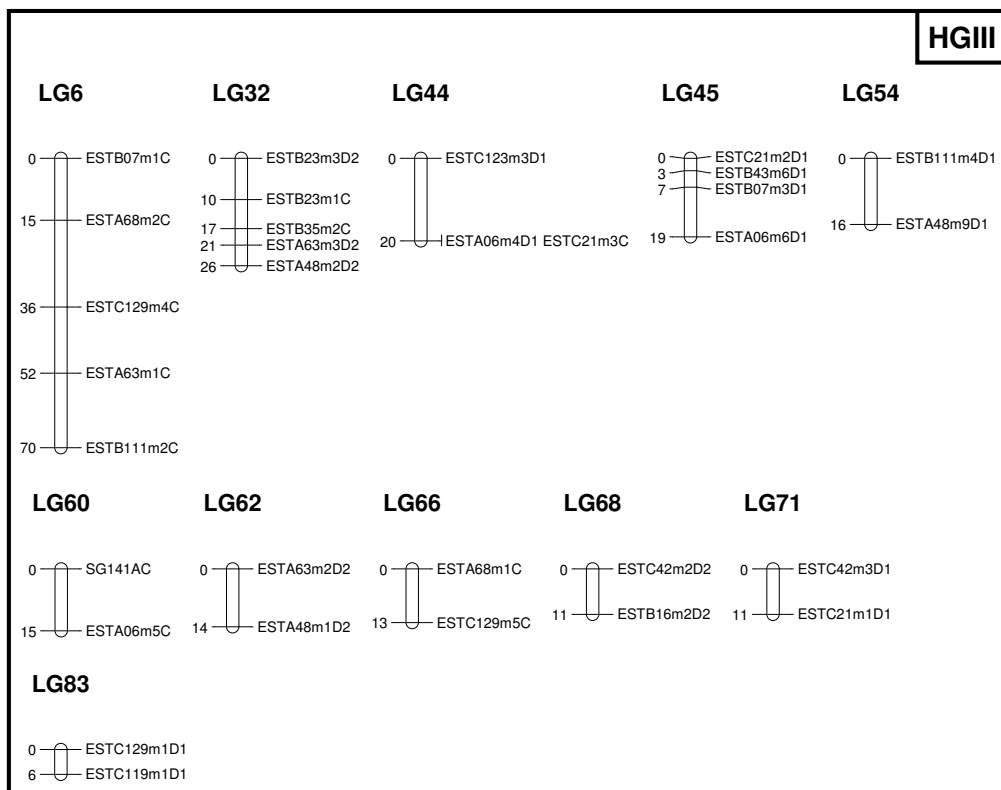


Figura 3 – Mapa genético integrado de um cruzamento comercial de cana-de-açúcar (SP80-180 × SP80-4966) baseado em 100 indivíduos. Distâncias no mapa são apresentadas em centimorgans (Kosambi). LG: do inglês *linkage group*. HG: do inglês *homology group*

(Continuação)

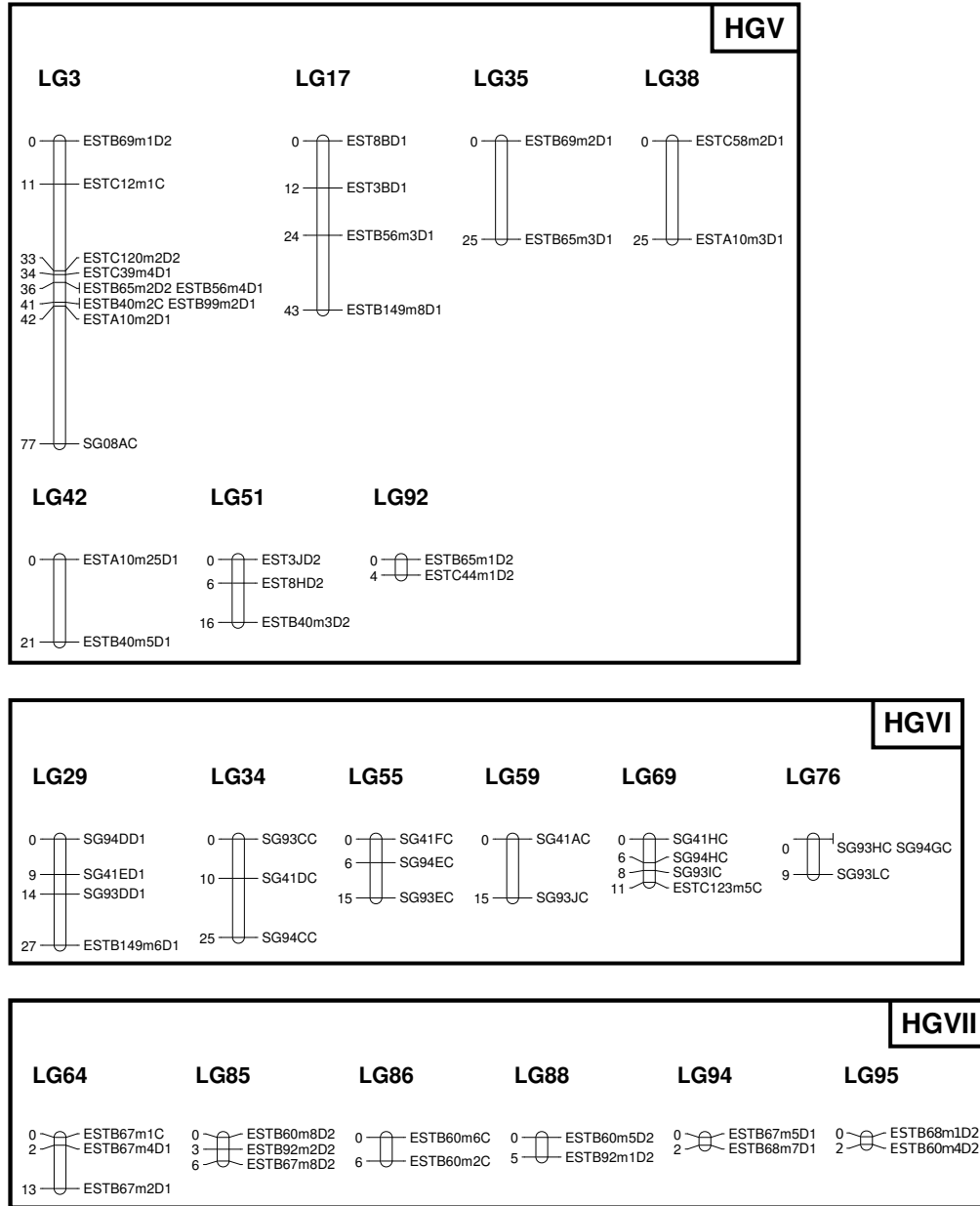


Figura 3 – Mapa genético integrado de um cruzamento comercial de cana-de-açúcar (SP80-180 × SP80-4966) baseado em 100 indivíduos. Distâncias no mapa são apresentadas em centimorgans (Kosambi). LG: do inglês *linkage group*. HG: do inglês *homology group*

(Continuação)



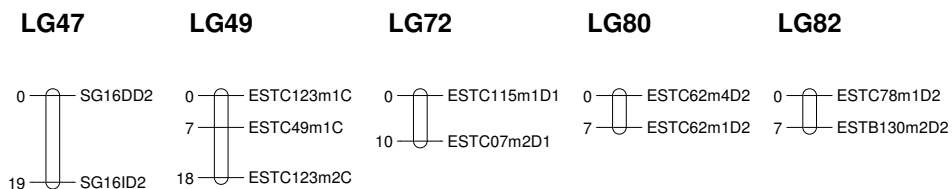
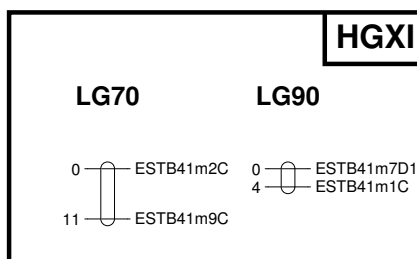
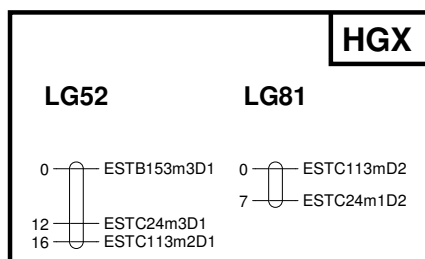
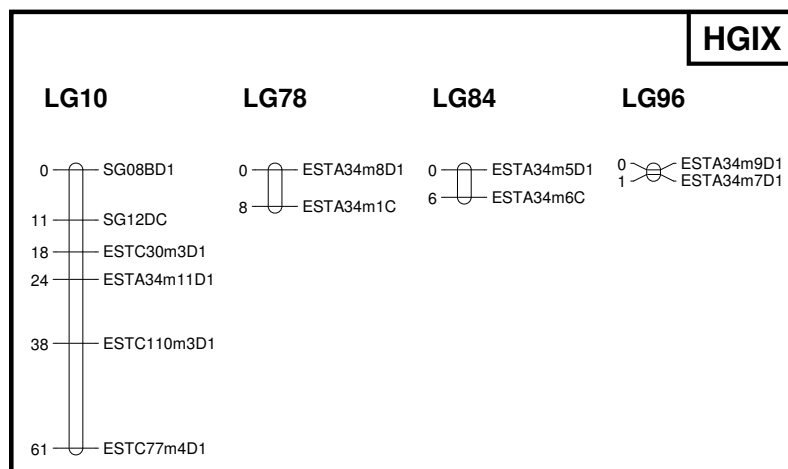
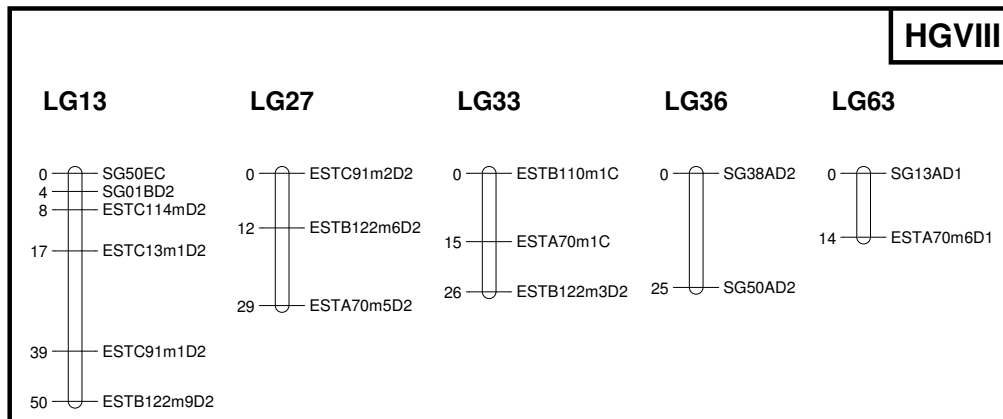


Figura 3 – Mapa genético integrado de um cruzamento comercial de cana-de-açúcar (SP80-180 × SP80-4966) baseado em 100 indivíduos. Distâncias no mapa são apresentadas em centiMorgans (Kosambi). LG: do inglês *linkage group*. HG: do inglês *homology group*

(Conclusão)

para caracteres e outra matriz para ambientes, de modo que a matriz  $G$  é um produto direto (de Kronecker) das duas matrizes componentes (SMITH et al., 2007; MALOSETTI et al., 2008).

O modelo (a) considera que os elementos na diagonal da matriz, isto é, as variâncias genéticas, são heterogêneos (DIAG), mas assume ausência de correlações genéticas entre “caracteres”. Já o modelo (b) permite que as variâncias genéticas sejam heterogêneas, enquanto considera uma covariância genética comum entre todos os “caracteres” ( $CS_{Het}$ ). O modelo (c), denominado fator analítico de primeira ordem (FA1), é um modelo multiplicativo que permite heterogeneidade de variâncias e covariâncias, ou seja, tenta aproximar o modelo completamente não-estruturado, porém com a vantagem de requerer um número menor de parâmetros (OMAN, 1991; GOGEL; CULLIS; VERBYLA, 1995; PIEPHO, 1997, 1998, 2000; EEUWIJK et al., 2001; SMITH; CULLIS; THOMPSON, 2001, 2005; MALOSETTI et al., 2004, 2006, 2008; BOER et al., 2007). Este último caso é exatamente o do modelo (d), o qual considera uma matriz de variâncias-covariâncias com elementos específicos para cada combinação caráter-ambiente (US).

Por fim, os modelos (e-n) consideram essas mesmas estruturas isoladamente para caracteres e ambientes, em diferentes combinações. No entanto, quando modelou-se ausência de correlação entre ambientes (modelo DIAG), os ajustes foram demasiadamente inadequados, conforme evidenciado por valores excessivamente elevados de AIC e BIC (resultados não mostrados). Assim, faz-se uma ressalva para o fato de que apenas um desses modelos é apresentado na Tabela 3, a saber: modelo (e), do tipo  $DIAG \otimes DIAG$ .

