

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

**Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs
em progênie de irmãos completos, com aplicação em cana-de-açúcar**

Rodrigo Gazaffi

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

Piracicaba
2009

Rodrigo Gazaffi
Engenheiro Agrônomo

Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênie de irmãos completos, com aplicação em cana-de-açúcar

Orientador:
Prof. Dr. ANTONIO AUGUSTO FRANCO GARCIA

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

Piracicaba
2009

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

Gazaffi, Rodrigo

Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênie de irmão completos, com aplicação em cana-de-açúcar / Rodrigo Gazaffi. - - Piracicaba, 2009.
103 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2009.
Bibliografia.

1. Cana-de-açúcar 2. Genética estatística 3. Marcador Molecular 4. Melhoramento genético vegetal 5. Verossimilhança I. Título

CDD 633.61
G289d

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

À Maria Ines,
minha mãe;
se não fosse por ela,
não teria chegado tão longe...

DEDICO

Agradecimentos

A Deus (óbvio).

A Jó (aquele mesmo)

A CAPES pela bolsa no início do curso e, conseqüentemente, pela obrigatoriedade de fazer o PAE (uma das melhores experiências que tive durante todos esses anos na ESALQ).

Ao CNPq, pela bolsa de estudos.

A FAPESP, pelo financiamento do Projeto de Iniciação Tecnológica (PITE) que permitiu a obtenção dos dados usados no presente trabalho e em especial à coordenadora do mesmo, Dra. Anete Pereira de Souza.

Ao professor Antonio Augusto Franco Garcia, pela orientação, pela amizade e por ter me encontrado no meio de uma aula de Cálculo Diferencial e Integral, além da iniciação científica (a origem de tudo) e das ótimas conversas sobre futebol.

À trupe do laboratório de Genética-Estatística: Priscilla Karen Sabadin, Maria Marta Pastina, Marcelo Mollinari, Renato (que está chegando, mas já é espiritualmente do grupo), Gabriel Rodrigues Alves Margarido (os nomes estão na ordem cronológica em que os conheci, por isso ninguém precisa ficar com ciúmes por estar mais na frente ou mais no final da lista).

Em especial, para o Marcelo, por ter já feito uma dissertação no \LaTeX , o que facilitou absurdamente as coisas para mim e também pela solidariedade em fazer comigo curso de verão no IME/USP.

Ao Millor pela amizade e pelas galinhas (do mapeamento de QTLs) que também colaboraram muito na minha formação.

Ao André Boldrin Beltrame pela amizade, companheirismo e pela solidariedade nas disciplinas de graduação (melhor não entrar em maiores detalhes) e, óbvio, pelo ótimo show do Pearl Jam no Pacaembu.

Ao grande amigo José Manoel Colombari Filho, companheiro de disciplinas na pós-graduação.

Ao Marcelo Fernandes de Oliveira pelo notebook; sempre agradei sua valiosa ajuda na compra desse brinquedinho, principalmente quando acabava a força e o mesmo continuava funcionando firme e forte na bateria.

À CamiloxLelis, Renato, Érika, Rafael, Ivan Moura, Lúcia, Karina, Ivan Pezoa pela amizade e pelas experiências que vivi e aprendi com cada um.

À Dra. Heloisa que me ajudou demais nesse período final, ouvindo todas as minhas besteiras e devaneios, além dos 'ansiolíticos' e 'agitolíticos'.

Pode parecer estranho, mas também agradeço a todas as coisas ruins que aconteceram comigo, pois com elas aprendi a dar o valor exato para as coisas boas.

Aos familiares, avô, tio, tia, tios-avós, primos, agregados, gato, cachorro e papagaios...

As professoras Araci e Janete, excelentes professoras de matemática e biologia (respectivamente), com quem aprendi a gostar dessa duas áreas... e não é que estou estudando algo que permite unir essas duas coisas! É só dar uma olhadinha na página 36 que será fácil entender o que estou falando.

À todos que se propuseram a escrever uma apostila sobre L^AT_EX, porque escrever tudo o que você verá daqui para frente, no 'wordão da vida' não dá...

Ao pessoal do João Batista do Departamento de Ciências Florestais que apesar de não conhece-los, criaram o 'thesisllystyle.cls' e o divulgaram na página do laboratório na internet; isso ajudou muito a criar coragem e escrever a tese no L^AT_EX.

Ao pessoal do Emacs (agora entendi a expressão: 'poderoso processador de texto').

A todo pessoal do R (afinal colocar uma análise de regressão e formatar gráficos, tudo dentro de um único `for`, não é tarefa para qualquer programa estatístico). Agradeço também os pioneiros que desenvolveram os pacotes: *MASS*, *stats*, *car*, *xtable*, *nlme*, *lme4*, pois deram contribuição substancial. E não me esquecendo do mais importante de todos (JABÁ), o importantíssimo e substancialíssimo pacote *OneMap* com todos os 'íssimos' acrescidos nos adjetivos que indicam qualidades positivas. Afinal, o software que permitiu o cálculo das probabilidades multiponto desse trabalho, pois sem elas alguns resultados não teriam sido possíveis.

Por fim, mas não menos importante por estarem no final da página, agradeço aos dois que estão comigo, ao Seo Antonio '...quando faz... faz bem feito...', e a minha mãe, que não preciso falar nada qualquer coisa que eu diga não seria justo para expressar o meu agradecimento.

Great power comes with great responsibilities

(Ben Parker; personagem de Stan Lee)

O plantio é facultativo, mas a colheita é obrigatória

(Desconheço a citação correta)

Sumário

RESUMO	9
ABSTRACT	10
1 INTRODUÇÃO	11
2 REVISÃO BIBLIOGRÁFICA	14
2.1 Aspectos gerais da cana-de-açúcar	14
2.2 Mapeamento de QTLs	15
2.2.1 Mapeamento de QTLs em populações experimentais	16
2.2.2 Mapeamento de QTLs em progênies de irmão completos	17
2.2.3 Abordagem de Lin et al. (2003)	19
2.3 Mapeamento de QTLs em cana-de-açúcar	27
Referências	30
3 DESENVOLVIMENTO DE MODELO GENÉTICO-ESTATÍSTICO PARA MAPEAMENTO DE QTLs EM UMA POPULAÇÃO DE IRMÃOS COMPLETOS	41
Resumo	41
Abstract	42
3.1 Introdução	43
3.2 Metodologia	45
3.2.1 Modelo estatístico	45
3.2.2 Verossimilhança e estimadores	47
3.2.3 Testes de Hipóteses	49
3.2.4 Fases de ligação e segregação dos QTLs	49
3.3 Simulação	53
3.4 Resultados	55
3.5 Discussão	61
3.6 Conclusão	63
Referências	64
4 MAPEAMENTO DE QTLs EM CANA-DE-AÇÚCAR UTILIZANDO MAPEAMENTO POR INTERVALO COMPOSTO E MAPA GENÉTICO INTEGRADO	71

Resumo	71
Abstract	72
4.1 Introdução	73
4.2 Materiais e Métodos	75
4.2.1 População de mapeamento e medidas fenotípicas	75
4.2.2 Marcadores moleculares e construção do mapa de ligação	76
4.2.3 Análise de marcas individuais	77
4.2.4 Mapeamento por intervalo composto	77
4.3 Resultados	79
4.3.1 Dados fenotípicos	79
4.3.2 Marcadores analisados individualmente	80
4.3.3 Mapeamento por intervalo composto	82
4.4 Discussão	93
4.5 Conclusão	97
Referências	98

Resumo

Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênie de irmãos completos, com aplicação em cana-de-açúcar

A cana-de-açúcar é uma espécie de elevada importância econômica, uma vez que o país é o maior produtor mundial desta cultura. Atualmente, recebe importância maior devido as questões econômico-ambientais geradas pelo seu principal produto, o etanol. Neste contexto, há um cenário de potencial crescimento para cultura, em que o melhoramento genético pode contribuir com o desenvolvimento de cultivares mais produtivos e eficientes para diversas condições de cultivo. Os programas de melhoramento genético poderiam utilizar as informações obtidas com o mapeamento de QTLs para tornarem-se mais eficientes. Para tanto, faz-se necessário o desenvolvimento de abordagem que permita o mapeamento de QTLs em cana-de-açúcar com maior precisão. Os modelos genéticos-estatísticos desenvolvidos para o mapeamento de QTLs, normalmente são baseados em populações experimentais originárias de linhagens endogâmicas e espécies diploides. No entanto, para espécies como cana-de-açúcar, eucalipto, maracujá, a obtenção deste tipo de material é impraticável pela alta depressão por endogâmia existente. Dessa forma, estudos de mapeamento são conduzidos com uso de progênies de irmãos completos, normalmente fazendo adaptações para que as metodologias existentes possam ser usadas, o que induz a várias desvantagens. Assim, o trabalho apresentou dois objetivos: i) desenvolvimento de uma abordagem para mapeamento de QTLs em progênies de irmãos completos, utilizando mapa integrado e mapeamento por intervalo composto. ii) aplicação do presente modelo em caracteres relacionados a produtividade em cana-de-açúcar. Verificou-se que o novo modelo foi superior as abordagens já existentes na literatura, pois em estudos de simulação os QTLs foram detectados com maior poder estatístico e também com maior precisão. A aplicação do modelo em dados de cana-de-açúcar permitiu a identificação de maior número de QTLs do que a abordagem comumente utilizada para esta cultura. O maior poder estatístico desta metodologia contribuiu diretamente para detecção de novas regiões que apresentam associação com caracteres de produção. Vale destacar que o modelo detectou QTLs com diferentes padrões de segregação, além de indicar a provável existência de QTLs que se expressam ao longo dos cortes.

Palavras-chave: CIM; Modelos de mistura; Análises multiponto; Verossimilhança; Segregação; Fase de ligação; EST-SSR; EST-RFLP

Abstract

Development and application of genetic model for QTL mapping in a full sib family

Sugarcane has great economic importance to Brazil, which is world's largest producer. Currently, sugarcane receives worldwide attention due to the economic and environmental issues generated by its main product, ethanol. In this context, there is a scenario of potential growth for the culture, in which breeding can contribute to the development of new cultivars, more productive and effective to explore environmental conditions. Breeding programs could use the information obtained from QTL mapping to become more efficient. Thus, it is necessary to develop new statistical methods that allow QTL mapping in sugarcane with more precision. Statistical models developed for QTL mapping are usually based on experimental populations obtained from inbred lines. However, for species such as sugarcane, eucalyptus, passion fruit, inbred lines are not available and then mapping studies are conducted using full-sib families, usually making adaptations to existing methodologies, leading to several disadvantages since the assumptions are usually unrealistic. This work had two objectives: i) develop an approach for mapping QTLs in full-sibs, using integrated map and composite interval mapping. ii) use this model to map QTL for traits related to productivity in sugarcane. Results showed that the new model provided better results than existing approaches in literature. In simulation studies, all QTL were detected with more statistical power and precision. The application of the model on data from sugarcane has allowed the identification of more QTLs than single marker analysis, which is commonly used. The statistical power contributed to detection of new regions were associated with variation of quantitative traits. The model also detected QTLs with different patterns of segregation, as well as QTL expressed across harvests.

Keywords: CIM; Mixture models; Multipoint analysis; Likelihood; Segregation; Linkage phases; EST-SSR; EST-RFLP

1 INTRODUÇÃO

O Brasil destaca-se no cenário internacional como o maior produtor e exportador de álcool e açúcar. Atualmente, o álcool desperta um crescente interesse por ser uma fonte de energia renovável, sendo sua produção estruturada em culturas ricas em amido ou sacarose, com grande destaque para a cana-de-açúcar. A importância do mercado de etanol pode ser observado através da produção mundial que triplicou entre 2000 e 2007 chegando a 50 bilhões de litros, podendo chegar a 125 bilhões de litros em 2017. O potencial de crescimento do Brasil seria maior ainda, pois o volume de etanol produzido aumentaria de 17 bilhões de litros (2007) para 40,5 bilhões de litros (2017) (ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA AGRICULTURA E ALIMENTAÇÃO - FAO, 2008).

Este aumento estaria diretamente relacionado com o potencial para produção de etanol produzido a partir de cana-de-açúcar sem o comprometimento de áreas utilizadas para alimentação humana e animal, além de fatores climáticos e experiência no programa de produção de etanol a partir de cana-de-açúcar iniciado por volta de 1970. Neste contexto, o etanol brasileiro é o único biocombustível competitivo em relação a gasolina e ao óleo cru (FAO, 2008; GOLDENBERG, 2007).

Ao considerar um quadro de crescimento para a produção de cana-de-açúcar, infere-se que o melhoramento genético é um fator crítico nesse sistema, pois a produtividade pode ser condicionada as técnicas agrícolas, como correções de solo e aplicação de defensivos, ou através de fatores intrínsecos dos cultivares, tais como potencial de produção, resistência a pragas, doenças e estresses abióticos. Estes últimos somente podem ser modificados com o melhoramento genético.

A obtenção de novos cultivares em cana-de-açúcar é dificultado pelo fato das variedades modernas apresentarem uma alta complexidade genética, devido ao seu elevado nível de ploidia, com frequente aneuploidia (HEINZ; TEW, 1987), o que demanda muito tempo e esforço (SOUZA, 2003). Esta complexidade genômica é resultado de intensos cruzamentos interespecíficos entre espécies do gênero *Saccharum*, seguidos de intensa seleção com base em fenótipos favoráveis para as principais características de importância comercial e agrônômica (MATSUOKA; GARCIA; ARIZONO, 1999). Normalmente, estes caracteres apresentam influência

ambiental e também são controlados por muitos genes.

O desenvolvimento e utilização de marcadores moleculares possibilitaram o estudo dos caracteres quantitativos sob um outro aspecto, denominado mapeamento de QTLs (quantitative trait loci), o qual está fundamentado no estabelecimento de associações entre caráter quantitativo (fenótipo) e marcadores moleculares (genótipo). De forma geral, QTLs podem ser definidos como regiões cromossômicas que influenciam a variação de caracteres quantitativos (LYNCH; WALSH, 1998; LIU, 1998) e o mapeamento de QTLs é o procedimento que permite a estimação de posição, efeitos e interações (epistasia) entre diferentes regiões (ZENG; KAO; BASTEN, 1999), considerando um ou mais caracteres e ambientes simultaneamente. Em princípio, essa informação pode ser utilizada em programas de melhoramento genético para a realização da seleção assistida por marcadores (FALCONER; MACKAY, 1996). Dessa forma, mapear QTLs em cana-de-açúcar pode ser muito útil.

Várias metodologias para mapear QTLs foram desenvolvidas nos últimos anos, entre as quais destacam-se o mapeamento por intervalo (LANDER; BOTSTEIN, 1989) o mapeamento por intervalo composto, ou CIM (ZENG, 1994) e o mapeamento por múltiplos intervalos, ou MIM (KAO; ZENG; TEASDALE, 1999). Esses procedimentos comumente baseiam-se em populações experimentais provenientes de linhagens endogâmicas. No entanto, algumas espécies de grande importância econômica, como cana-de-açúcar, eucalipto, citros e maracujá não possuem linhagens endogâmicas à disposição. Nesse caso, as análises genético-estatísticas relacionadas à construção de mapas genéticos e ao mapeamento de QTLs baseiam-se em populações desenvolvidas pelo cruzamento biparental entre indivíduos não homocigóticos, gerando uma progênie segregantes com indivíduos que são irmãos completos.

Métodos para construção de mapas de ligação e o mapeamento de QTLs têm sido desenvolvidos para esse tipo de população por diversos autores (ARUS et al, 1994; JOHNSON; JANSEN; ARENDONK, 1999; LIN et al., 2003; RIDOUT et al., 1998; RITTER; SALAMINI, 1996; SONG; SOLLER; GENIZI, 1999), sendo que uma dificuldade inerente está na impossibilidade de caracterizar corretamente as fases de ligação existente entre marcadores e QTLs. Isto ocorre devido aos diversos padrões de segregação existentes nos cruzamentos e pelas diversas formas que os marcadores podem estar ligados. Lu, Cui e Wu (2004), Wu et al. (2002a), Wu et al. (2002b) propuseram abordagens para construção de mapas de ligação que permitem estimar corretamente a ligação e a ordem entre marcadores, assim como suas

respectivas fases de ligação, mapas desse tipo podem dar suporte ao mapeamento de QTLs.

Lin et al. (2003) expandiram os conceitos apresentados por Wu et al. (2002a) ao mapeamento de QTLs em progênies de irmãos completos. No entanto, existem limitações, uma vez que a abordagem considera apenas intervalos formados por dois marcadores adjacentes, sem utilizar métodos multiponto. Também foram consideradas como parâmetros as fases de ligação entre marcadores e QTLs, o que resulta em um modelo de misturas de distribuições tendo como variáveis latentes os genótipos dos QTLs e fases de ligação. Isso torna as análises extremamente complexas e ineficientes. Por considerar apenas um intervalo de marcadores para a realização do mapeamento de QTLs, tal abordagem apresenta as mesmas limitações observadas ao modelo proposto por Lander e Botstein (1989), a saber: i) a estatística do teste pode indicar a presença de um QTL num intervalo, mesmo que o QTL esteja localizado em intervalos adjacentes; isto ocorre pois os intervalos não são independentes; ii) se houver dois ou mais QTLs posicionados no mesmo cromossomo as estimativas para os efeitos aditivos e dominância são viesados; iii) o mapeamento considerando apenas a informação de dois marcadores não é eficiente, levando a redução do número de QTLs mapeados (ZENG, 1994). Assim, Zeng (1994) apresentou o mapeamento por intervalo composto, com modelo mapeando QTLs dentro de um intervalo formado por dois marcadores como cofatores no modelo visando controlar efeitos de outros QTL. Isso resultou em uma melhoria significativa no poder de mapeamento de forma que o CIM é atualmente o procedimento mais usado no mapeamento de plantas.