Uma inspeção da Tabela 3 mostra que, de acordo com o critério AIC, o modelo a ser selecionado foi o completamente não estruturado para combinações de caráter-ambiente (modelo d), já que este foi o modelo que apresentou menor valor desse critério. Já o BIC indicou que o melhor ajuste foi obtido com o modelo (n),  $US \otimes US$ . Tal resultado está de acordo com o esperado teoricamente, visto que o AIC selecionou o modelo em que a matriz  $G$  possuía maior número de parâmetros (171), enquanto que o BIC selecionou um modelo mais parsimonioso, com apenas 26 parâmetros estimados no total das duas matrizes  $G$  componentes (BURNHAM; ANDERSON, 2004). Devido à maior simplicidade do modelo, o que facilita a interpretação dos parâmetros estimados e resulta em análises de QTL mais rápidas e eficientes, decidiu-se escolher o modelo indicado pelo critério BIC nesta análise fenotípica para ser usado nas análises subsequentes. Em linhas gerais, nota-se que os modelos de menor BIC foram aqueles que consideraram uma matriz não-estruturada para variâncias-covariâncias genéticas entre ambientes, indicando que as variâncias e covariâncias entre

ambientes são suficientemente distintas para justificar a inclusão de um parâmetro individual para cada uma delas. Em essência, tal fato também foi sugerido pelo critério AIC, o qual adicionalmente sugere a inclusão de parâmetros individuais para cada combinação caráter-ambiente.

Tabela 3 – Diferentes modelos para a matriz de variâncias-covariâncias genéticas ( $M = C \times A$ , em que  $C$  é o número de caracteres e  $A$  é o número de ambientes) e valores correspondentes de AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano)

Matriz $G$	Modelo	$n_{PAR}$ em $G^a$	AIC	BIC
$G = G_{M \times M}^{caráter-amb}$	a) DIAG	18	10129,24	10223,02
	b) $CS_{Het}$	19	9956,91	10053,30
	c) FA1	36	9158,88	9299,56
	d) US	171	<b>7867,76</b>	8360,13
$G = G_{C \times C}^{caráter} \otimes G_{A \times A}^{amb}$	e) DIAG $\otimes$ DIAG	$(3 + 6) - 1 = 8$	10161,84	10229,57
	f) DIAG $\otimes$ $CS_{Het}$	$(3 + 7) - 1 = 9$	8294,01	8364,35
	g) DIAG $\otimes$ FA1	$(3 + 12) - 1 = 14$	8244,70	8328,06
	h) DIAG $\otimes$ US	$(3 + 21) - 1 = 23$	8086,46	8193,27
	i) $CS_{Het} \otimes CS_{Het}$	$(4 + 7) - 1 = 10$	8288,79	8361,73
	j) $CS_{Het} \otimes$ FA1	$(4 + 12) - 1 = 15$	8238,37	8324,34
	k) $CS_{Het} \otimes$ US	$(4 + 21) - 1 = 24$	8082,51	8191,92
	l) US $\otimes$ $CS_{Het}$	$(6 + 7) - 1 = 12$	8266,55	8344,70
	m) US $\otimes$ FA1	$(6 + 12) - 1 = 17$	8218,84	8310,02
	n) US $\otimes$ US	$(6 + 21) - 1 = 26$	8064,89	<b>8179,51</b>

$G$ : matriz de variâncias-covariâncias genéticas; DIAG: diagonal;  $CS_{Het}$ : simetria composta heterogênea; FA1: fator analítico de primeira ordem; e US: não-estruturado. Os menores valores de AIC e BIC são evidenciados em negrito.  
<sup>a</sup> O número total de parâmetros de variância-covariância no modelo é igual ao valor mostrado para a matriz  $G$  mais 18 do número de parâmetros na matriz  $R$

### 4.3 Análise de QTLs

Os resultados para o ajuste do modelo MIM final estão apresentados na Tabela 4. A ordem de apresentação dos QTLs é a mesma em que eles foram encontrados, usando a estratégia de mapeamento descrita na seção 3.2.4.1. No total, foram detectados 21 efeitos significativos, dos quais 13 corresponderam a efeitos principais e 8 foram interações epistáticas. Dos 13 efeitos principais, 8 foram detectados em grupos de ligação e 5 em marcadores não ligados, isto é, não alocados no mapa de ligação. É importante ressaltar que esses efeitos localizados em marcadores não ligados não equivalem a QTLs detectados pelo método de marcas individuais, visto que marcadores não alocados no mapa foram incluídos como parte integral do processo de busca do método MIM, ou

seja, como se fossem um grupo de ligação com apenas um marcador. De fato, é possível notar que 6 dos 8 efeitos epistáticos envolvem QTLs detectados nesses marcadores avulsos.

Um aspecto interessante é que todos os QTLs interagiram com locais e/ou cortes, de modo que não houve efeitos constantes em todas as condições ambientais consideradas. Do total de 21 efeitos, 3 interagiram apenas com local, mais especificamente o QTL XI, posicionado no grupo LG10; o QTL XII, associado ao marcador ESTB64m3C; e o efeito epistático entre os QTLs VII e XI. Nestes casos, os efeitos foram diferentes para os dois locais, mas exibiram o mesmo padrão nos três cortes. Por outro lado, 5 dos 21 efeitos exibiram interação significativa com corte, mas não com local, a saber: o QTL IV, localizado no grupo LG72; a epistasia entre os QTLs V e VI; o QTL VIII no grupo LG7; e os efeitos de interação entre os QTLs IX e XIII e entre III e IX. Para esses efeitos, houve um padrão semelhante de expressão nos dois locais, mas diferenças ao longo dos cortes. Por fim, os 13 efeitos restantes foram significativamente distintos ao longo das combinações de local e corte, isto é, a interação  $QTL \times L \times C$  foi significativa. Esses últimos QTLs incluem: QTL I (grupo LG92), QTL II (grupo LG55), QTL III (grupo LG11), QTL V (marcador ESTB45m6D1), QTL VI (marcador EST8CC), epistasia  $II \times V$ , QTL VII (grupo LG64), QTL IX (marcador ESTB40m1D2), QTL X (marcador ESTC19m4C), QTL XIII (grupo LG49) e interações epistáticas  $VI \times XII$ ,  $III \times VI$  e  $VII \times XIII$ . Tais resultados contrastam com os de Pastina (2010), em que poucas interações significativas com locais foram detectadas.

Valores em negrito na Tabela 4 indicam efeitos (em valor absoluto) maiores do que duas vezes o desvio-padrão correspondente, ou seja:  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ . Malosetti et al. (2008) propuseram o uso deste limiar como critério de significância para efeitos individuais, específicos para cada combinação entre caráter e ambiente. Isto auxilia na verificação do padrão de expressão de cada QTL, pois fornece uma ideia da magnitude dos efeitos dos QTLs para cada caráter, em cada local e cada corte. Não obstante, a significância de cada QTL foi calculada conjuntamente para todos os efeitos, através do teste de Wald, o que determinou sua manutenção ou remoção do modelo. De fato, nota-se que a última interação epistática (entre os QTLs III e IX) não exibe efeitos significativos de acordo com esse critério de Malosetti et al. (2008), apesar de o teste marginal de Wald ter sido significativo em nível de 5% de probabilidade.

Na Figura 4 são mostrados os perfis obtidos pelo método MIM da estatística  $-\log_{10}(p\text{-valor})$ , para os grupos de ligação em que os efeitos significativos foram detectados. É importante observar que esses gráficos são utilizados no método MIM apenas para exibir o padrão de significância de

Tabela 4 – Estimativas dos efeitos dos QTLs sob o modelo MIM final e seus desvios-padrão correspondentes entre parênteses (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare;  $\alpha_{p_{cjk}}$ ,  $\alpha_{q_{cjk}}$  e  $\beta_{(\tau_1\tau_2)_{cjk}}$  são os efeitos aditivos dos genitores  $P$  e  $Q$  e a interação epistática, respectivamente, específicos para cada combinação caráter ( $c$ ), local ( $j$ ) e corte ( $k$ ). Valores em negrito são significativos pelo critério:  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ )

(Continua)

QTL	LG (efeito)	Marcadores	Posição (cM)	Caráter	Local	Corte			
						1	2	3	
I	92 ( $\alpha_{q_{cjk}}$ )	ESTB65m1D2/ ESTC44m1D2	4,0	Fiber	1	<b>-0,250</b> (0,118)	<b>-0,401</b> (0,130)	<b>-0,373</b> (0,137)	
					2	<b>-0,356</b> (0,109)	<b>-0,280</b> (0,127)	-0,243 (0,128)	
					POL	1	0,067 (0,126)	0,009 (0,129)	-0,041 (0,134)
						2	0,046 (0,116)	0,238 (0,127)	<b>0,478</b> (0,150)
					TCH	1	<b>-3,988</b> (1,801)	<b>-5,527</b> (1,986)	<b>-7,308</b> (2,069)
						2	-1,571 (1,680)	-2,296 (1,886)	-3,145 (1,919)
II	55 ( $\alpha_{p_{cjk}}$ )	SG41FC	0,0	Fiber	1	-0,215 (0,240)	-0,358 (0,265)	-0,381 (0,279)	
					2	<b>-0,615</b> (0,223)	-0,300 (0,259)	-0,487 (0,261)	
					POL	1	0,431 (0,255)	0,415 (0,262)	<b>0,579</b> (0,273)
						2	0,470 (0,236)	<b>0,849</b> (0,258)	<b>1,191</b> (0,303)
					TCH	1	-0,313 (3,654)	-0,276 (4,033)	0,607 (4,205)
						2	-2,899 (3,429)	-1,999 (3,846)	-1,801 (3,926)
III	11 ( $\alpha_{q_{cjk}}$ )	ESTA55m3D2	0,0	Fiber	1	0,024 (0,153)	0,018 (0,169)	0,117 (0,178)	
					2	0,098 (0,142)	-0,130 (0,165)	-0,153 (0,166)	
					POL	1	-0,080 (0,163)	0,122 (0,166)	-0,026 (0,174)
						2	0,142 (0,150)	-0,086 (0,164)	<b>-0,428</b> (0,192)
					TCH	1	<b>9,596</b> (2,331)	<b>12,364</b> (2,567)	<b>13,198</b> (2,673)
						2	<b>7,229</b> (2,182)	<b>6,750</b> (2,442)	<b>5,076</b> (2,489)
IV	72 ( $\alpha_{p_{cjk}}$ )	ESTC115m1D1	0,0	Fiber	1	<b>0,247</b> (0,103)	<b>0,247</b> (0,114)	<b>0,254</b> (0,120)	
					2	<b>0,247</b> (0,103)	<b>0,247</b> (0,114)	<b>0,254</b> (0,120)	
					POL	1	-0,067 (0,107)	-0,073 (0,115)	0,018 (0,123)
						2	-0,067 (0,107)	-0,073 (0,115)	0,018 (0,123)
					TCH	1	-0,225 (1,575)	2,978 (1,724)	3,327 (1,803)
						2	-0,225 (1,575)	2,978 (1,724)	3,327 (1,803)
V	NL ( $\alpha_{p_{cjk}}$ )	ESTB45m6D1		Fiber	1	<b>-0,312</b> (0,153)	-0,336 (0,170)	-0,306 (0,178)	
					2	<b>-0,432</b> (0,146)	<b>-0,414</b> (0,167)	<b>-0,393</b> (0,171)	
					POL	1	0,142 (0,162)	0,163 (0,168)	0,001 (0,178)
						2	0,054 (0,153)	0,205 (0,167)	0,330 (0,190)
					TCH	1	<b>-4,895</b> (2,337)	<b>-8,138</b> (2,574)	<b>-8,278</b> (2,687)
						2	-3,836 (2,242)	<b>-6,317</b> (2,494)	<b>-5,866</b> (2,568)
VI	NL ( $\alpha_{p_{cjk}}$ )	EST8CC		Fiber	1	0,026 (0,242)	0,090 (0,264)	0,291 (0,276)	
					2	0,079 (0,225)	0,175 (0,257)	0,226 (0,258)	
					POL	1	-0,237 (0,256)	-0,074 (0,261)	-0,274 (0,272)
						2	0,182 (0,236)	<b>-0,630</b> (0,256)	<b>-1,206</b> (0,296)
					TCH	1	6,769 (3,679)	5,669 (4,011)	3,342 (4,165)
						2	<b>7,454</b> (3,450)	5,782 (3,817)	3,683 (3,888)

Tabela 4 – Estimativas dos efeitos dos QTLs sob o modelo MIM final e seus desvios-padrão correspondentes entre parênteses (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare;  $\alpha_{pcjk}$ ,  $\alpha_{qck}$  e  $\beta_{(\tau_1\tau_2)ck}$  são os efeitos aditivos dos genitores  $P$  e  $Q$  e a interação epistática, respectivamente, específicos para cada combinação caráter ( $c$ ), local ( $j$ ) e corte ( $k$ ). Valores em negrito são significativos pelo critério:  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ )

(Continuação)

QTL	LG (efeito)	Marcadores	Posição (cM)	Caráter	Local	Corte		
						1	2	3
II	$\beta_{(\alpha_p\alpha_p)ck}$			Fiber	1	-0,376 (0,246)	<b>-0,758</b> (0,271)	<b>-0,649</b> (0,285)
					2	-0,360 (0,229)	-0,450 (0,264)	-0,340 (0,267)
V				POL	1	<b>0,525</b> (0,261)	0,227 (0,268)	0,051 (0,280)
					2	0,126 (0,242)	0,313 (0,264)	-0,131 (0,307)
				TCH	1	<b>-7,833</b> (3,752)	<b>-12,658</b> (4,121)	<b>-13,899</b> (4,290)
					2	-2,498 (3,524)	-6,484 (3,930)	<b>-8,272</b> (4,010)
V	$\beta_{(\alpha_p\alpha_p)ck}$			Fiber	1	0,207 (0,152)	0,284 (0,168)	0,188 (0,177)
					2	0,207 (0,152)	0,284 (0,168)	0,188 (0,177)
VI				POL	1	<b>-0,363</b> (0,157)	-0,271 (0,169)	-0,213 (0,182)
					2	<b>-0,363</b> (0,157)	-0,271 (0,169)	-0,213 (0,182)
				TCH	1	4,065 (2,311)	<b>9,195</b> (2,542)	<b>9,519</b> (2,663)
					2	4,065 (2,311)	<b>9,195</b> (2,542)	<b>9,519</b> (2,663)
VII	$64 (\alpha_{pcjk})$	ESTB67m4D1	1,7	Fiber	1	-0,113 (0,121)	-0,230 (0,133)	-0,098 (0,139)
					2	-0,045 (0,112)	-0,074 (0,129)	-0,011 (0,130)
				POL	1	<b>0,393</b> (0,129)	0,154 (0,131)	0,108 (0,137)
					2	<b>0,301</b> (0,119)	0,181 (0,129)	<b>0,333</b> (0,150)
				TCH	1	-1,087 (1,848)	-2,841 (2,021)	-2,311 (2,100)
					2	-1,195 (1,728)	-2,640 (1,919)	-1,726 (1,954)
VIII	$7 (\alpha_{qck})$	ESTA61m7D2	68,0	Fiber	1	<b>0,298</b> (0,102)	<b>0,299</b> (0,113)	<b>0,299</b> (0,119)
					2	<b>0,298</b> (0,102)	<b>0,299</b> (0,113)	<b>0,299</b> (0,119)
				POL	1	<b>-0,351</b> (0,106)	-0,191 (0,114)	-0,035 (0,122)
					2	<b>-0,351</b> (0,106)	-0,191 (0,114)	-0,035 (0,122)
				TCH	1	2,893 (1,559)	<b>5,723</b> (1,711)	<b>6,146</b> (1,792)
					2	2,893 (1,559)	<b>5,723</b> (1,711)	<b>6,146</b> (1,792)
IX	NL ( $\alpha_{qck}$ )	ESTB40m1D2		Fiber	1	0,241 (0,132)	<b>0,333</b> (0,145)	0,151 (0,153)
					2	0,226 (0,123)	0,185 (0,142)	0,166 (0,144)
				POL	1	<b>-0,340</b> (0,139)	-0,133 (0,144)	0,003 (0,150)
					2	0,004 (0,130)	-0,079 (0,142)	<b>-0,362</b> (0,165)
				TCH	1	<b>7,410</b> (2,005)	<b>10,740</b> (2,210)	<b>10,649</b> (2,304)
					2	<b>4,674</b> (1,892)	<b>8,540</b> (2,116)	<b>7,683</b> (2,165)
X	NL ( $\alpha_{pcjk}$ )	ESTC19m4C		Fiber	1	0,182 (0,148)	<b>0,352</b> (0,164)	0,329 (0,172)
					2	<b>0,357</b> (0,137)	<b>0,329</b> (0,160)	0,251 (0,161)
				POL	1	<b>-0,556</b> (0,157)	-0,267 (0,161)	-0,166 (0,168)
					2	-0,160 (0,145)	<b>-0,332</b> (0,159)	<b>-0,722</b> (0,188)
				TCH	1	4,404 (2,256)	<b>7,663</b> (2,490)	<b>8,622</b> (2,595)
					2	<b>5,277</b> (2,107)	<b>5,344</b> (2,366)	3,757 (2,411)

Tabela 4 – Estimativas dos efeitos dos QTLs sob o modelo MIM final e seus desvios-padrão correspondentes entre parênteses (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare;  $\alpha_{p_{cjk}}$ ,  $\alpha_{q_{cjk}}$  e  $\beta_{(\tau_1\tau_2)_{cjk}}$  são os efeitos aditivos dos genitores  $P$  e  $Q$  e a interação epistática, respectivamente, específicos para cada combinação caráter ( $c$ ), local ( $j$ ) e corte ( $k$ ). Valores em negrito são significativos pelo critério:  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ )

(Continuação)

QTL	LG (efeito)	Marcadores	Posição (cM)	Caráter	Local	Corte		
						1	2	3
XI	$10 (\alpha_{p_{c_j}}$ )	ESTA34m11D1/ ESTC110m3D1	36,0	Fiber	1	<b>0,372</b> (0,120)	<b>0,372</b> (0,120)	<b>0,372</b> (0,120)
					2	<b>0,342</b> (0,111)	<b>0,342</b> (0,111)	<b>0,342</b> (0,111)
				POL	1	<b>-0,391</b> (0,123)	<b>-0,391</b> (0,123)	<b>-0,391</b> (0,123)
					2	<b>-0,355</b> (0,114)	<b>-0,355</b> (0,114)	<b>-0,355</b> (0,114)
				TCH	1	1,646 (1,810)	1,646 (1,810)	1,646 (1,810)
					2	<b>3,595</b> (1,686)	<b>3,595</b> (1,686)	<b>3,595</b> (1,686)
XII	$NL (\alpha_{p_{c_j}}$ )	ESTB64m3C		Fiber	1	-0,443 (0,227)	-0,443 (0,227)	-0,443 (0,227)
					2	<b>-0,461</b> (0,210)	<b>-0,461</b> (0,210)	<b>-0,461</b> (0,210)
				POL	1	-0,234 (0,234)	-0,234 (0,234)	-0,234 (0,234)
					2	0,034 (0,216)	0,034 (0,216)	0,034 (0,216)
				TCH	1	4,877 (3,442)	4,877 (3,442)	4,877 (3,442)
					2	1,448 (3,201)	1,448 (3,201)	1,448 (3,201)
XIII	$49 (\alpha_{p_{c_{jk}}})$	ESTC123m1C/ ESTC49m1C	1,0	Fiber	1	-0,099 (0,245)	-0,333 (0,271)	-0,404 (0,285)
					2	-0,340 (0,227)	-0,169 (0,265)	-0,334 (0,267)
				POL	1	<b>0,813</b> (0,260)	0,402 (0,267)	0,301 (0,279)
					2	0,076 (0,240)	<b>0,621</b> (0,264)	0,096 (0,310)
				TCH	1	0,857 (3,730)	2,064 (4,118)	0,969 (4,294)
					2	3,303 (3,496)	4,996 (3,924)	0,127 (4,003)
VI × XII	$\beta_{(\alpha_p\alpha_p)_{cjk}}$			Fiber	1	0,021 (0,231)	0,125 (0,249)	-0,117 (0,260)
					2	0,000 (0,214)	0,008 (0,241)	-0,077 (0,242)
				POL	1	0,058 (0,244)	0,100 (0,248)	0,234 (0,257)
					2	0,092 (0,225)	<b>0,541</b> (0,241)	0,482 (0,276)
				TCH	1	-2,255 (3,520)	-1,100 (3,793)	-1,338 (3,919)
					2	-6,425 (3,287)	-4,104 (3,591)	-2,291 (3,644)
IX × XIII	$\beta_{(\alpha_q\alpha_p)_{c_k}}$			Fiber	1	-0,308 (0,240)	-0,242 (0,266)	-0,126 (0,279)
					2	-0,308 (0,240)	-0,242 (0,266)	-0,126 (0,279)
				POL	1	-0,493 (0,248)	-0,436 (0,267)	-0,050 (0,287)
					2	-0,493 (0,248)	-0,436 (0,267)	-0,050 (0,287)
				TCH	1	<b>-7,952</b> (3,654)	<b>-10,809</b> (4,013)	<b>-10,957</b> (4,202)
					2	<b>-7,952</b> (3,654)	<b>-10,809</b> (4,013)	<b>-10,957</b> (4,202)
VII × XI	$\beta_{(\alpha_p\alpha_p)_{c_j}}$			Fiber	1	-0,191 (0,119)	-0,191 (0,119)	-0,191 (0,119)
					2	<b>-0,258</b> (0,110)	<b>-0,258</b> (0,110)	<b>-0,258</b> (0,110)
				POL	1	0,219 (0,123)	0,219 (0,123)	0,219 (0,123)
					2	0,152 (0,113)	0,152 (0,113)	0,152 (0,113)
				TCH	1	<b>-4,857</b> (1,807)	<b>-4,857</b> (1,807)	<b>-4,857</b> (1,807)
					2	-2,311 (1,676)	-2,311 (1,676)	-2,311 (1,676)

Tabela 4 – Estimativas dos efeitos dos QTLs sob o modelo MIM final e seus desvios-padrão correspondentes entre parênteses (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare;  $\alpha_{p_{cjk}}$ ,  $\alpha_{q_{cjk}}$  e  $\beta_{(\tau_1\tau_2)_{cjk}}$  são os efeitos aditivos dos genitores  $P$  e  $Q$  e a interação epistática, respectivamente, específicos para cada combinação caráter ( $c$ ), local ( $j$ ) e corte ( $k$ ). Valores em negrito são significativos pelo critério:  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$  (Conclusão)

QTL	LG (efeito)	Marcadores	Posição (cM)	Caráter	Local	Corte		
						1	2	3
III ×	$\beta_{(\alpha_q\alpha_p)_{cjk}}$			Fiber	1	-0,072 (0,149)	-0,128 (0,164)	-0,173 (0,173)
					2	-0,193 (0,139)	-0,070 (0,160)	0,008 (0,162)
VI				POL	1	-0,182 (0,158)	-0,100 (0,162)	-0,053 (0,169)
					2	-0,228 (0,146)	-0,021 (0,160)	0,227 (0,187)
				TCH	1	<b>-5,434</b> (2,269)	<b>-6,908</b> (2,499)	<b>-8,239</b> (2,603)
					2	<b>-4,730</b> (2,129)	-3,964 (2,382)	-2,960 (2,431)
VII ×	$\beta_{(\alpha_p\alpha_p)_{cjk}}$			Fiber	1	-0,481 (0,249)	<b>-0,641</b> (0,276)	<b>-0,617</b> (0,290)
					2	-0,437 (0,232)	-0,461 (0,269)	-0,358 (0,271)
XIII				POL	1	-0,121 (0,265)	-0,038 (0,272)	0,091 (0,284)
					2	-0,425 (0,245)	0,073 (0,268)	<b>0,720</b> (0,314)
				TCH	1	-3,375 (3,802)	-4,244 (4,192)	-1,648 (4,368)
					2	0,256 (3,566)	-3,229 (3,995)	-1,813 (4,077)
III ×	$\beta_{(\alpha_q\alpha_q)_{ck}}$			Fiber	1	-0,003 (0,109)	0,070 (0,120)	0,152 (0,126)
					2	-0,003 (0,109)	0,070 (0,120)	0,152 (0,126)
IX				POL	1	-0,149 (0,112)	0,054 (0,121)	0,084 (0,129)
					2	-0,149 (0,112)	0,054 (0,121)	0,084 (0,129)
				TCH	1	-1,143 (1,652)	-0,152 (1,810)	-0,367 (1,894)
					2	-1,143 (1,652)	-0,152 (1,810)	-0,367 (1,894)

LG: grupo de ligação; NL: marcador não-ligado.



cada QTL ao longo do grupo de ligação correspondente, de modo que eles foram obtidos apenas na etapa final de refinamento de posição. Tal aplicação é marcadamente distinta do que é feito nos métodos IM e CIM, em que os gráficos de perfil de verossimilhança são efetivamente utilizados para declarar um possível QTL como significativo ou não (MING et al., 2002a; AL-JANABI et al., 2007).