Esses fatos expostos contextualizam as pesquisas aqui apresentadas. O mapeamento de cana-de-açúcar tem sido estudado sob diferentes aspectos visando entender a arquitetura genética dos caracteres quantitativos, no entanto, para ser possível realizar o mapeamento de QTLs de forma satisfatória, faz-se necessário o desenvolvimento de uma metodologia de mapeamento semelhante à proposta por Lin et al. (2003), mas baseada em uma metodologia de mapeamento superior ao mapeamento por intervalo. O presente trabalho teve então dois objetivos: 1) expandir a abordagem proposta por Lin et al. (2003) para o contexto do mapeamento por intervalo composto; 2) mapear QTLs para produtividade, teor de açúcar e fibra em uma progênie de irmãos completos de cana-de-açúcar.

2 REVISÃO BIBLIOGRÁFICA

2.1 Aspectos gerais da cana-de-açúcar

A cana-de-açúcar pertence a família Poaceae e gênero *Saccharum*, cujo sistema de reprodução é predominantemente alogamia. Atualmente, considera-se a presença de seis espécies neste gênero: *Saccharum spontaneum* L. ($2n = 60 - 128$) e *S. robustum* Brandes e Jeswiet ex Grassl ($2n = 60 - 205$), *S. officinarum* L. ($2n = 80$), *S. barberi* Jeswiet ($2n = 81 - 124$), *S. sinense* Roxb ($2n = 111 - 120$), e *S. edule* Hassk ($2n = 60 - 80$). As espécies do gênero *Saccharum*, juntamente com *Erianthus* sect. *Ripidium*, *Sclerostachya* e *Narenga*, formam o chamado complexo *Saccharum*, isto é, um grupo de inter cruzamento muito próximo, o qual originou a cana-de-açúcar. No entanto, estudos moleculares e morfológicos apontariam a existência de duas espécies *S. officinarum* e *Saccharum spontaneum* (IRVINE, 1999).

S. officinarum é uma espécie poliploide com centro de diversidade localizado na Nova Guiné. Esta espécie também é denominada 'cana-nobre', devido ao alto conteúdo de açúcar e boas características gerais de industrialização. *S. espontaneum* é uma espécie altamente polimórfica e apresenta alta adaptabilidade, podendo ser encontrada em diversas condições ambientais. Em relação ao melhoramento, é a espécie que contribui para inclusão de resistência a pragas e doenças e perfilhamento, dentre outras características agrícolas. As demais espécies já foram utilizadas em programas de melhoramento, mas atualmente são apenas consideradas como fonte de variabilidade genética (MATSUOKA; GARCIA; ARIZONO, 1999).

A obtenção dos primeiros cultivares melhorados ocorreu com a produção de híbridos interespecíficos entre *S. officinarum* e outras espécies do gênero *Saccharum*. Para a recuperação das propriedades apresentadas em *S. officinarum* (genitor feminino) várias gerações de retrocruzamentos foram conduzidos até a obtenção do material de interesse, no processo denominado por nobilização. No entanto, verificou-se que nos retrocruzamentos a contribuição genética entre os genitores não era proporcional, isto é, o genitor materno apresentava maior contribuição para o número de cromossomos nestes novos materiais. O resultado foi a obtenção de híbridos com alto níveis de ploidia e também aneuploides (HEINZ; TEW, 1987).

Atualmente, o melhoramento genético baseia-se em processos de seleção e clonagem de genótipos superiores de populações segregantes, obtidas através de cruzamentos entre indiví-

duos diferentes, pois a cana-de-açúcar apresenta alta capacidade de propagação vegetativa e o florescimento ocorre em condições climáticas específicas. Nas etapas iniciais as seleções não apresentam alta intensidade, pois há poucas repetições para cada genótipo, o que torna as medidas fenotípicas muito sujeitas à influência ambiental, comprometendo a seleção de materiais promissores. Portanto, a liberação de uma nova variedade somente ocorre após diversos experimentos, em diferentes locais e anos de cultivo, o que em média demandaria 10 anos para a obtenção de novo cultivar (MATSUOKA; GARCIA; ARIZONO, 1999; MATSUOKA et al., 1999).

Assim, o desenvolvimento de métodos que possibilitem uma seleção mais eficiente ou precoce, otimizaria os programas de melhoramento genético, incrementando o ganho genético mais rapidamente. O mapeamento de QTL's pode ser uma importante ferramenta, pois permitiria o conhecimento da arquitetura genética dos caracteres quantitativos, auxiliando na condução de seleção assistida por marcadores (DEKKERS; HOSPITAL, 2002).

2.2 Mapeamento de QTLs

Com o desenvolvimento dos marcadores moleculares, métodos genético-estatísticos e ferramentas computacionais, tornou-se possível detalhar a herança e arquitetura genética dos caracteres complexos, ou seja, estudar regiões do genoma que controlam caracteres quantitativos (QTL - quantitative trait loci) (FALCONER; MACKAY, 1996; MACKAY, 2001).

Ao longo dos anos, várias metodologias foram propostas para a detecção e estimação dessas regiões, o que permitiu um acúmulo considerável de informações sobre base genética de caracteres quantitativos de importância agrônômica para as principais culturas, como milho (CARDINAL et al., 2001; MANGOLIN et al., 2004; SABADIN et al., 2008; SIBOV et al., 2003; ZHAO et al., 2006; WASSOM, 2008a, 2008b), soja (CHAPMAN et al. 2003; LI; PFEIFFER; CORNELIUS, 2008; TASMA et al. 2001), tomate (CHEN et al. 1999; GORGUET et al., 2008; MAZZUCATO et al., 2008), dentre outras. Entretanto, ao observar as propriedades biológicas dos materiais em estudo, os trabalhos podem ser classificados em dois grupos: i) mapeamento de QTLs para populações experimentais (com linhagens endogâmicas); ii) mapeamento em populações obtidas com cruzamento de indivíduos não endogâmicos, como por exemplo as progênes de irmãos completos (LIN et al., 2003).

O mapeamento de QTLs em populações experimentais caracteriza-se pelo desenvolvi-

mento de linhagens endogâmicas e obtenção de populações segregantes, tais como populações F_2 ou de retrocruzamento. Uma vez que a segregação dos marcadores é conhecida e a presença de desequilíbrio de ligação é maximizada nesses delineamentos, associações entre QTLs e marcadores podem ser estabelecidas com sucesso. O desenvolvimento de modelos para mapeamento de QTL's neste tipo de população é relativamente simples, pois há apenas dois alelos segregando em cada loco e as frequências alélicas e as fases de ligação entre marcador e QTL's são conhecidas. Por outro lado, o mapeamento de QTL's também pode ser feito para progênies de irmãos completos. Neste caso, a população segregante apresenta marcadores com diversos padrões de segregação, o número de QTLs varia de um a quatro (em diploides) e a fase de ligação entre os marcadores é desconhecida, assim como entre QTL e marcas (LIN et al., 2003).

2.2.1 Mapeamento de QTLs em populações experimentais

Diversos autores apresentam detalhes sobre como o mapeamento de QTLs é realizado nessas situações (BROMAN, 2001; DOERGE, 2002; DOERGE; ZENG; WEIR, 1997; FALCONER; MACKAY, 1996; JANSEN, 2000; LIU, 1998; LYNCH; WALSH, 1998; MACKAY, 2001). De forma geral, para a realização do mapeamento é necessário uma população que apresente variabilidade genética e elevado desequilíbrio de ligação. Populações experimentais do tipo F_2 , retrocruzamentos e linhagens puras recombinantes são utilizadas neste tipo de estudo (DOERGE, 2002). Esta escolha depende das características da espécie estudada, tais como, sistema de reprodução, ciclo de vida da geração, etc. A análise dos dados obtidos pode ser realizada através de várias metodologias que abrangem desde técnicas simples, como análise de marcas individuais, ou até modelos que incluem múltiplas regiões do genoma simultaneamente (JANSEN; STAM, 1994; KAO; ZENG; TEASDALE, 1999; ZENG, 1994).

O mapeamento de QTLs, utilizando análise de marcas individuais baseia-se na comparação entre as médias das classes genotípicas dos marcadores, o que pode ser realizado através de testes estatísticos, como teste t , análise de variância, regressão linear simples (LYNCH; WALSH, 1998) e teste da razão de verossimilhança (WELLER, 1986). As principais vantagens desses métodos são a sua simplicidade e rapidez de execução, e podem ser facilmente estendidos para múltiplos locos, não sendo necessário ter um mapa de ligação para essas análises. As principais desvantagens são: não distinção entre posição e efeitos dos QTLs;

necessidade de descartar indivíduos com genótipo perdido para um determinado marcador; redução do poder do teste quando os marcadores não saturam completamente o genoma, uma vez que o QTL pode estar a uma distância considerável do marcador (BROMAN, 2001).

O mapeamento por intervalo foi apresentado por Lander e Botstein (1989). O método considera um par de marcas adjacentes para inferir se há um QTL contido nesta região. Para tanto, é necessário o conhecimento do mapa de ligação. Este modelo é considerado como um modelos de misturas, o que permitiu o posterior desenvolvimento de metodologias mais elaboradas, como o mapeamento por intervalo composto univariado (ZENG, 1993, 1994), multivariado (JIANG; ZENG, 1995) e mapeamento de múltiplos caracteres (KAO; ZENG; TEASDALE, 1997; KAO; ZENG; TEASDALE, 1999) ou ainda abordagem de quadrados mínimos (HALEY; KNOTT, 1994; EEUWIJK et al., 2005). Uma característica dos métodos derivados do mapeamento por intervalo é a busca contínua de QTL ao longo do genoma, através da obtenção de curvas para a estatística utilizada para os testes de hipóteses, resultando na detecção de regiões significativas dentro de grupos de ligação (DOERGE, 2002; LYNCH; WALSH, 1998).

O mapeamento por intervalo composto (JANSEN; STAM, 1994; ZENG, 1994), além de considerar os pares de marcadores adjacentes, também inclui marcadores localizados em outras regiões do genoma (cofatores), com intuito de controlar os efeitos de QTLs externos ao intervalo investigado (ZENG, 1994). A inclusão de marcas com efeitos significativos sobre o QTL do intervalo analisado diminui o resíduo do modelo, aumentando sensivelmente o poder de detecção de QTL's, além de melhorar consideravelmente a precisão das estimativas de efeito e posição. Contudo, a inclusão excessiva de vários marcadores como cofatores pode reduzir sensivelmente a precisão do método (LYNCH; WALSH, 1998), em função de superparametrizações.

2.2.2 Mapeamento de QTLs em progênie de irmão completos

Determinados grupos de plantas não possuem linhagens endogâmicas à disposição, em função de diversos fatores, tais como alta depressão por endogamia, elevado tempo para a obtenção dessas linhagens (longo intervalo entre gerações), ou devido a complexidade do genoma (como a cana-de-açúcar). Assim, o mapeamento de QTLs pode ser feito a partir do cruzamento entre indivíduos não aparentados, gerando progênie segregante de irmãos

completos.

Algumas particularidades ocorrem neste tipo de progênie. Por exemplo, devido ao fato dos locos não estarem em homozigose, diversos padrões de segregação podem ocorrer. Wu et al. (2002a) afirmam que é possível a ocorrência de até 18 tipos distintos de segregação, as quais podem ser completamente informativas (1:1:1:1) ou parcialmente informativas (1:2:1, 3:1 e 1:1). Com relação ao QTLs, a situação é mais complexa, pois o genótipo do QTL não é observável e o número de alelos pode variar ou interagir de forma a obter diferentes tipos de segregações, que devem ser inferidos, a partir dos dados experimentais. Ainda, o não conhecimento da fase de ligação entre QTL e marcador também dificulta o mapeamento, pois a fase de ligação é fundamental para determinar a origem dos alelos de interesse. O mapeamento de QTLs nessa situação normalmente é conduzido através de três abordagens: marcas analisadas individualmente, *duplo pseudo testcross* e mapeamento de QTLs em progênies de irmãos completos baseado em mapa integrado.

De forma similar ao realizado em populações experimentais, a análise individual de marcas permite a associação entre marcadores e fenótipos. Pode ser útil, para populações de irmãos completos, pois pode ser aplicada para marcadores com qualquer padrão de segregação. O estudo de epistasia é relativamente simples, porém apresenta baixo poder estatístico.

O *duplo pseudo-testcross* (GRATTAPAGLIA; SEDEROFF, 1994) propõem a utilização de marcadores com segregação 1:1 para a construção do mapa de ligação e o mapeamento de QTLs. Essencialmente, são construídos dois mapas de ligação, um para cada genitor, em função de qual deles possui o alelo heterozigoto para o loco do marcador. O resultado final é a obtenção de dois mapas, no qual cada um apresenta marcadores que segregam em cada genitor. Esta é uma abordagem muito utilizada na prática (AITKEN et al., 2008; AL-JANABI et al., 2007; BUNDOCK et al., 2008; CARNEIRO et al., 2002; CHEN et al., 2008; DOLIGEZ et al., 2006; GAUDET et al., 2008; LANTERI et al., 2006; LOPES et al., 2006; MYBURG et al. 2003; PIPERIDIS et al. 2008), pois permite o pronto uso das metodologias e softwares desenvolvidos para populações experimentais. A desvantagem está em não utilizar marcadores com outros tipos de segregação e não conseguir integrar em único mapa os marcadores que segregam para cada genitor. Com relação ao mapeamento de QTLs, somente é possível detectar sem viés QTLs que segreguem em um dos genitores e que não possuam efeito de dominância.

Apesar de haver muitas metodologias que visam a construção de mapas genéticos integrados (BUTCHER et al., 2002; LU; CUI; WU, 2004; WU et al., 2002a; WU et al., 2002b), não há muitos modelos permitindo o mapeamento de QTLs neste contexto (LIN et al., 2003). A seguir será detalhada a abordagem que foi usada como base para o desenvolvimento do novo modelo a ser apresentado.

2.2.3 Abordagem de Lin et al. (2003)

A abordagem apresentada por Lin et al. (2003) baseia-se em um cruzamento biparental entre indivíduos diploides, P e Q , cujos cromossomos homólogos são denominados por 1 e 2. Para um dado loco, considera-se que o cruzamento entre os genótipos P_1P_2 e Q_1Q_2 resultam em 4 possíveis arranjos de cromossomos na progênie, P_1Q_1 , P_1Q_2 , P_2Q_1 e P_2Q_2 . Wu et al. (2002a) mostram que é possível observar 18 tipos de segregações para um loco marcador, as quais podem estar agrupadas em sete classes (Tabela 1).

Supondo a existência de um QTL contido dentro de um intervalo definido por dois marcadores adjacentes e que a fase de ligação entre os marcadores seja fixa (mapa de ligação previamente conhecido), podem ser encontrados quatro possíveis configurações genotípicas, isto é, diferentes fases de ligação entre marcador e QTL (Figura 1).

Desconsiderando as possíveis fases de ligação entre marcadores e QTL, os genitores podem formar oito tipos de gametas haploides, cuja probabilidade de ocorrência de cada genótipo depende da fase de ligação. Por exemplo, ao considerar o arranjo Φ_{11} , as probabilidades para os oito tipos de gametas possíveis estão descritas na Tabela 2, em que g_{111} (por exemplo) indica o genótipo do gameta formado para o genitor P : $M_1^{k+1}P_1M_1^{k+1}$. Esses oito gametas são gerados para ambos os pais e podem se unir ao acaso, portanto é possível a formação de 64 tipos genotípicos distintos na progênie. Esses tipos podem estar expresso em uma matriz $G^{k(k+1)}$, em que $k(k+1)$ representa o intervalo formado pelos marcadores adjacentes k e $k+1$. As probabilidades de ocorrência de um determinado gameta, considerando apenas o genótipo dos marcadores (alelo do QTL não observável), podem ser obtidas e estão apresentadas na Tabela 3, sendo por exemplo $g_{1.1}$ o gameta formado considerando apenas o genótipo dos marcadores do loco k e $k+1$, para o genitor P : $M_1^{k+1}M_1^{k+1}$. Neste caso, é possível formar 16 tipos de genótipos distintos na progênie, os quais compõem um vetor $M^{k(k+1)}$. A matriz $G^{k(k+1)}$ e o vetor $M^{k(k+1)}$ estão apresentados na Tabela 4 para as quatro configurações

Tabela 1 – Possíveis configurações genóticas para marcadores moleculares, bandas observadas e progênie segregante (WU et al., 2002a)