A Figura 4 também exibe os perfis de  $-\log_{10}(p\text{-valor})$  para cada caráter analisado individualmente. Uma análise visual desses perfis pode fornecer evidências em favor da hipótese de pleiotropia, quando os pontos de máximo para os vários caracteres ocorrem na mesma posição, ou em prol da hipótese de QTLs ligados, quando os picos estão distantes entre si. Para todos os efeitos aqui detectados, não houve evidências suficientes em favor da hipótese de QTLs ligados, seja porque os picos para todos os caracteres ocorreram em posições muito próximas (como para o QTL XI, no grupo LG10), porque o pico foi significativo apenas para um caráter (como observado para o QTL XIII, localizado no grupo LG49) ou porque o grupo de ligação correspondente é muito pequeno (como no caso do QTL I, grupo LG92). De qualquer forma, no contexto dos modelos mistos conforme empregados neste trabalho, não há um teste formal que permita a comparação do modelo com apenas um QTL pleiotrópico contra o modelo com mais de um QTL, em que cada um controla apenas um caráter, pois os QTLs foram incluídos como efeitos fixos e os modelos não são aninhados. Caso os preditores genéticos sejam incluídos como efeitos aleatórios, os critérios de seleção AIC e BIC podem ser utilizados para comparação desses dois modelos, pois neste caso a parte fixa de ambos é a mesma (EEUWIJK, comunicação pessoal).

As posições dos QTLs nos respectivos grupos de ligação, bem como as associações significativas com marcas não ligadas, são mostradas na Figura 5. Esta figura apresenta os caracteres influenciados por cada QTL, segundo o critério de Malosetti et al. (2008), além dos efeitos específicos para cada um dos dois locais e três cortes, sendo esses efeitos separados em duas linhas e três colunas, respectivamente. Efeitos significativos são indicados por um sinal de positivo ou negativo, caso a presença do alelo aumente ou reduza o valor do caráter quantitativo, respectivamente. O símbolo “0” representa um efeito não significativo. De fato, trata-se de uma representação gráfica dos resultados numéricos apresentados na Tabela 4. Não obstante, facilita a observação de alguns pontos interessantes. Primeiramente, fica evidente que nove dos treze QTLs com efeitos principais são pleiotrópicos e sempre afetam a porcentagem de fibras e a produção de cana no mesmo sentido, enquanto que o teor de sacarose é influenciado no sentido oposto. É também possível observar que

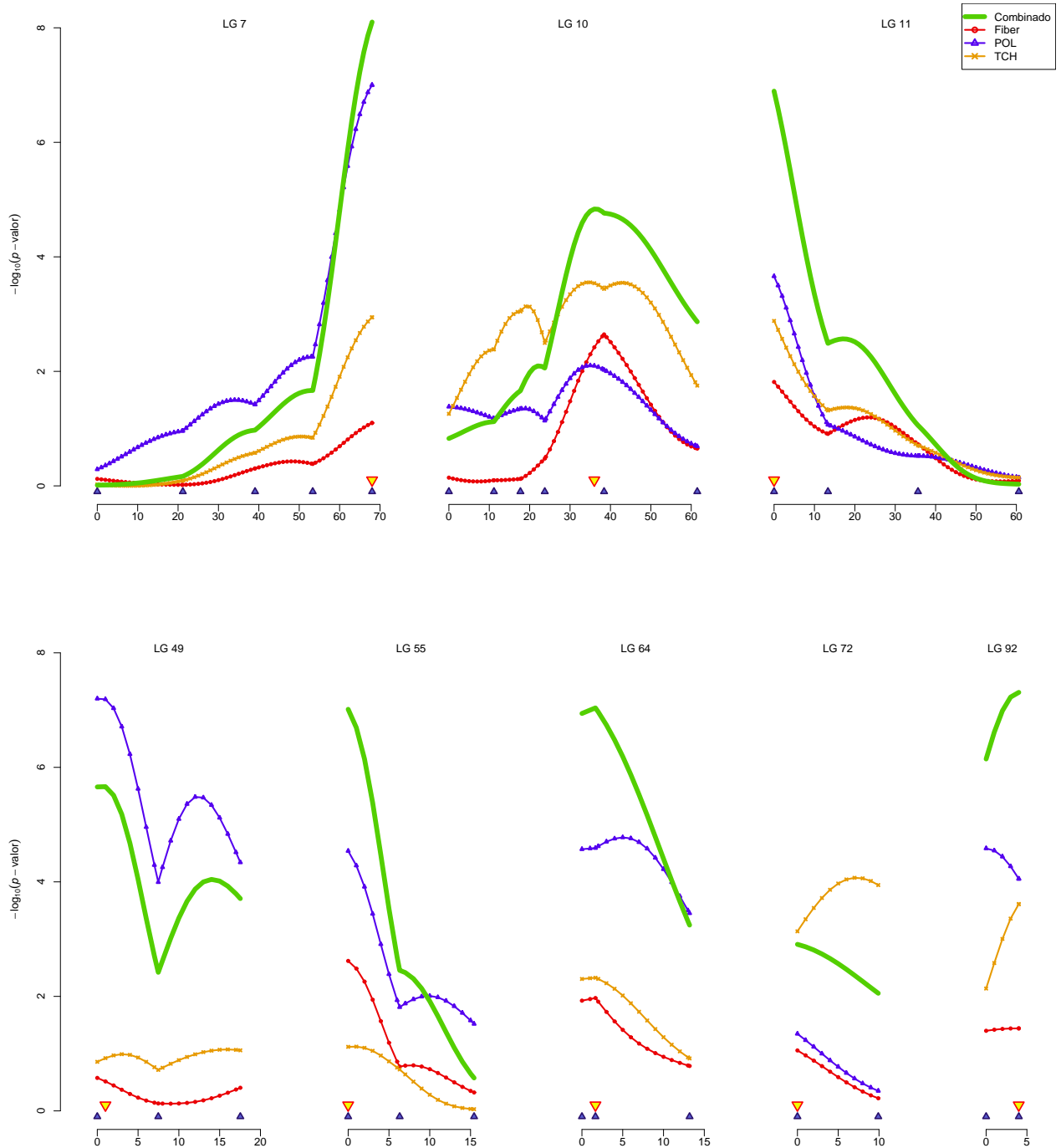


Figura 4 – Resultados do método MIM indicando as posições dos QTLs (triângulos amarelos) e os perfis de  $-\log_{10}(p\text{-valor})$  ao longo dos grupos de ligação (LG) para a análise conjunta dos três caracteres e para cada caráter separadamente (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare. Triângulos azuis: posições dos marcadores moleculares no mapa de ligação. Valores de distância em cM utilizando a função de mapeamento de Kosambi)

os QTLs dos grupos de ligação LG7 e LG72 apresentam apenas uma linha de efeitos para os respectivos caracteres, pois não exibem interação com locais. Analogamente, o QTL do grupo LG10 e aquele associado à marca ESTB64m3C não interagem com cortes, o que é refletido pela existência

de apenas uma coluna de efeitos para cada caráter. Por fim, nota-se também uma aparente falta de estabilidade de expressão dos locos, visto que muitas vezes os efeitos de um QTL são observados apenas em um dado local e corte.

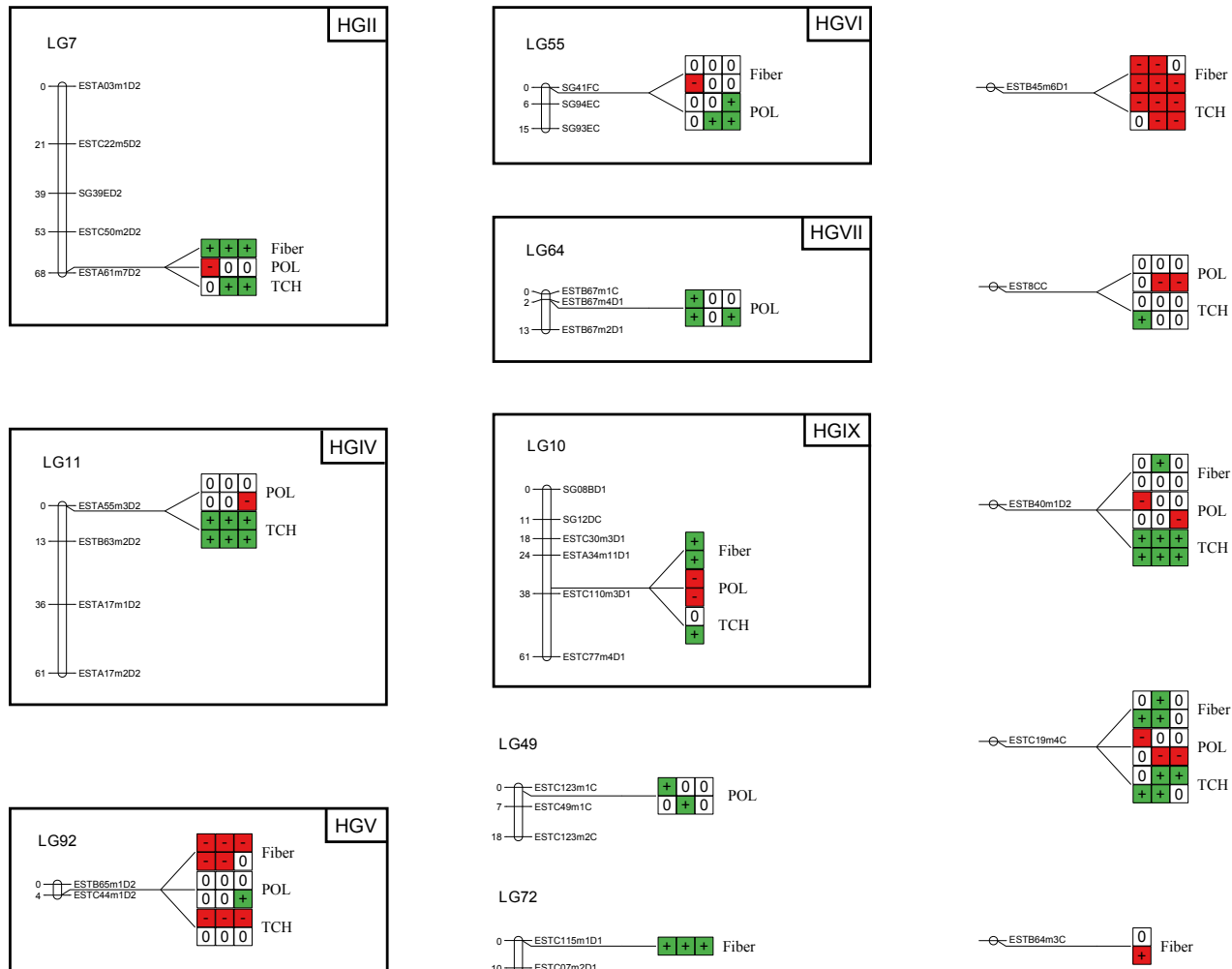


Figura 5 – Grupos de ligação com QTLs detectados e efeitos significativos pelo critério:  $|\text{efeito}| \geq 2 \times$  desvio-padrão. Duas linhas e/ou três colunas de efeitos para cada caráter indicam efeitos distintos para locais e/ou cortes, respectivamente. Efeitos significativos são indicados por cor de fundo verde ou vermelha, caso a presença do alelo aumente ou diminua a expressão do caráter, respectivamente (Fiber: porcentagem de fibra; POL: conteúdo de sacarose; TCH: toneladas de cana por hectare. Valores de distância em cM utilizando a função de mapeamento de Kosambi)

## 5 DISCUSSÃO

Programas de melhoramento genético usualmente fazem uso de dados coletados em muitas localidades, ao longo de vários anos, e diversos caracteres são mensurados simultaneamente. Portanto, naturalmente espera-se que as variabilidades genética e residual sejam diferentes nas condições ambientais em estudo, assim como espera-se que haja correlações entre caracteres e entre ambientes, tanto de origem genética quanto residual. Tal estrutura intrínseca dos dados automaticamente torna-os propícios a uma análise de modelos mistos, os quais permitem o ajuste de matrizes de variância-covariância adequadas a cada situação. Esta prática, por sua vez, reduz a probabilidade de erro tipo I, isto é, de declaração de falsos positivos, tornando os resultados mais confiáveis (PIEPHO, 2005).

No presente estudo, o modelo selecionado explorou o produto direto de matrizes de covariância separadas para caracteres e ambientes, sendo ambas não estruturadas (SMITH et al., 2007). Malosetti et al. (2008) relataram que, para um conjunto de dados MTME de milho, o uso de produtos diretos de matrizes para construir matrizes de variância e covariância resultou em bom ajuste aos dados, ao mesmo tempo em que reduziu consideravelmente o número de parâmetros. Por outro lado, Malosetti et al. (2006; 2008) argumentam que o uso de modelos não estruturados pode ser problemático quando o número de caracteres e/ou ambientes torna-se apenas moderadamente elevado, devido a dificuldades numéricas no ajuste do modelo. Neste trabalho, foram estudados apenas três caracteres e seis ambientes, de modo que o número de parâmetros não tornou proibitivas as análises de busca por QTLs. Não obstante, o produto de uma matriz não estruturada para caracteres por uma matriz não estruturada para ambientes reduziu o número de parâmetros de 171, referente ao modelo não estruturado para combinações caracteres  $\times$  ambientes, para apenas 26 (Tabela 3). Assim, nota-se que a estratégia proposta por Smith et al. (2007) melhorou o ajuste do modelo fenotípico aos dados e facilitou a realização das etapas subsequentes.