Marcadores	Genitores			Progênie		
	Cruzamento	Genótipo Observado	Bandas observadas	Segregação		
<i>A</i>	1	$ab \times cd$	$ab \times cd$	ac, ad, bc, bd	1:1:1:1	
	2	$ab \times ac$	$ab \times ac$	a, ac, ba, bc	1:1:1:1	
	3	$ab \times co$	$ab \times c$	ac, a, bc, b	1:1:1:1	
	4	$ao \times bo$	$a \times b$	ab, a, b, o	1:1:1:1	
<i>B</i>	B_1	5	$ab \times ao$	$ab \times a$	$ab, 2a, b$	1:2:1
	B_2	6	$ao \times ab$	$a \times ab$	$ab, 2a, b$	1:2:1
	B_3	7	$ab \times ab$	$ab \times ab$	$a, 2ab, b$	1:2:1
<i>C</i>	8	$ao \times ao$	$a \times a$	$3a, o$	3:1	
<i>D</i>	D_1	9	$ab \times cc$	$ab \times c$	ac, bc	1:1
		10	$ab \times aa$	$ab \times a$	a, ab	1:1
		11	$ab \times oo$	$ab \times o$	a, b	1:1
		12	$bo \times aa$	$b \times a$	ab, a	1:1
		13	$ao \times oo$	$a \times o$	a, o	1:1
	D_2	14	$cc \times ab$	$c \times ab$	ac, bc	1:1
		15	$aa \times ab$	$a \times ab$	a, ab	1:1
		16	$oo \times ab$	$o \times ab$	a, b	1:1
		17	$aa \times bo$	$a \times b$	ab, a	1:1
		18	$oo \times ao$	$o \times a$	a, o	1:1

$$\begin{array}{l}
\Phi_{11} = \begin{array}{c} M_1^k \quad M_2^k \\ P_1 \quad P_2 \\ M_1^{k+1} \quad M_2^{k+1} \end{array} \times \begin{array}{c} N_1^k \quad N_2^k \\ Q_1 \quad Q_2 \\ N_1^{k+1} \quad N_2^{k+1} \end{array} \Rightarrow P_1Q_1, P_1Q_2, P_2Q_1, P_2Q_2 \\
\Phi_{12} = \begin{array}{c} M_1^k \quad M_2^k \\ P_1 \quad P_2 \\ M_1^{k+1} \quad M_2^{k+1} \end{array} \times \begin{array}{c} N_1^k \quad N_2^k \\ Q_2 \quad Q_1 \\ N_1^{k+1} \quad N_2^{k+1} \end{array} \Rightarrow P_1Q_2, P_1Q_1, P_2Q_2, P_2Q_1 \\
\Phi_{21} = \begin{array}{c} M_1^k \quad M_2^k \\ P_2 \quad P_1 \\ M_1^{k+1} \quad M_2^{k+1} \end{array} \times \begin{array}{c} N_1^k \quad N_2^k \\ Q_1 \quad Q_2 \\ N_1^{k+1} \quad N_2^{k+1} \end{array} \Rightarrow P_2Q_1, P_2Q_2, P_1Q_1, P_1Q_2 \\
\Phi_{22} = \begin{array}{c} M_1^k \quad M_2^k \\ P_2 \quad P_1 \\ M_1^{k+1} \quad M_2^{k+1} \end{array} \times \begin{array}{c} N_1^k \quad N_2^k \\ Q_2 \quad Q_1 \\ N_1^{k+1} \quad N_2^{k+1} \end{array} \Rightarrow P_2Q_2, P_2Q_1, P_1Q_2, P_1Q_1
\end{array}$$

Figura 1 – Possíveis fases de ligação (ϕ) entre marcadores e QTL. O QTL está contido num intervalo definido por duas marcas k e $k+1$, com fase de ligação conhecida. Os alelos para os marcadores são representados por M_1 e M_2 , N_1 e N_2 , originários dos genitores P e Q , respectivamente. Os alelos para os QTLs estão apresentados por P_1 , P_2 , Q_1 e Q_2 (LIN et al., 2003)

Tabela 2 – Tipos e frequência dos gametas formados pelos parentais ao considerar um intervalo com dois marcadores e um QTL para a fase de ligação Φ_{11} . As taxas de recombinações entre QTL e os locos k e $k + 1$ são, respectivamente, r_1 e r_2 (LIN et al., 2003)

Tipos de Gameta	Ocorrência do gameta
$g_{111} = \Pr(M_1^k P_1 M_1^{k+1}) = \Pr(N_1^k Q_1 N_1^{k+1})$	$\frac{1}{2}(1 - r_1)(1 - r_2)$
$g_{112} = \Pr(M_1^k P_1 M_2^{k+1}) = \Pr(N_1^k Q_1 N_2^{k+1})$	$\frac{1}{2}(1 - r_1)r_2$
$g_{121} = \Pr(M_1^k P_2 M_1^{k+1}) = \Pr(N_1^k Q_2 N_1^{k+1})$	$\frac{1}{2}r_1 r_2$
$g_{122} = \Pr(M_1^k P_2 M_2^{k+1}) = \Pr(N_1^k Q_2 N_2^{k+1})$	$\frac{1}{2}r_1(1 - r_2)$
$g_{211} = \Pr(M_2^k P_1 M_1^{k+1}) = \Pr(N_2^k Q_1 N_1^{k+1})$	$\frac{1}{2}r_1(1 - r_2)$
$g_{212} = \Pr(M_2^k P_1 M_2^{k+1}) = \Pr(N_2^k Q_1 N_2^{k+1})$	$\frac{1}{2}r_1 r_2$
$g_{221} = \Pr(M_2^k P_2 M_1^{k+1}) = \Pr(N_2^k Q_2 N_1^{k+1})$	$\frac{1}{2}(1 - r_1)r_2$
$g_{222} = \Pr(M_2^k P_2 M_2^{k+1}) = \Pr(N_2^k Q_2 N_2^{k+1})$	$\frac{1}{2}(1 - r_1)(1 - r_2)$

Tabela 3 – Tipos e probabilidades de ocorrência dos gametas formados pelos parentais ao considerar um intervalo com apenas dois marcadores para a fase de ligação Φ_{11} . A taxa de recombinações entre os locos é r (LIN et al., 2003)

Tipos de Gameta	Ocorrência do gameta
$g_{1.1} = \Pr(M_1^k M_1^{k+1}) = \Pr(N_1^k N_1^{k+1})$	$\frac{1}{2}(1 - r)$
$g_{1.2} = \Pr(M_1^k M_2^{k+1}) = \Pr(N_1^k N_2^{k+1})$	$\frac{1}{2}r$
$g_{2.1} = \Pr(M_2^k M_1^{k+1}) = \Pr(N_2^k N_1^{k+1})$	$\frac{1}{2}r$
$g_{2.2} = \Pr(M_2^k M_2^{k+1}) = \Pr(N_2^k N_2^{k+1})$	$\frac{1}{2}(1 - r)$

possíveis.

Com base nos valores existentes no vetor $M^{k(k+1)}$ e na matriz $G^{k(k+1)}$ pode-se obter uma matriz $H^{k(k+1)}$ que contém a probabilidade condicional dos genótipos dos QTLs de acordo com o Teorema de Bayes:

$$H^{k(k+1)} = G^{k(k+1)} \oslash M^{k(k+1)} \quad (1)$$

em que, \oslash representa a divisão elemento a elemento entre o vetor $M^{k(k+1)}$ e cada coluna da matriz $G^{k(k+1)}$.

A matriz $H^{k(k+1)}$ é formada por 16 linhas e 4 colunas e pode ser dividida em quatro vetores $h_{11}^{k(k+1)}$, $h_{12}^{k(k+1)}$, $h_{21}^{k(k+1)}$ e $h_{22}^{k(k+1)}$, que representam os possíveis pareamentos cromossômicos parentais na progênie, ou seja, cada vetor $h^{k(k+1)}$ contém a informação referente as probabilidades genotípicas, condicionados a uma determinada fase de ligação. Nota-se que esta abordagem é baseada em probabilidades usando apenas dois marcadores flanqueadores,

Tabela 4 – Vetor $M^{k(k+1)}$ e matriz $G^{k(k+1)}$, utilizados para cálculo das probabilidades condicionais do genótipo dos QTL contido em um intervalo definido pelos marcadores k e $k + 1$ (LIN et al., 2003)