Euwijk et al. (2001) afirmam que o modelo não estruturado pode ser apropriado quando há diferenças no conjunto de genes responsáveis pela performance em cada ambiente. Segundo esses autores, um conjunto comum de genes é expresso até certo grau em todos os ambientes, o que faz com que as correlações genéticas sejam de moderadas a altamente positivas. Além desses genes envolvidos nessa resposta básica, conjuntos de genes particulares para cada ambiente podem ser necessários para adaptações específicas, ao passo que diferentes formas alélicas podem ser melhor

ajustadas às distintas condições ambientais e, assim, dar origem a correlações genéticas negativas. Notadamente, este raciocínio também é válido para correlações entre caracteres (EEUWIJK et al., 2001).

A decisão de considerar cada efeito do modelo como fixo ou aleatório depende da natureza das observações, dos objetivos do trabalho e até mesmo da preferência dos pesquisadores (BOER et al., 2007). Como os 100 genótipos utilizados para o mapeamento de QTLs foram amostrados de uma progênie segregante oriunda de um cruzamento biparental, não havia interesse específico em cada um deles, mas sim na estimação da variação genética como um todo. Por esta razão, o efeito de genótipos foi aqui considerado como aleatório. Por outro lado, não havia o objetivo de estudar o comportamento médio desses genótipos ao longo de diversos ambientes, de modo que as duas cidades e os três anos agrícolas utilizados para avaliação da população segregante não necessitariam representar uma amostra maior de ambientes. Então, os fatores local e corte foram considerados como efeitos fixos. Claramente, essa decisão estende-se aos efeitos dos marcadores moleculares, ou seja, dos preditores genéticos.

A literatura de mapeamento de QTLs no contexto de modelos mistos mostra que diferentes autores tomam QTLs como fixos ou aleatórios em função dos objetivos do estudo (PIEPHO, 2000; VERBYLA et al., 2003; BOER et al., 2007). Marcadores moleculares incluídos como efeitos aleatórios geralmente têm como objetivo simplesmente controlar a variação genética residual, conforme feito por Wang et al. (1999), ou são utilizados para modelar a interação  $QTL \times E$ , como empregado por Piepho et al. (2000) e Verbyla et al. (2003). Já preditores genéticos considerados como termos fixos representam posições de interesse específico no genoma e cujos efeitos deseja-se estimar, conforme utilizados por Piepho (2000), Malosetti et al. (2004) e Boer et al. (2007), entre outros. O presente estudo considerou QTLs como termos fixos, o que, aliado ao fato de locais e cortes também serem fixos, permitiu a estimação de efeitos de QTLs específicos para cada local e corte. Tal abordagem é mais adequada para o estudo da interação  $QTL \times E$ , a qual era de fato a principal meta, do que para estudos de estabilidade de QTLs (PIEPHO, 2000). No entanto, devido à ausência de “encolhimento” dos efeitos fixos, visto que nos modelos mistos apenas os efeitos aleatórios são estimados pelo método BLUP, é importante ressaltar que as estimativas dos efeitos dos QTLs incluídos como termos fixos são superestimadas e, portanto, demasiadamente otimistas (BOER et al., 2007).

Smith, Cullis e Thompson (2005) e Welham et al. (2010) detalharam e compararam duas abor-

dagens concorrentes para a análise de interação  $G \times E$  a partir de dados de múltiplos genótipos avaliados em múltiplos ambientes, isto é, dados do tipo MET. O primeiro método reúne todas as etapas da análise em um único estágio e, por isto, é denominado de análise *one-stage*, em inglês. Neste caso, dados de parcelas individuais são utilizados como entrada para o ajuste de um modelo estatístico, o qual utiliza as informações de todos os ambientes simultaneamente para estimar efeitos e realizar testes de hipóteses. Esses autores mostram que esta abordagem proporciona máximo poder estatístico e ausência de viés na estimação dos efeitos. Já no segundo método, denominado análise *two-stage*, inicialmente ajustam-se modelos separados para cada ambiente, a partir dos quais são obtidos BLUPs para cada genótipo, utilizados para compor um tabela de médias de genótipos por ambientes. A segunda etapa da análise consiste em utilizar esta tabela, acompanhada ou não por pesos que refletem a variância residual e, por conseguinte, a precisão experimental de cada experimento, para modelar a interação  $G \times E$ . Esta segunda abordagem, apesar de acelerar as análises e permitir o tratamento de uma quantidade muito maior de dados experimentais, pode tornar as análises viesadas e resultar em redução do poder estatístico, e deve ser vista como uma aproximação para a abordagem mais eficiente de análise *one-stage*. No presente trabalho, havia 4.032 observações, no total de genótipos, testemunhas, caracteres, locais e cortes, número razoavelmente pequeno que viabilizou a realização de todas as análises em um único passo. Assim, garantiu-se que as conclusões fossem tomadas a partir de resultados não viesados e com elevado poder estatístico (SMITH; CULLIS; THOMPSON, 2005; WELHAM et al., 2010).

As análises para cada caráter separadamente não forneceram, em grupo de ligação algum, evidências em favor da hipótese de QTLs ligados ao invés de apenas um QTL pleiotrópico (Figura 4). Por conseguinte, as posições foram mantidas inalteradas e os efeitos foram estimados, em cada uma delas, simultaneamente para todos os caracteres. De acordo com o critério de significância  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$  (Tabela 4), os 21 efeitos detectados exibiram diferentes padrões de pleiotropia. Apenas 6 deles exibiram alguma influência sobre os três caracteres, a saber: QTL I (grupo de ligação LG92), interação entre os QTLs II e V, QTL VIII (grupo LG64), QTL IX (grupo LG7), QTL X (marcador ESTC19m4C) e QTL XI (grupo LG10). Do total restante, os 7 efeitos a seguir influenciaram dois caracteres: o QTL II (grupo LG55) e a interação VII  $\times$  XIII apresentaram efeitos significativos para Fiber e POL; o QTL V (marcador ESTB45m6D1) e a interação VII  $\times$  XI influenciaram os caracteres Fiber e TCH; o QTL III (grupo LG11), QTL VI (marcador EST8CC) e a interação V  $\times$  VI participaram do controle dos caracteres POL e TCH. Finalmente, 7 efeitos

afetaram apenas um caráter: os QTL IV (grupo LG72) e XII (marcador ESTB64m3C) exibiram significância apenas para o caráter Fiber; QTL VII (grupo LG64), QTL XIII (grupo LG49) e a epistasia VI  $\times$  XII influenciaram o caráter POL; as interações IX  $\times$  XIII e III  $\times$  VI apresentaram efeitos significativos para TCH.

Pastina (2010), ao utilizar modelos mistos univariados para mapear QTLs para o mesmo conjunto de dados, detectou um efeito significativo para o caráter TCH no grupo de ligação LG72, o qual apresentou influência apenas sobre o caráter Fiber no presente estudo (QTL IV na Tabela 4). Por outro lado, observa-se na Figura 4 que este mesmo grupo exibiu o pico mais pronunciado exatamente para o caráter TCH. Esta aparente discrepância pode ser devida ao fato de que o critério de significância individual proposto por Malosetti et al. (2008) pode não refletir adequadamente a significância do teste de Wald, o qual testa todos os efeitos conjuntamente. Fato bastante semelhante ocorreu para o QTL XII, associado ao marcador ESTB64m3C, o qual aqui exibiu efeitos significativos apenas para Fiber, enquanto que Pastina (2010) detectou significância para o caráter TCH. Consistência entre os resultados foi observada para o QTL II localizado no grupo LG55, o qual influenciou o caráter Fiber em ambos os trabalhos. É interessante ressaltar que, apesar do número relativamente elevado de QTLs encontrados em ambos os trabalhos, houve pequena sobreposição entre eles. De fato, apenas os três efeitos acima mencionados foram detectados em comum nos dois estudos. Isto evidencia a expressiva dificuldade da tarefa de localizar locos associados a caracteres quantitativos, pois mesmo análises bastante sofisticadas como as empregadas em ambos os trabalhos resultaram em conclusões sensivelmente distintas.

O estudo aqui realizado tinha entre seus principais objetivos verificar a ocorrência de interação QTL  $\times$  E e, de fato, algumas causas dessa interação puderam ser investigadas. Inicialmente, é essencial observar que todos os QTLs detectados, tanto os efeitos principais quanto os epistáticos, exibiram algum tipo de interação, havendo três possibilidades: QTL  $\times$  L, QTL  $\times$  C ou QTL  $\times$  L  $\times$  C. Assim, não houve QTL algum com expressão constante em todos os ambientes utilizados, o que pode ter implicações importantes para a seleção assistida por esses marcadores. Mais especificamente, não seria possível selecionar um marcador associado a um QTL com o mesmo efeito nos dois locais, ou com efeito inalterado ao longo dos anos. De fato, seria necessário selecionar marcadores específicos para cada local, e cujos efeitos possivelmente oscilariam em função de variações em cada ano agrícola. Isto parece estar de acordo com o que é amplamente discutido pelos melhoristas, ou seja, que as interações estão sempre presentes e que, por conseguinte, é difícil ou

mesmo impossível encontrar genótipos amplamente adaptados a muitas condições ambientais.

Apesar de não haver QTLs com efeitos constantes, é interessante notar que não foram observadas variações nos sinais dos efeitos dentro de um mesmo local ao longo dos cortes, ou entre os locais para um mesmo corte (Figura 5). Assim, a seleção para um QTL qualquer em um determinado local não afetaria negativamente a performance dos genótipos no outro local. Em contraste, os resultados de Pastina (2010) indicam que alguns QTLs expressaram-se de maneira consistente em todas as combinações local  $\times$  corte, quando cada caráter foi analisado individualmente. Na realidade, nota-se na Tabela 4 que alguns QTLs exibiram efeitos bastante semelhantes nos seis ambientes, para caracteres isolados. Por exemplo, os efeitos do QTL V, associado ao marcador ESTB45m6D1, variaram apenas entre  $-0,432$  e  $-0,306$  para o caráter Fiber. Por outro lado, os efeitos deste mesmo marcador foram muitos mais variáveis para os caracteres POL (de  $0,001$  a  $0,330$ ) e TCH (de  $-8,278$  a  $-3,836$ ). Logo, como o teste de Wald verifica a significância das interações para os três caracteres simultaneamente, é provável que a ocorrência de interação para pelo menos um caráter tenha causado a significância estatística dessa interação para os efeitos em conjunto.

Além desses resultados dos testes de Wald para interações, a Tabela 4 mostra que, para diversos QTLs, um único efeito foi significativo para um dado caráter, ou seja, o QTL expressou-se em apenas um ano e em apenas um local. Um exemplo pode ser visto para o QTL I (grupo LG92), cujo efeito para POL foi significativo apenas no terceiro corte do local 2 (Jaú). Isto foi verificado para 9 dos 21 efeitos detectados, se incluído o último termo epistático, que não exibiu efeitos significativos pelo critério  $|\text{efeito}| \geq 2 \times \text{desvio-padrão}$ . Esta observação deixa evidente que esses resultados devem ser utilizados com cautela, seja porque o padrão de expressão dos QTLs tenha efetivamente sido errático, o que reduziria sua aplicabilidade em programas de seleção assistida, seja porque a análise estatística não tenha possuído poder o suficiente para detectar efeitos de menor magnitude. Em ambos os casos, avaliações complementares seriam necessárias para confirmar a ocorrência ou não desses efeitos de natureza irregular. Uma ressalva deve ser feita para o fato de que esse critério de significância não considera a possibilidade de ocorrência de falsos positivos como consequência dos múltiplos testes - no presente caso foram realizados até 18 testes para cada efeito incluído no modelo. Por outro lado, o reduzido tamanho amostral pode dificultar a detecção de QTLs de pequeno efeito.

Passando à correlação entre caracteres, Malosetti et al. (2008) afirmaram que uma análise integrada de QTLs fornece informações importantes a respeito da determinação genética e relação entre



caracteres. Conforme descrito anteriormente, 13 dos 21 efeitos detectados (61,9%) influenciaram pelo menos dois dos três caracteres estudados, ou seja, a maioria dos QTLs exibiu algum grau de pleiotropia. Ademais, é marcante o fato de virtualmente todos os QTLs pleiotrópicos exibirem o seguinte padrão de sinais dos efeitos: + para Fiber, - para POL e + para TCH; ou - para Fiber, + para POL e - para TCH (Tabela 4 e Figura 5). Quaisquer desvios deste padrão ocorreram apenas para efeitos de menor magnitude, os quais não foram significativos. Portanto, observa-se que todos os QTLs pleiotrópicos contribuíram no sentido de causar correlação positiva entre Fiber e TCH, mas negativa entre Fiber e POL e entre POL e TCH. De fato, a correlação fenotípica entre Fiber e TCH foi igual a 0,391, de modo que os efeitos dos QTLs acompanharam este mesmo sentido. Por outro lado, o caráter POL apresentou-se fenotipicamente não correlacionado com os demais, tendo exibido correlação igual a -0,005 com Fiber e 0,066 com TCH, de tal maneira que os QTLs detectados não se mostraram condizentes com tal observação. Não obstante, o fato de a correlação genética entre POL e os outros dois caracteres ter sido negativa está de acordo com o que é observado empiricamente pelos melhoristas. Portanto, foram aqui obtidas informações importantes sobre os mecanismos responsáveis por ao menos parte da correlação entre os caracteres. Caso tivessem sido encontrados QTLs com diferentes padrões de sinais, ou seja, com contribuições distintas à correlação genética, esta informação seria valiosa para a seleção assistida, pois abriria a possibilidade de se quebrar parcialmente a correlação, isto é, selecionar genótipos com valores aumentados de todos os caracteres. Evans (2002) demonstrou teoricamente que QTLs pleiotrópicos atuando em sentido oposto à correlação fenotípica são detectados com maior facilidade, visto que os testes de hipóteses têm maior poder estatístico nesta situação. O fato de todos os QTLs detectados exibirem o mesmo padrão de correlações pode significar que realmente não havia efeitos de outra natureza na progênie estudada, ou pode ser uma consequência de baixo poder estatístico para detecção desse outro tipo de efeito.