Marcador		$G^{k(k+1)}$				
		ϕ_{11}	P_1Q_1	P_1Q_2	P_2Q_1	P_2Q_2
Genótipo	$M^{k(k+1)}$	ϕ_{12}	P_1Q_2	P_2Q_1	P_2Q_2	P_1Q_1
		ϕ_{21}	P_2Q_1	P_2Q_2	P_1Q_1	P_1Q_2
		ϕ_{22}	P_2Q_2	P_2Q_1	P_1Q_2	P_1Q_1
			$g_{11}^{k(k+1)}$	$g_{12}^{k(k+1)}$	$g_{21}^{k(k+1)}$	$g_{22}^{k(k+1)}$
1	$M_1^k N_1^k M_1^{k+1} N_1^{k+1}$	$g_{1.1}^2$	g_{111}^2	$g_{111}g_{121}$	$g_{121}g_{111}$	g_{121}^2
2	$M_1^k N_1^k M_1^{k+1} N_2^{k+1}$	$g_{1.1}g_{1.2}$	$g_{111}g_{112}$	$g_{111}g_{122}$	$g_{121}g_{112}$	$g_{121}g_{122}$
3	$M_1^k N_1^k M_2^{k+1} N_1^{k+1}$	$g_{1.1}g_{1.2}$	$g_{111}g_{112}$	$g_{112}g_{121}$	$g_{122}g_{111}$	$g_{121}g_{122}$
4	$M_1^k N_1^k M_2^{k+1} N_2^{k+1}$	$g_{1.2}^2$	g_{112}^2	$g_{112}g_{122}$	$g_{122}g_{112}$	g_{122}^2
5	$M_1^k N_2^k M_1^{k+1} N_1^{k+1}$	$g_{1.1}g_{2.1}$	$g_{111}g_{211}$	$g_{111}g_{221}$	$g_{121}g_{211}$	$g_{121}g_{221}$
6	$M_1^k N_2^k M_1^{k+1} N_2^{k+1}$	$g_{1.1}g_{2.2}$	$g_{111}g_{212}$	$g_{111}g_{222}$	$g_{121}g_{212}$	$g_{121}g_{222}$
7	$M_1^k N_2^k M_2^{k+1} N_1^{k+1}$	$g_{1.2}g_{2.1}$	$g_{112}g_{211}$	$g_{112}g_{221}$	$g_{122}g_{211}$	$g_{122}g_{221}$
8	$M_1^k N_2^k M_2^{k+1} N_2^{k+1}$	$g_{1.2}g_{2.2}$	$g_{112}g_{212}$	$g_{112}g_{222}$	$g_{122}g_{212}$	$g_{122}g_{222}$
9	$M_2^k N_1^k M_1^{k+1} N_1^{k+1}$	$g_{2.1}g_{1.1}$	$g_{211}g_{111}$	$g_{211}g_{121}$	$g_{221}g_{111}$	$g_{221}g_{121}$
10	$M_2^k N_1^k M_1^{k+1} N_2^{k+1}$	$g_{2.1}g_{1.2}$	$g_{211}g_{112}$	$g_{211}g_{122}$	$g_{221}g_{112}$	$g_{221}g_{122}$
11	$M_2^k N_1^k M_2^{k+1} N_1^{k+1}$	$g_{2.2}g_{1.1}$	$g_{212}g_{111}$	$g_{212}g_{121}$	$g_{222}g_{111}$	$g_{222}g_{121}$
12	$M_2^k N_1^k M_2^{k+1} N_2^{k+1}$	$g_{2.2}g_{1.2}$	$g_{212}g_{112}$	$g_{212}g_{122}$	$g_{222}g_{112}$	$g_{222}g_{122}$
13	$M_2^k N_2^k M_1^{k+1} N_1^{k+1}$	$g_{2.1}^2$	g_{211}^2	$g_{211}g_{221}$	$g_{221}g_{211}$	g_{221}^2
14	$M_2^k N_2^k M_1^{k+1} N_2^{k+1}$	$g_{2.1}g_{2.2}$	$g_{211}g_{212}$	$g_{211}g_{222}$	$g_{221}g_{212}$	$g_{221}g_{222}$
15	$M_2^k N_2^k M_2^{k+1} N_1^{k+1}$	$g_{2.1}g_{2.2}$	$g_{212}g_{211}$	$g_{212}g_{221}$	$g_{222}g_{211}$	$g_{221}g_{222}$
16	$M_2^k N_2^k M_2^{k+1} N_2^{k+1}$	$g_{2.2}^2$	g_{212}^2	$g_{212}g_{222}$	$g_{222}g_{212}$	g_{222}^2

ou seja, não utiliza as vantagens da abordagem multiponto.

Pode-se assumir que um indivíduo i da progênie recebe alelos u e v do QTL, oriundos dos genitores P e Q , respectivamente. As probabilidades condicionais para os genótipos dos QTLs neste indivíduo estão contidos nos vetores $h_{uv}^{k(k+1)}$, dada as possíveis fases de ligação. A probabilidade na qual uma possível fase ocorra nos genitores P e Q é denominada de p e q , respectivamente. Portanto, as probabilidades de ocorrência das quatro possíveis fases de ligação (associação-associação, associação-repulsão, repulsão-associação, repulsão-repulsão) são $\phi_{11} = pq$, $\phi_{12} = p(1 - q)$, $\phi_{21} = (1 - p)q$ e $\phi_{22} = (1 - p)(1 - q)$. É possível portanto definir um modelo de misturas, considerando como variáveis latentes os genótipos dos QTL e fases de ligação.

Assumindo que os valores fenotípicos y tem distribuição normal com média μ_{uv} e variância σ^2 , tem-se que $f_{uv}(y_i) = \frac{1}{2\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_i - \mu_{uv})^2}{2\sigma^2}\right]$. Logo, a função de verossimilhança é (LIN et al. 2003):

$$L(\Omega|y) = \prod_{i=1}^n \left[(h_{11i}\phi_{11} + h_{12i}\phi_{12} + h_{21i}\phi_{21} + h_{22i}\phi_{22})f_{11}(y_i) + \right. \\ (h_{12i}\phi_{11} + h_{11i}\phi_{12} + h_{22i}\phi_{21} + h_{21i}\phi_{22})f_{12}(y_i) + \\ (h_{21i}\phi_{11} + h_{22i}\phi_{12} + h_{11i}\phi_{21} + h_{12i}\phi_{22})f_{21}(y_i) + \\ \left. (h_{22i}\phi_{11} + h_{21i}\phi_{12} + h_{12i}\phi_{21} + h_{11i}\phi_{22})f_{22}(y_i) \right] \quad (2)$$

$$L(\Omega|y) = \prod_{i=1}^n \sum_{u=1}^2 \sum_{v=1}^2 \pi_{uvi} f_{uv}(y_i) \quad (3)$$

Nesse caso, $\Omega = (\mu_{uv}, \sigma^2, p, q)^T$ é o vetor de parâmetros para o modelo de misturas $\pi_{11i} = h_{11i}\phi_{11} + h_{12i}\phi_{12} + h_{21i}\phi_{21} + h_{22i}\phi_{22}$, $\pi_{12i} = h_{12i}\phi_{11} + h_{11i}\phi_{12} + h_{22i}\phi_{21} + h_{21i}\phi_{22}$, $\pi_{21i} = h_{21i}\phi_{11} + h_{22i}\phi_{12} + h_{11i}\phi_{21} + h_{12i}\phi_{22}$ e $\pi_{22i} = h_{22i}\phi_{11} + h_{21i}\phi_{12} + h_{12i}\phi_{21} + h_{11i}\phi_{22}$. Estimativas dos parâmetros contidos em Ω podem ser obtidas através do uso de algoritmos iterativos, como por exemplo o algoritmo EM (Expectation-maximization) (DEMPSTER; LAIRD; RUBIN, 1977).

O efeito dos possíveis QTLs pode ser decomposto em efeitos aditivos e efeitos de domi-

nância, através do seguinte modelo linear:

$$\mu_{uv} = \mu + \alpha_u + \beta_v + \delta_{uv}; \quad u, v = 1, 2 \quad (4)$$

sendo, μ : média geral; α_u e β_v : efeitos aditivos dos QTLs provenientes dos parentais P e Q , respectivamente; δ_{uv} : efeito da interação entre alelos (dominância).

Ao considerar todos os possíveis alelos e as possíveis combinações entre si, há quatro possíveis efeitos aditivos ($\alpha_1, \alpha_2, \beta_1$ e β_2 , dois em cada genitor), assim como quatro efeitos de dominância ($\delta_{11}, \delta_{12}, \delta_{21}$ e δ_{22}), os quais não são independentes e por isso não são estimáveis. Para contornar isso, Lin et al. (2003) fazem uma reparametrização, $\alpha = \alpha_1 = -\alpha_2, \beta = \beta_1 = -\beta_2$ e $\delta = \delta_{11} = -\delta_{12} = -\delta_{21} = \delta_{22}$, denotando

$$\hat{a} = \mathbf{D}^{-1}\hat{m} \quad (5)$$

sendo, $\hat{a} = (\mu, \alpha, \beta, \delta)^T$: vetor que contém as estimativas para média geral e os valores genéticos (aditivos e dominantes) para os QTLs; $\hat{m} = (\mu_{uv})_{4 \times 1}$ vetor com estimativas de máxima verossimilhança para as quatro médias do modelo de misturas, obtidos via algoritmo EM; \mathbf{D}^{-1} : matriz inversa de

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad (6)$$

As fases de ligação são inferida com base nas probabilidades ϕ_{uv} , que são funções dos parâmetro p e q . Assim, infere-se qual é a fase mais provável (maior valor de ϕ). A fase de ligação é importante para o reconhecimento da configuração do cruzamento dos genitores e também fundamental para das estimativas dos efeitos aditivos e de dominância, pois dependendo da fase de ligação o vetor \hat{m} assume diferentes configurações:

$$\begin{aligned}
\phi_{11}: \quad pq \quad \hat{m} &= \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{21} & \mu_{22} \end{bmatrix} \\
\phi_{12}: \quad p(1-q) \quad \hat{m} &= \begin{bmatrix} \mu_{12} & \mu_{11} & \mu_{22} & \mu_{21} \end{bmatrix} \\
\phi_{21}: \quad (1-p)q \quad \hat{m} &= \begin{bmatrix} \mu_{21} & \mu_{22} & \mu_{11} & \mu_{12} \end{bmatrix} \\
\phi_{22}: \quad (1-p)(1-q) \quad \hat{m} &= \begin{bmatrix} \mu_{22} & \mu_{21} & \mu_{12} & \mu_{11} \end{bmatrix}
\end{aligned}$$

Assim, os estimadores de máxima verossimilhança para p e q usando o algoritmo EM são:

$$p = \frac{\sum_{i=1}^N (\pi_{12}\pi_{22}\Delta_1 K_{i2} + \pi_{11}\pi_{21}\Delta_2 J_{i2})}{\sum_{i=1}^N [\pi_{12}\pi_{22}\Delta_1 (K_{i2} - K_{i1}) - \pi_{11}\pi_{21}\Delta_2 (J_{i2} - J_{i1})]}$$

$$q = \frac{\sum_{i=1}^N (\pi_{12}\pi_{22}\Delta'_1 K'_{i2} + \pi_{11}\pi_{21}\Delta'_2 J'_{i2})}{\sum_{i=1}^N [\pi_{12}\pi_{22}\Delta'_1 (K'_{i2} - K'_{i1}) - \pi_{11}\pi_{21}\Delta'_2 (J'_{i2} - J'_{i1})]}$$

$$\Delta_1 = q(h_{11} - h_{21}) + (1-q)(h_{12} - h_{22})$$

$$\Delta_2 = q(h_{12} - h_{22}) + (1-q)(h_{11} - h_{21})$$

$$K_{i1} = q(h_{21}\Pi_{i11} - h_{11}\Pi_{i21}) + (1-q)(h_{22}\Pi_{i11} - h_{12}\Pi_{i21})$$

$$K_{i2} = q(h_{11}\Pi_{i11} - h_{21}\Pi_{i21}) + (1-q)(h_{12}\Pi_{i11} - h_{22}\Pi_{i21})$$

$$J_{i1} = q(h_{22}\Pi_{i12} - h_{12}\Pi_{i22}) + (1-q)(h_{21}\Pi_{i12} - h_{11}\Pi_{i22})$$

$$J_{i2} = q(h_{12}\Pi_{i12} - h_{22}\Pi_{i22}) + (1-q)(h_{11}\Pi_{i12} - h_{21}\Pi_{i22})$$

$$\Delta'_1 = p(h_{11} - h_{12}) + (1-p)(h_{21} - h_{22})$$

$$\Delta'_2 = p(h_{21} - h_{22}) + (1-p)(h_{11} - h_{22})$$

$$K'_{i1} = p(h_{12}\Pi_{i11} - h_{11}\Pi_{i12}) + (1-p)(h_{22}\Pi_{i11} - h_{21}\Pi_{i12})$$

$$K'_{i2} = p(h_{11}\Pi_{i11} - h_{21}\Pi_{i12}) + (1-p)(h_{21}\Pi_{i11} - h_{22}\Pi_{i12})$$

$$J'_{i1} = p(h_{22}\Pi_{i21} - h_{21}\Pi_{i22}) + (1-p)(h_{12}\Pi_{i21} - h_{11}\Pi_{i22})$$

$$J'_{i2} = p(h_{21}\Pi_{i21} - h_{22}\Pi_{i22}) + (1-p)(h_{11}\Pi_{i21} - h_{12}\Pi_{i22})$$

A existência de um QTL em uma dada posição do genoma é testada usando o teste da razão de verossimilhança (LRT), cuja hipótese de nulidade (H_0) considera ausência de QTL no intervalo:

$$LRT = -2 \left[\log_e L_0(\mu = \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}, \sigma^2, p, q) - \log_e L_1(\hat{\Omega}) \right]$$

apresentando distribuição de χ^2 com 4 graus de liberdade, sendo L_0 e L_1 a verossimilhança do modelo sob H_0 e H_a (usando as estimativas obtidas), respectivamente. Por fim, para a

identificação de outros tipos de segregações (1:2:1 e 1:1), é necessário testar se $\alpha = \beta$ e $\alpha = 0$, ou ainda se $\beta = 0$.

2.3 Mapeamento de QTLs em cana-de-açúcar

Devido ao elevado grau de poliploidia apresentada pelos cultivares modernos de cana-de-açúcar, assim como a alta depressão por endogâmia, a obtenção de linhagens é praticamente inviável, sendo a estratégia para condução de estudos de mapeamento de QTLs também baseada na utilização de progênies de irmãos completos. A poliploidia verificada em cana-de-açúcar faz ainda surgir outros padrões de segregação não encontrados em espécies diploides, tais como 7:1, 7:2, 11:1, dentre outros (GRIVET et al., 1996; EDMÉ; GLYNN; COMSTOCK, 2006), como consequência do fato dos apresentarem diferentes doses para os alelos.

Neste contexto, Wu et al. (1992) propuseram a construção de mapas de ligação apenas considerando marcadores que segregavam nas proporções 1:1 e 3:1, pois são segregações resultados de pareamento de bivalentes mesmo em elevados níveis de ploidia. Nos poliploides a segregação 1:1 ocorre quando em um genitor há somente uma cópia do alelo no loco (alelo simplex) e não há cópia deste alelo no outro genitor (alelo nuliplex). O padrão 3:1 ocorre para os locos que ambos os genitores apresentam alelos simplex. Neste caso o alelo simplex em poliploides corresponderia a um alelo heterozigoto para as espécies diploides e o nuliplex corresponderia ao loco homozigoto nulo.

Desta forma, a construção dos mapas de ligação através da utilização de marcadores em dose única permitiram a construção de mapas de ligação e o mapeamento de QTLs através da abordagem do *duplo pseudo testcross*, o que permite a utilização de modelos e programas computacionais desenvolvidos para populações experimentais. Já os marcadores não posicionados nos mapas de ligação podem também ser comumente utilizados no mapeamento com a análise de marcas individuais (AITKEN; JACKSON; McINTYRE, 2005; AITKEN; JACKSON; McINTYRE, 2006; AITKEN et al. 2008; AL-JANABI et al. 2007; HOARAU et al., 2001; HOARAU et al., 2002; McINTYRE et al., 2005a; MING et al. 2002a; MING et al., 2002b; RABOIN et al., 2006). Porém, também há estudos que utilizam misturas de segregação para a construção de mapas de ligação integrados (GARCIA et al., 2006; OLIVEIRA et al., 2007).

Dentre os principais trabalhos para construção de mapas de ligação destacam-se os pri-

meiros mapas construídos em cana utilizando RFLP (da SILVA et al., 1993) e RAPD (AL-JANABI et al., 1993); posteriormente, estes dados foram re-analisados em um único mapa (da SILVA et al., 1995). Hoarau et al. (2001) construíram o maior mapa de ligação já publicado, com 295 indivíduos e 887 marcadores AFLP, com 5849 cM e cobrindo cerca de 1/3 do genoma do cruzamento considerado.

O mapeamento de QTLs em cana-de-açúcar fundamentalmente é feito para caracteres relacionados à produtividade, tais como produção de açúcar, peso, diâmetro, número de colmos, percentual de fibra e resistências a doenças (AITKEN et al. 2008; McINTYRE et al., 2005a, 2005b; MING et al., 2001, 2002b). A principal abordagem é *pseudo duplo testcross*, com análise de marcas individuais, o que possivelmente explique a dificuldade em usar tais informações em programas de seleção assistida. Alguns resultados serão apresentados a seguir.

Ao considerar o cruzamento entre Green German e IND81-146 Ming et al. (2001) construíram o mapa genético com marcadores RFLP (1:1) e mapearam QTLs através do mapeamento por intervalo e análise individual de marcas (primeiro corte). Para TPH foram encontrados 14 QTLs que explicavam 65,5% da variação fenotípica. Para o cruzamento biparental PIN84-1 e Muntok Java foram detectados 22 QTLs para TPH, os quais apresentavam $R^2 = 68,3\%$. Estes resultados são referentes ao primeiro corte com dados de campo e casa de vegetação.

Ming et al. (2002b) utilizando estes mesmos cruzamento e encontraram para Green German e IND81-146 43 QTLs distribuídos entre TPH (3), conteúdo de fibra (19), POL (2), cinzas (7), número de colmos (2), peso de colmos (10), com R^2 variando entre 18,4% (TPH) a 62,7% (peso de colmo). Para o cruzamento PIN84-1 e IND81-146 foram detectados sete QTLs para produção de açúcar, sendo R^2 com todos os marcadores de 30%. Foram também encontrados um QTL para conteúdo de Fibra ($R^2 = 7,0\%$), 12 QTLs para POL ($R^2 = 40,0\%$), 13 para conteúdo de cinzas ($R^2 = 41,4\%$), um para número de colmos ($R^2 = 6,1\%$) e 24 para peso de colmo ($R^2 = 37,8\%$).

McIntyre et al. (2005a, 2005b) realizaram o mapeamento de QTLs para locos de resistência a doenças. Para tanto, utilizaram o cruzamento biparental Q117 e 74C42, usando marcadores AFLP, RFLP e SSR (1:1 e 3:1) com marcas analisadas individualmente. Foram encontrados sete QTLs para podridão radicular, quatro QTLs para resistência à ferrugem marrom. Os QTLs explicavam cerca de 6 a 18% da variação fenotípica.