O método MIM é o único a permitir a inclusão de termos epistáticos diretamente no processo de busca por QTLs (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999). Dos 21 efeitos detectados, 8 corresponderam a interações entre QTLs, o que representa praticamente 40% do total. Outro ponto interessante é que 9 dos 13 efeitos principais (69,23%) estiveram envolvidos em interações epistáticas, a saber: QTLs II, III, V, VI, VII, IX, XI, XII e XIII (Tabela 4). Isto mostra claramente que modelos mais realistas sobre os fenômenos genéticos não podem ignorar o fenômeno da epistasia. O QTL VI foi o que participou do maior número de interações - três, com os QTLs III, V e XII. Podlich

et al., 2004, e Cooper et al., 2005 apud Boer et al. (2007) mostraram que interações epistáticas entre QTLs e o *background*, em combinação com a interação QTL  $\times$  E, podem ser importantes para programas de seleção assistida por marcadores. Realmente, à semelhança do que foi observado para efeitos principais, a Tabela 4 evidencia que os efeitos epistáticos não foram consistentes ao longo dos locais ou ambientes, de tal forma que os valores genéticos dos indivíduos devem ser calculados especificamente para cada local, caso haja interesse em utilizar essa informação para seleção. Fica evidente, no entanto, a complexidade do controle genético dos caracteres quantitativos.

Alguns autores realizaram buscas por efeitos epistáticos através de abordagens digênicas e modelos de análise de variância ou regressão (SILLS et al., 1995; HOARAU et al., 2002; AITKEN; JACKSON; McINTYRE, 2006; AITKEN et al., 2008), mas tal abordagem sem controle da variação genética residual geralmente resulta em altas taxas de falsos positivos (Li, 1997 e Li et al., 1997a, 1997b apud Wang et al., 1999). Os autores deste último estudo ainda afirmaram que, assim como a inclusão de cofatores com efeitos principais no método CIM controla com sucesso a influência da variação genética residual no mapeamento de QTLs na ausência de epistasia, a inclusão de marcadores interatuantes ligados a QTLs epistáticos aumenta o poder, acurácia e precisão do mapeamento de QTLs. Em seu trabalho, esses marcadores foram considerados como aleatórios, de modo a controlar a variação sem causar superparametrização do modelo. Em contraste, como o presente estudo utilizou o método MIM de mapeamento de QTLs, as epistasias também foram incluídas como efeitos fixos, pois havia interesse explícito nos efeitos de cada interação.

À exceção do trabalho descrito por Pastina (2010), o qual também fez uso de modelos mistos e modelou diretamente as interações QTL  $\times$  L, QTL  $\times$  C e QTL  $\times$  L  $\times$  C, não é trivial fazer comparações entre os resultados aqui obtidos e aqueles publicados por outros autores. Conforme descrito na seção 2.3.2, em alguns trabalhos foram feitas análises de mapeamento de QTLs em vários locais e/ou cortes, mas sempre com uma abordagem aproximada, em que as análises foram conduzidas separadamente para cada ambiente e os resultados foram visualmente comparados, a fim de se verificar a existência de QTLs com efeitos estáveis ou exibindo interação com o ambiente. Mesmo a comparação com as conclusões de Pastina (2010) devem ser feitas com cautela, já que as análises foram realizadas para cada caráter individualmente em seu trabalho, enquanto que aqui foi utilizada uma análise multivariada. Isto pode ser verificado, por exemplo, para o grupo LG35, no qual Pastina (2010) detectou QTLs com efeitos opostos para os caracteres POL e Fiber, possivelmente representando um único QTL pleiotrópico, ao passo que no presente trabalho não foram detectados

efeitos significativos nesse grupo de ligação. Outra observação importante é que, por ter realizado análises separadas para cada caráter, Pastina (2010) pode ter encontrado maior número de falsos positivos, já que cada posição do genoma foi testada para presença de QTLs quatro vezes, uma para cada caráter, ou seja, um número maior de testes foi realizado. Apesar de limiares conservativos fornecerem alguma proteção contra múltiplos testes, isto foi feito sem considerar que múltiplos caracteres foram estudados (Caliński et al., 2000). É também interessante observar que Pastina (2010) buscou QTLs tanto com efeitos específicos para cada combinação local  $\times$  corte, quanto com efeitos principais, isto é, efeitos para a média de todos os ambientes. Malosetti et al. (2004) argumentaram que esta última abordagem é mais apropriada para estudos de QTLs em um contexto multi-ambiente e, por conseguinte, também em um contexto multi-caráter. Outra potencial fonte de discrepâncias entre o presente estudo e o de Pastina (2010) é que neste último a ocorrência de epistasia não foi considerada. Como o método MIM a incorpora diretamente no processo de busca, a inclusão de uma interação epistática pode subsequentemente influenciar a detecção de outros efeitos, inclusive efeitos principais.

Este é o único trabalho, de nosso conhecimento, a fazer uso de um modelo misto multivariado para busca de QTLs pleiotrópicos e/ou interatuantes com o ambiente em cana-de-açúcar. Tal abordagem traz muitas vantagens, com destaque para o aumento do poder estatístico e a redução da ocorrência de falsos positivos, benefícios que, em conjunto, tornam as conclusões mais confiáveis (PIEPHO, 2005; MALOSETTI et al., 2008). A realização de todas as análises em uma única etapa, com modelagem direta das variâncias e covariâncias genéticas entre caracteres e entre ambientes, elimina vieses e torna as estimativas dos efeitos dos QTLs válidas, se considerado seu possível uso para seleção assistida (WELHAM et al., 2010). Além disso, o ajuste de múltiplos QTLs, tanto com efeitos principais quanto epistáticos, em um único modelo MIM, permite a obtenção de estimativas dos valores genotípicos dos indivíduos, os quais podem ser diretamente canalizados para programas de melhoramento genético (ZENG, 2001). Isto também representa uma importante vantagem do modelo MIM em relação às metodologias precedentes, as quais estimam os efeitos de cada QTL individualmente, mesmo se considerada a inclusão de cofatores no modelo CIM (BROMAN; SEN, 2009).

De qualquer maneira, apesar de este trabalho utilizar modelos mais sofisticados para o mapeamento de QTLs em cana-de-açúcar e representar ganhos e benefícios em diversos aspectos, muitos fatores influenciam a confiabilidade dos resultados e muito ainda resta a ser feito. Em primeiro

lugar, o modelo de mapeamento aqui utilizado foi desenvolvido para espécies diplóides, em que não há possibilidade de ocorrer alelos em múltiplas doses, de tal forma que as conclusões apresentadas também são aproximações. Inicialmente, é preciso desenvolver metodologias de marcadores moleculares que possibilitem a genotipagem dos indivíduos com estimativas precisas do número de cópias dos alelos. Em seguida, métodos de construção de mapas genéticos específicos para poliplóides serão necessários para tratamento desses dados moleculares. Por fim, modelos de mapeamento de QTLs capazes de estimar efeitos quantitativos das múltiplas doses deverão ser delineados.

Especificamente no caso dos dados aqui utilizados, é essencial ressaltar que há também algumas limitações. O primeiro aspecto de grande importância diz respeito ao tamanho populacional. Claramente, uma amostra de 100 indivíduos não representa o caso ideal para mapeamento de QTLs, visto que grandes populações são necessárias para sua detecção e localização com maior poder e precisão, principalmente se considerada a inclusão da interação  $QTL \times E$  no modelo (BOER et al., 2007; MACKAY; STONE; AYROLES, 2009). Boer et al. (2007) afirmaram que populações com aproximadamente 500 indivíduos são suficientes para detecção de QTLs. Esses autores também enfatizaram a importância de se trabalhar com número mais elevado de ambientes, de modo a reduzir os erros nas estimativas dos efeitos. Trabalhos futuros em cana-de-açúcar certamente requerirão progênies maiores e avaliações mais extensivas em termos de ambientes amostrados.

Outro ponto crucial está relacionado à qualidade do mapa de ligação. A integração de marcadores com segregação 1:1 e 3:1 em um único mapa, válido para ambos os genitores, representa um avanço em relação à metodologia do duplo *pseudo-testcross*, pois resulta em maior saturação do mapa, melhor representação do polimorfismo gerado pelo cruzamento, maior poder estatístico e possibilidade de estimação de efeitos para os dois genitores no mesmo grupo de ligação (PASTINA, 2010). Além disso, como o software *OneMap* faz uso de metodologia multiponto, o mapa integrado utilizado possui maior verossimilhança do que os previamente publicados para o mesmo cruzamento (GARCIA et al., 2006; OLIVEIRA et al., 2007). Em contrapartida, apenas marcadores em dose única foram empregados, o que reduziu a cobertura do genoma, devido ao descarte de marcadores com outros padrões de segregação. A reduzida saturação pode também ser notada pelo fato de que muitos grupos de ligação não puderam ser integrados, em função da presença apenas de marcadores do tipo  $D_1$  ou  $D_2$ . O número de marcadores por grupo de ligação foi baixo, o comprimento dos grupos de ligação foi bastante pequeno, em média, e houve ainda muitos marcadores não ligados a grupo de ligação algum. Assim, é provável que diversos grupos de ligação representem

porções não unidas de grupos maiores. Em conjunto, essas considerações impediram a realização plena das vantagens do método MIM.

Não obstante as limitações inerentes ao conjunto de dados aqui empregado, a realização de uma análise MTME em uma única etapa deu origem a resultados não viesados e com máximo poder estatístico. A modelagem sob o prisma dos modelos mistos permitiu considerar variâncias genéticas heterogêneas e covariâncias, eliminando assim a adoção de suposições irrealistas. Por fim, a fundamentação biológica do método MIM culminou em resultados mais facilmente interpretáveis e de aplicabilidade direta em programas de melhoramento genético. A união das vantagens de cada uma dessas frentes de conhecimento torna o método aqui desenvolvido genuinamente apropriado para o tratamento de dados dessa natureza.

## 6 CONCLUSÃO

O método de mapeamento delineado no presente trabalho foi eficiente para o estudo da arquitetura genética de três caracteres quantitativos da cana-de-açúcar, tendo sido encontrados alguns QTLs com efeitos consistentes ao longo dos cortes e locais, além de alguns QTLs exibindo forte interação com o ambiente. Também foram obtidas informações a respeito da base genética da correlação entre esses caracteres. O uso do método MIM permitiu a detecção de epistasia durante o processo de busca por QTLs, o que é extremamente interessante tanto do ponto de vista estatístico quanto biológico. A utilização dos modelos mistos permitiu a extração do máximo de informação dos dados de múltiplos caracteres e ambientes, além de possibilitar uma modelagem adequada das fontes de variação experimentais.



## REFERÊNCIAS

AgraFNP. Cana. In \_\_\_\_\_. **AGRIANUAL 2011**: Anuário Estatístico da Agricultura Brasileira. São Paulo, 2011. p. 215-236.

AHN, S.; TANKSLEY, S.D. Comparative linkage maps of the rice and maize genomes. **Proceedings of the National Academy of Science of the United States of America**, Washington, v. 90, p. 7980-7984, 1993.

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 a sugarcane cultivar. **Theoretical and Applied Genetics**, New York, v. 110, 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**, New York, v. 112, p. 1306-1317, 2006.

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

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**, New York, v. 117, p. 1191-1203, 2008.

AKAIKE, H. A new look at the statistical model identification. **The Institute of Electrical and Electronics Engineers Transactions on Automatic Control**, Notre Dame, v. 19, p. 716-723, 1974.

AL-JANABI, S.M.; HONEYCUTT, R.J.; McCLELLAND, M.; SOBRAL, B.W.S. A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. **Genetics**, Bethesda, v. 134, 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 (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. **Molecular Breeding**, Berlin, v. 19, p. 1-14, 2007.

ALWALA, S.; KIMBENG, C.A. Molecular genetic linkage mapping in *Saccharum*: strategies, resources and achievements. In: HENRY, R.J. (Ed.) **Genetics, Genomics and Breeding of Sugarcane**. New Hampshire: Science Publishers, 2010. 1 v.

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**, Wageningen, v. 164, p. 37-51, 2008.

ANDRU, S.; PAN, Y.B.; THONGTHAWEE, S.; BURNER, D.M.; KIMBENG, C.A. Genetic analysis of the sugarcane (*Saccharum* spp.) cultivar 'LCP 85-384'. I. Linkage mapping using AFLP, SSR, and TRAP markers. **Theoretical and Applied Genetics**, New York, v. 123, p. 77-93, 2011.

ASNAGHI, C.; D'HONT, A.; GLASZMANN, J.C.; ROTT, P. Resistance of sugarcane cultivar R 570 to *Puccinia melanocephala* isolates from different geographic locations. **Plant Disease**, Sant Paul, v. 85, p. 282-286, 2001.

BALZARINI, M. Applications of mixed models in plant breeding. In: KANG, M.S. (Ed.) **Quantitative genetics, genomics and plant breeding**. New York: CABI Publishing. 2001. p. 353-363.

BASTEN, C.J.; WEIR, B.S.; ZENG, Z.B. **QTL Cartographer**: a reference manual and tutorial for QTL mapping. Raleigh, 2005. Disponível em: <<http://statgen.ncsu.edu/qtlcart>>. Acesso em: 25 maio 2011.

- BELSLEY, D.A.; KUH, E.; WELSCH, R.E. **Regression diagnostics**: identifying influential data and sources of collinearity. New York: John Wiley & Sons, 1980. 320 p.
- BOER, M.P.; WRIGHT, D.; FENG, L.; PODLICH, D.W.; LUO, L.; COOPER, M.; EEUWIJK, F.A. van. A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. **Genetics**, Bethesda, v. 177, p. 1801-1813, 2007.
- BROMAN, K.W. Review of statistical methods for QTL mapping in experimental crosses. **Lab Animal**, New York, v. 30, p. 44-52, 2001.
- BROMAN, K.W. SEN, S. **A guide to QTL mapping with R/qtl**. New York: Springer, 2009. 396 p.
- BROMAN, K.W.; WU, H.; SEN, S.; CHURCHILL, G.A. R/qtl: QTL mapping in experimental crosses. **Bioinformatics**, Oxford, v. 19, p. 889-890, 2003.
- BURNHAM, K.P.; ANDERSON, D.R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods & Research**, London, v. 33, p. 261-304, 2004.
- CALIŃSKI, T.; KACZMAREK, Z.; KRAJEWSKI, P.; FROVA, C.; SARI-GORLA, M. A multivariate approach to the problem of QTL localization. **Heredity**, London, v. 84, p. 303-310, 2000.
- CULLIS, B.; GOGEL, B.; VERBYLA, A.; THOMPSON, R. Spatial analysis of multi-environment early generation variety trials. **Biometrics**, Washington, v. 54, p. 1-18, 1998.
- CROSSA, J.; VARGAS, M.; EEUWIJK, F.A. van; JIANG, C.; EDMEADES, G.O.; HOISINGTON, D. Interpreting genotype  $\times$  environment interaction in tropical maize using linked molecular markers and environmental covariables. **Theoretical and Applied Genetics**, New York, v. 99, p. 611-625, 1999.

DAUGROIS, J.H.; GRIVET, L.; ROQUES, D.; HOARAU, J.Y.; LOMBARD, 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, p. 1059-1064, 1996.

DEKKERS, J.C.M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, New York, v.3, p.22-32, 2002.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society B**, London, v. 39, p. 1-38, 1977.

D'HONT, A.; LU, Y.H.; GONZÁLES de LEON, D.; GRIVET, L.; FLEDMANN, P.; LANAUD, C.; GLASZMANN, J.C. A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. **Genome**, Ottawa, v. 37, p. 222-230, 1994.

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 Genetics and Genomics**, Berlin, v. 250, p. 405-413, 1996.

D'HONT, A.; ISON, D.; ALIX, K.; GLASZMANN, J.C. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. **Genome**, Ottawa, v. 41, p. 221-225, 1998.

DOERGE, R.W. Mapping and analysis of quantitative trait loci in experimental populations. **Nature Reviews Genetics**, New York, v. 3, p. 43-52, 2002.

DUARTE, J.B.; VENCOSKY, R. Interação genótipos × ambientes: uma introdução à análise "AMMI". **Série Monografias, Sociedade Brasileira de Genética**, Ribeirão Preto, 1999. 60 p.

- ECKERMANN, P.J.; VERBYLA, A.P.; CULLIS, B.R.; THOMPSON, R. The analysis of quantitative traits in wheat mapping populations. **Australian Journal of Agricultural Research**, Melbourne, v. 52, p. 1195-1206, 2001.
- EEUWIJK, F.A. van; BOER, M.; TOTIR, L.R.; BINK, M.; WRIGHT, D.; WINKLER, C.R.; PODLICH, D.; BOLDMAN, K.; BAUMGARTEN, A.; SMALLEY, M.; ARBELBIDE, M.; ter BRAAK, C.J.F.; COOPER, M. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. **Theoretical and Applied Genetics**, New York, v. 120, 429-440, 2010.
- EEUWIJK, F.A. van; COOPER, M.; DeLACY, I.H.; CECCARELLI, S.; GRANDO, S. Some vocabulary and grammar for the analysis of multi-environment trials, as applied to the analysis of FPB and PPB trials. **Euphytica**, Wageningen, v. 122, p. 477-490, 2001.
- EEUWIJK, F.A. van; MALOSETTI, M.; YIN, X.; STRUIK, P.C.; STAM, P. Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. **Australian Journal of Agricultural Research**, Melbourne, v. 56, p. 883-894, 2005.
- EDWARDS, M.D.; STUBER, C.W.; WENDEL, J.F. Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. **Genetics**, New York, v. 116, p. 113-125, 1987.
- EVANS, D.M. The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between variables. **American Journal of Human Genetics**, Baltimore, v. 70, p. 1599-1602, 2002.
- FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. 4.ed. Essex, UK: Longman, 1996. 464 p.
- FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: EMBRAPA-CENARGEN, 1998. 220 p.

FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Melbourne, v. 14, p. 742-754, 1963.

FNP Consultoria & Comércio. Cana. In \_\_\_\_\_. **AGRIANUAL 2006**: Anuário Estatístico da Agricultura Brasileira. São Paulo, 2006. p. 227-248.

FRARY, A.; NESBITT, T.C.; FRARY, A.; GRANDILLO, S.; KNAAP, E. van der; CONG, B.; LIU, J.; MELLER, J.; ELBER, R.; ALPERT, K.B.; TANKSLEY, S.D. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. **Science**, Washington, v. 289, p. 85-88, 2000.

GABRIEL, K.R. Least squares approximation of matrices by additive and multiplicative models. **Journal of the Royal Statistical Society B**, London, v. 40, p. 186-196, 1978.

GARCIA, A.A.F.; KIDO, E.A.; MEZA, A.N.; SOUZA, H.M.B.; PINTO, L.R.; PASTINA, M.M.; LEITE, C.S.; SILVA, J.A.G. da; 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 for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, New York, v. 112, p. 298-314, 2006.

GAZAFFI, R. **Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênie de irmãos completos, com aplicação em cana-de-açúcar**. 2009. 103 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2009.

GILMOUR, A.R.; CULLIS, B.R.; VERBYLA, A.P. Accounting for natural and extraneous variation in the analysis of field experiments. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 2, p. 269–293, 1997.

GILMOUR, A.R.; THOMPSON, R.; CULLIS, B.R. Average Information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Washington, v. 51, p. 1440-1450, 1995.

GOGEL, B.J.; CULLIS, B.R.; VERBYLA, A.P. REML estimation of multiplicative effects in multi-environment variety trials. **Biometrics**, Washington, v. 51, p. 744-749, 1995.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. **Genetics**, Bethesda, v. 137, p. 1121-1137, 1994.

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, p. 987-1000, 1996.

GUIMARÃES, C.T.; SILLS, G.R.; SOBRAL, B.W.S. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Science of the United States of America**, Washington, v. 94, p. 14261-14266, 1997.

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, p. 125-132, 1999.

GUO, Z.; NELSON, J.C. Multiple-trait quantitative trait locus mapping with incomplete phenotypic data. **BMC Genetics**, London, v. 9, p. 82-93, 2008.

HALEY, C.S.; KNOTT, S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity**, London, v. 69, p. 315-324, 1992.

HEINZ, D.J.; TEW, T.L. Hybridization procedures. In: HEINZ, D.J. (Ed.) **Sugarcane improvement through breeding**. Amsterdam: Elsevier. 1987. p. 313-342.

HENDERSON, C.R. Statistical methods in animal improvement: historical overview. In: GIANOLA, D.; HAMMOND, K. (Ed.) **Advances in statistical methods for genetic improvement of livestock**. New York: Springer-Verlag, 1990. p. 1-14.

HOARAU, J.Y.; OFFMANN, B.; D'HONT, A.; RISTERUCCI, A.M.; ROQUES, D.; GLASZMANN, 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, p. 84-97, 2001.

HOARAU, J.Y.; GRIVET, L.; OFFMANN, B.; RABOIN, L.M.; DIORFLAR, J.P.; PAYET, J.; HELLMANN, M.; D'HONT, A.; GLASZMANN, 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, p. 1027-1037, 2002.

IRVINE, J.E. *Saccharum* species as horticultural classes. **Theoretical and Applied Genetics**, New York, v. 98, p. 186-194, 1999.

JANSEN, R.C.; STAM, P. High resolution of quantitative traits into multiple loci via interval mapping. **Genetics**, Bethesda, v. 136, p. 1447-1455, 1994.

JIANG, C.; ZENG, Z.B. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Bethesda, v. 140, p. 1111-1127, 1995.

JOHNSON, D.L.; JANSEN, R.C.; ARENDONK, J.A.M. van. Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. **Genetical Research**, London, v. 73, p. 75-83, 1999.

JORDAN, D.R.; CASU, R.E.; BESSE, P.; CARROLL, B.C.; BERDING, N.; McINTYRE, C.L. Markers associated with stalk number and suckering in sugarcane collocate with tillering and rhizomatousness QTLs in sorghum. **Genome**, Ottawa, v. 47, p. 988-993, 2004.

KAO, C.H.; ZENG, Z.B. General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. **Biometrics**, Washington, v. 53, p. 653-665, 1997.

KAO, C.H., ZENG, Z.B.; TEASDALE, R D. Multiple interval mapping for quantitative trait loci. **Genetics**, Bethesda, v. 152, p. 1203-1216, 1999.

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. 164 p. Tese (Doutorado em Energia na Agricultura) – CENA-Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2003.

KNOTT, S.A.; HALEY, C.S. Multitrait least squares for quantitative trait loci detection. **Genetics**, Bethesda, v. 156, 899-911, 2000.

KOROL, A.B.; RONIN, Y.I.; NEVO, E.; HAYES, P.M. Multi-interval mapping of correlated trait complexes. **Heredity**, London, v. 80, p. 273-284, 1998.

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

LANDER; E.S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Bethesda, v. 121, p. 185-199, 1989.

LANDER, E.S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M.J.; LINCON S.E.; NEWBURG, L. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics**, Orlando, v. 1, p. 174-181, 1987.

LEWIN, H.A.; LARKIN, D.M.; PONTIUS, J.; O'BRIEN, S.J. Every genome sequence needs a good map. **Genome Research**, Woodbury, v. 19, p. 1925-1928, 2009.

LIN, M.; LOU, X.Y.; CHANG, M.; WU, R. A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. **Genetics**, Bethesda, v. 165, p. 901-913, 2003.



LIU, B.H. **Statistical genomics: linkage, mapping, and QTL analysis**. Boca Raton: CRC Press, 1998. 611p.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980p.

MACKAY, T.F.C. Quantitative trait loci in *Drosophila*. **Nature Reviews Genetics**, New York, v. 2, p. 11-20, 2001a.

MACKAY, T.F.C. The genetic architecture of quantitative traits. **Annual Review of Genetics**, Palo Alto, v. 35, p. 303-339, 2001b.

MACKAY, T.F.C.; STONE, E.A.; AYROLES, J.F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews Genetics**, New York, v. 10, p. 565-577, 2009.

MALIEPAARD, C.; JANSEN, J.; VAN OOIJEN, J.W. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, Cambridge, v. 70, p. 237-250, 1997.

MALOSETTI, M.; RIBAUT, J.M.; VARGAS, M.; CROSSA, J.; EEUWIJK, F.A. van. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Wageningen, v. 161, p. 241-257, 2008.

MALOSETTI, M.; VOLTAS, J.; ROMAGOSA, I.; ULLRICH, S.E.; EEUWIJK, F.A. van. Mixed models including environmental covariables for studying QTL by environment interaction. **Euphytica**, Wageningen, v. 137, p. 139-145, 2004.

MALOSETTI, M.; BOER, M.P.; BINK, M.C.A.M.; EEUWIJK, F.A. van. Multi-trait QTL analysis based on mixed models with parsimonious covariance matrices. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings...**, Belo Horizonte: WCGALP, 2006. Abstract 25-04.

MARGARIDO, G.R.A.; SOUZA, A.P.; GARCIA, A.A.F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, Lund, v. 144, p. 78-79, 2007.

MARTÍNEZ, O.; CURNOW, R.N. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. **Theoretical and Applied Genetics**, New York, v. 85, p. 480-488, 1992.

MATHEWS, K.L.; MALOSETTI, M.; CHAPMAN, S.; McINTYRE, L.; REYNOLDS, M.; SHORTER, R.; EEUWIJK, F.A. van. Multi-environment QTL mixed models for drought stress adaptation in wheat. **Theoretical and Applied Genetics**, New York, v. 117, p. 1077-1091, 2008.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. (Ed.) **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999. p. 205-252.

MATSUOKA, S.; GARCIA, A.A.F.; CALHEIROS, G.C. Hibridação em cana-de-açúcar. In: BORÉM, A. (Ed.) **Hibridação artificial em plantas**. Viçosa: UFV, 1999. p. 221-256.

McCULLOCH, C.E.; SEARLE, S.R. **Generalized, linear, and mixed models**. New York: John Wiley & Sons, 2001. 325 p.

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**, Berlin, v. 16, p. 151-161, 2005a.

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

McINTYRE, C.L.; JACKSON, M.; CORDEIRO, G.M.; AMOUYAL, O.; HERMANN, S.; AITKEN, K.S.; ELIOTT, F.; HENRY, R.J.; CASU, R.E.; BONNETT, G.D. The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. **Molecular Breeding**, Berlin, v. 18, p. 39-50, 2006.

MEZA, A.N. **Construção de um mapa genético para híbridos interespecíficos de variedades comerciais de cana-de-açúcar (*Saccharum spp.*) empregando-se marcadores moleculares do tipo RFLP**. 2002. 102 p. Tese (Doutorado em Genética Vegetal e Melhoramento) – Instituto de Biologia, Universidade Estadual de Campinas, Campinas, 2002.

MING, R.; DEL MONTE, T.A.; HERNANDEZ, E.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. **Genome**, Ottawa, v. 45, p. 794-803, 2002a.

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

MING, R.; LIU, S.C.; LIN, Y.R.; SILVA, J. da; WILSON, W.; BRAGA, D.; DEYNZE, A. van; WENSLAFF, T.F.; WU, K.K.; MOORE, P.H.; BURNQUIST, W.; SORRELLS, M.E.; IRVINE, J.E.; PATERSON, A.H. Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. **Genetics**, Bethesda, v. 150, p. 1663-1682, 1998.

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, p. 332-345, 2002c.

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**, London, v. 11, p. 2075-2084, 2001.

MOHAN, M.; NAIR, S.; BHAGWAT, A.; KRISHNA, T.G.; YANO, M.; BHATIA, C.R.; SASAKI, T. Genome mapping, molecular markers and markers-assisted selection in crop plants. **Molecular Breeding**, Berlin, v. 3, p. 87-103, 1997.

MOLLINARI, M.; MARGARIDO, G.R.A.; VENCOVSKY, R.; GARCIA, A.A.F. Evaluation of algorithms used to order markers on genetic maps. **Heredity**, London, v. 103, p. 494-502, 2009.

MUDGE, J.; ANDERSEN, W.R.; KEHRER, R.L.; FAIRBANKS, D.J. A RAPD genetic map of *Saccharum officinarum*. **Crop Science**, Madison, v. 36, p. 1362-1366, 1996.

OKADA, M.; LANZATELLA, C.; SAHA, M.C.; BOUTON, J.; WU, R.; TOBIAS, C.M. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. **Genetics**, Bethesda, v. 185, p. 745-760, 2010.

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, p. 189-208, 2007.

OMAN, S.D. Multiplicative Effects in Mixed Model Analysis of Variance. **Biometrika**, Cambridge, v. 78, p. 729-739, 1991.

PASTINA, M.M. **Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar, usando a abordagem de modelos mistos**. 2010. 89 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2010.

PASTINA, M.M.; PINTO, L.R.; OLIVEIRA, K.M.; SOUZA, A.P.; GARCIA, A.A.F. Molecular mapping of complex traits. In: HENRY, R.J. (Ed.) **Genetics, Genomics and Breeding of Sugarcane**. New Hampshire: Science Publishers, 2010. 1 v.

PATERSON, A.H.; BOWER, J.E.; BURROW, M.D.; DRAYE, X.; ELSIK, C.G.; JIANG, C.X.; KATSAR, C.S.; LAN, T.H.; LIN, Y.R.; MING, R.; WRIGHT, R.J. Comparative genomics of plant chromosomes. **The Plant Cell**, Baltimore, v. 12, p. 1523-1539, 2000.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, London, v. 58, p. 545-554, 1971.

PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. GenStat for Windows (12th Edition) Introduction. **VSN International**, Hemel Hempstead, 2009.

PIEPHO, H.P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, p. 761-766, 1997.

PIEPHO, H.P. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. **Theoretical and Applied Genetics**, New York, v. 97, p. 195-201, 1998.

PIEPHO, H.P. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. **Genetics**, Bethesda, v. 156, p. 2043-2050, 2000.

PIEPHO, H.P. Statistical tests for QTL and QTL-by-environment effects in segregating populations derived from line crosses. **Theoretical and Applied Genetics**, New York, v. 110, p. 561-566, 2005.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-Plus**. New York: Springer-Verlag. 2000. 528 p.

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**, Wageningen, v. 172, 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 associations. **Molecular Breeding**, Berlin, v. 21, p. 233-247, 2008.

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, 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, p. 1382-1391, 2006.

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

RIDOUT, M.S., TONG, S.; VOWDEN, C.J.; TOBUTT, K.R. Three-point linkage analysis in crosses of allogamous plant species. **Genetical Research**, London, v. 72, p. 111-121, 1998.

RIPOL, M.I.; CHURCHILL, G.A.; SILVA, J.A.G. da; SORRELLS, M. Statistical aspects of genetic mapping in autoployploids. **Gene**, Amsterdam, v. 235, p. 31-41, 1999.

RITTER, E.; SALAMINI, F. The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. **Genetical Research**, London, v.67, p.55-65, 1996.

RITTER, E.; GEBHARDT, C.; SALAMINI, F. Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. **Genetics**, Bethesda, v. 125, p. 645-654, 1990.

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**, Berlin, v. 269, p. 406-419, 2003.

SCHLÖTTERER, C. The evolution of molecular markers - just a matter of fashion? **Nature Reviews Genetics**, New York, v. 5, p. 63-69, 2004.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Philadelphia, v. 6, p. 461-464, 1978.

SEARLE, S.R. **Linear models**. New York: John Wiley & Sons, 1971. 532 p.

SEMAGN, K.; BJØRNSTAD, Å; NDJIONDJOP, M.N. Principles, requirements and prospects of genetic mapping in plants. **African Journal of Biotechnology**, Nairobi, v. 5, p. 2569-2587, 2006.

SIBOV, S.T.; SOUZA JR, C.L.; GARCIA, A.A.F.; GARCIA, A.F.; SILVA, A.R.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. **Hereditas**, Lund, v. 139, p. 96-106, 2003a.

SIBOV, S.T.; SOUZA JR, C.L.; GARCIA, A.A.F.; SILVA, A.R.; GARCIA, A.F.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 2. Quantitative trait loci (QTL) for grain yield, plant height, ear height and grain moisture. **Hereditas**, Lund, v. 139, p. 107-115, 2003b.

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

SILVA, J.A.G. da; BRESSIANI, J.A. Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, p. 294-298, 2005.

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

SILVA, J.A.G. da; HONEYCUTT, R.J.; BURNQUIST, W.; AL-JANABI, S.M.; SORRELLS, M.E.; TANKSLEY, S.D.; SOBRAL, B.W.S. *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP- and PCR-based markers. **Molecular Breeding**, Berlin, v. 1, p. 165-179, 1995.

SILVA, L.C. **Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses**. 2010. 208 p. Tese (Doutorado em Estatística) - North Carolina State University, Raleigh, 2010.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Washington, v. 57, p. 1138-1147, 2001.



SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **Journal of Agricultural Science**, Cambridge, v. 143, p. 449-462, 2005.

SMITH, A.B.; STRINGER, J.K.; WEI, X.; CULLIS, B.R. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, Wageningen, v. 157, p. 253-266, 2007.

SOBRAL, B.W.S.; HONEYCUTT, R.J. High output genetic mapping in polyploids using PCR-generated markers. **Theoretical and Applied Genetics**, New York, v. 86, p. 105-112, 1993.

SONG, J.Z.; SOLLER, M.; GENIZI, A. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. **Genetical Research**, London, v. 73, p. 61-73, 1999.

SOUZA, A.P. Biologia molecular aplicada ao melhoramento. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. (Ed.) **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT. 2001. p. 939-965.

SOUZA, H.M.B. **Mapeamento genético de híbridos comerciais de cana-de-açúcar (*Saccharum* spp.) utilizando marcadores moleculares do tipo microssatélite**. 2003. 95 p. Dissertação (Mestrado em Energia na Agricultura) – CENA-Centro de Energia Nuclear na agricultura, Universidade de São Paulo, Piracicaba, 2003.

STUBER, C.W.; EDWARDS, M.D.; WENDEL, J.F. Molecular-marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. **Crop Science**, Madison, v. 27, p. 639-648, 1987.

VAN OOIJEN, J.W. **MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species**. Wageningen: Kyazma B.V., 2009.

VARGAS, M.; EEUWIJK, F.A. van.; CROSSA, J.; RIBAUT, J.M. Mapping QTLs and QTL  $\times$  environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. **Theoretical and Applied Genetics**, New York, v. 112, p. 1009-1023, 2006.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer-Verlag, 2000, 568 p.

VERBYLA, A.P.; ECKERMAN, P.J.; THOMPSON, R.; CULLIS, B.R. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. **Australian Journal of Agricultural Research**, Melbourne, v. 54, p. 1395-1408, 2003.

VUYLSTEKE, M.; MANK, R.; ANTONISE, R.; BASTIAANS, E.; SENIOR, M.L.; STUBER, C.W.; MELCHINGER, A.E.; LÜBBERSTEDT, T.; XIA, X.C.; STAM, P.; ZABEAU, M.; KUIPER, M. Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. **Theoretical and Applied Genetics**, New York, v. 99, p. 921-935, 1999.

WANG, S.; BASTEN, C.J.; ZENG, Z.B. **Windows QTL Cartographer 2.5**. Raleigh, 2007. Disponível em: <<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>>. Acesso em: 25 maio 2011.

WANG, D.L.; ZHU, J.; LI, Z.K.; PATERSON, A.H. Mapping QTLs with epistatic effects and QTL $\times$ environment interactions by mixed linear model approaches. **Theoretical and Applied Genetics**, New York, v. 99, p. 1255-1264, 1999.

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

WELHAM, S.J.; GOGEL, B.J.; SMITH, A.B.; THOMPSON, R.; CULLIS, B.R. A comparison of analysis methods for late-stage variety evaluation trials. **Australian & New Zealand Journal of Statistics**, Oxford, v. 52, p. 125-149, 2010.

WELLER, J.I. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. **Biometrics**, Washington, v. 42, p. 627-640, 1986.

WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in Statistics - Simulation and Computation**, Zug, v. 22, p. 1079-1106, 1993.

WOUW, M. van de; HINTUM, T. van; KIK, C.; TREUREN, R. van; VISSER, B. Genetic diversity trends in twentieth century crop cultivars: a meta analysis. **Theoretical and Applied Genetics**, New York, v. 120, p. 1241-1252, 2010.

WU, R.; MA, C.X.; WU, S.S.; ZENG, Z.B. Linkage mapping of sex-specific differences. **Genetical Research**, London, v. 79, p. 85-96, 2002a.

WU, R.; MA, C.X.; PAINTER, I.; ZENG, Z.B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. **Theoretical Population Biology**, New York, v. 61, p. 349-363, 2002b.

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

YIM, Y.S.; DAVIS, G.L.; DURU, N.A.; MUSKET, T.A.; LINTON, E.W.; MESSING, J.W.; McMULLEN, M.D.; SODERLUND, C.A.; POLACCO, M.L.; GARDINER, J.M.; COE Jr., E.H. Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. **Plant Physiology**, Rockville, v. 130, p. 1686-1696, 2002.

ZENG, Z.B. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. **Proceedings of the National Academy of Science of the United States of America**, Washington, v. 90, p. 10972-10976, 1993.

ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, Bethesda, v. 136, p. 1457-1468, 1994.

ZENG, Z.B. **Statistical methods for mapping quantitative trait loci**. Raleigh: Department of Statistics, North Carolina State University, 2001. 128 p.

ZENG, Z.B., KAO, C.H.; BASTEN, C.J. Estimating the genetic architecture of quantitative traits. **Genetical Research**, London, v. 74, p. 279-289, 1999.