Aitken et al. (2008) utilizando marcadores AFLP e RFLP nas proporções 1:1 e 3:1,

realizaram a construção de dois mapas genéticos separados para a condução do mapeamento por intervalo composto e também realizaram a análise de marcas simples, avaliando três cortes. Foram mapeados QTLs para peso (18), altura (8), número (7) e diâmetro de colmos (19), além de seis QTLs para produtividade (TCH). Em todos os casos, os R^2 variaram entre 2% a 9%.

Referências

- 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.; 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.
- 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.
- ARUS, P., OLARTE, C.; ROMERO, M.; VARGAS, F. Linkage analysis of 10 isozyme genes in F_1 segregating almond progenies, **Journal of the American Society for Horticultural Science**, Geneva, v. 119, p. 339-344, 1994.
- BROMAN, K.W. Review of statistical methods for QTL mapping in experimental crosses. **Lab Animal**, New York, v. 30, n. 7, p. 44-52, 2001.
- BUNDOCK, P.C.; POTTS, B.M.; VAILLANCOURT, R.E.; Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*, **Tree Genetics & Genomes**, Berlin, v. 4, p. 85-95, 2008.

BUTCHER, A.; WILLIAMS, R.; WHITAKER, D.; LING, S.; SPEED, P.; MORAN F.

Improving linkage analysis in outcrossed forest trees - an example from *Acacia mangium*, **Theoretical and Applied Genetics**, New York, v. 104, p. 1185-1191, 2002.

CARDINAL, A.J.; LEE, M.; SHAROPOVA, N. WOODMAN-CLIKEMAN, W.L.; LONG,

M.L. Genetic Mapping and Analysis of Quantitative Trait Loci for Resistance to Stalk Tunneling by the European Corn Borer in Maize, **Crop Science**, Madison, v. 41, p. 835-845, 2001.

CARNEIRO, M.S.; CAMARGO, L.E.A.; COELHO, A.S.G.; VENCOVSKY, R.; LEITE,

R.P.; STENZEL, N.M.; VIEIRA, M.L.C. RAPD-Based Genetic Linkage Maps of Yellow Passion Fruit (*Passiflora edulis* Sims. f. *flavicarpa* Deg.), **Genome**, Ottawa, v. 45, p. 670-678, 2002.

CHAPMAN, A.; PANTALONE, V.R.; USTUN, A.; ALLEN, F.L.; LANDAU-ELLIS, D.;

TRIGIANO, R.N.; GRESSHOFF, P.M. Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. **Euphytica**, Wageningen, v. 129, p. 387-393, 2003.

CHEN, F.Q.; FOOLAD, M.R.; HYMAN, J.; CLAIR, D.A. ST.; BEELAMAN, R.B.

Mapping of QTL's for lycopene and other fruit traits in a *Lycopersicon esculentum* x *L. pimpinellifolium* cross and comparison of QTL's across tomato species. **Molecular Breeding**, New York, v. 5, p. 283-299, 1999.

CHEN, C.; KIM D. BOWMAN, K.D.; CHOI, Y.A.; DANG, P.M.; RAO, M.N.; HUANG, S.,

SONEJI, J.R.; McCOLLUM, T.G.; GMITTER, F.G.Jr. EST-SSR genetic maps for *Citrus sinensis* and *Poncirus trifoliata*, **Tree Genetics & Genomes**, Berlin, v. 4, p. 1-10, 2008.

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; RUBIN, D.B. Maximum Likelihood from Incomplete Data via EM algorithm, **J. R. Stat. Soc. Ser. B. Methodol**, London, v. 39, 1-38, 1977.

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

DOERGE, R.W.; ZENG, Z.B.; WEIR, B.S. Statistical issues in the search for genes affecting quantitative traits in experimental populations, **Statistical Science**, Hayward, v. 12, p. 195-219, 1997.

DOLIGEZ, A; ADAM-BLONDON, A.F.; CIPRIANI, G.; GASPERO, G. DI; LAUCOU, V.; MERDINOGLU, D.; MEREDITH, C.P.; RIAZ, S.; ROUX, C.; THIS, P. An integrated SSR map of grapevine based on five mapping populations, **Theoretical and Applied Genetics**, New York, v. 113, p. 369-382, 2006.

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

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to Quantitative Genetics**. 4th ed. Essex, UK: Longman, 1996. 464 p.

FAO. **Agricultural Outlook 2008-2017**. Disponível em:

<www.fao.org/es/ESC/en/15/120/highlight_123_p.html>. Acessado em: 25 jan. 2009.

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.

GAUDET, M.; JORGE, V.; PAOLUCCI, I.; BERITOGNOLO, I.;

SCARASCIA-MUGNOZZA, G.; SABATTI, M. Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait, **Tree Genetics & Genomes**, Berlin, v. 4, p. 25-36, 2008.

GOLDEMBERG; J. Ethanol for a sustainable energy future. **Science**, New York, v. 315, p. 808-810, New York, 2007

GORGUET, B.; EGGINK, P.M.; OCAÑA, J., TIWARI, A., SCHIPPER, D.; FINKERS, R.; VISSER, R.G.F.; HEUSDEN, A.W. van. Mapping and characterization of novel parthenocarpic QTLs in tomato, **Theoretical and Applied Genetics**, New York, v. 116, p. 755–767, 2008.

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.

HALEY, C.S.; KNOTT, S.A.; ELSEN J.M. Mapping quantitative trait loci in crosses between outbred lines using least squares. **Genetics**, Baltimore, v.136, p. 195–207, 1994.

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

- HOARAU, J.Y.; GRIVET, L.; OFFMAN, B.; RABOIN, L.M.; DIORFLAR, J.P.; PAYET, J.; HELLMAN, M.; D'HONT, A.; GLASZMANN, J.C. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTL's for yield components, **Theoretical and Applied Genetics**, New York, v. 105, p. 1027-1037, 2002.
- HOARAU, J.Y.; OFFMAN, B.; D'HONT, A.; RISTERUCCIO, 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.
- IRVINE, J.E. *Saccharum* species as horticultural classes, **Theoretical and Applied Genetics**, New York, v. 98, p. 186-194, 1999.
- JANSEN, R.C. Quantitative trait loci in inbred lines. In: BALDING, D.J.; BISHOP, M. CANNINGS, C. **Handbook of Statistical Genetics**. New York, NY: John Wiley & Sons, 2000. p. 567-597
- JANSEN, R.C.; STAM, P. 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., R. C. JANSEN, and J. A. M. VAN ARENDONK. Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach, **Genetical Research**, London, v. 73, p. 75-83, 1999.
- KAO, C.H.; ZENG, Z.B. General formulae 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.

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

LANTERI, S; ACQUADRO, A.; COMINO, C.; MAURO, R.; MAUROMICALE, G.; PORTIS, E. A first linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus* L.) based on AFLP, S-SAP, M-AFLP and microsatellite markers, **Theoretical and Applied Genetics**, New York, v. 112, p. 1532-1542, 2006.

LI, D.; PFEIFFER, T.W.; CORNELIUS, P.L. Soybean QTL for Yield and Yield Components Associated with Glycine soja Alleles, **Crop Science**, Madison, v. 48, p. 571–581, 2008.

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.

LOPES, R; LOPES, M.T.G.; CARNEIRO, M.S.; MATTA, F.P.; CAMARGO, L.E.A.; VIEIRA, M.L.C. Linkage and mapping of resistance genes to *Xanthomonas axonopodis* pv. *passiflorae* in yellow passion fruit, **Genome**, Ottawa, v. 49, p. 17-29, 2006.

LU, Q.; CUI, Y; WU, R. A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. **BMC Genetics**, London, v. 5, p. 5-20, 2004.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits**. Sunderland: Sinauer Associates, Inc., 1998. 980p.

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

MANGOLIN, C.A.; SOUZA JÚNIOR, C.L. de; GARCIA, A.A.F.; GARCIA, A.F.; SIBOV, S.T.; SOUZA, A.P. de. Mapping QTLs for kernel oil content in a tropical maize population, **Euphytica**, Wageningen, v. 137, p. 251–259, 2004.

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

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.

MAZZUCATO, A.; PAPA, R.; BITOCCHI E.,; MOSCONI, P.; NANNI, L.; NEGRI, V.; PICARELLA, M.E.; SILIGATO, F.; SORESSI, G.P.; TIRANTI, B.; VERONESI, F. Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. **Theoretical and Applied Genetics**, New York, v. 116, p. 657-669, 2008.

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. 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.

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.

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

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

MYBURG, A.A.; GRIFFIN, A.R.; SEDEROFF, R.R.; Whetten, R.W. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F_1 hybrid based on a double pseudo-backcross mapping approach, **Theoretical and Applied Genetics**, New York, v. 107, p. 1028-1042, 2003.

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

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.

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

RITTER, E. and F. SALAMINI. 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.

SABADIN, P.K.; SOUZA JR., C.L.; SOUZA, A.P.; GARCIA, A.A.F. QTL mapping for yield components in a tropical maize population using microsatellite markers, **Hereditas**, Lund, v. 145, p. 194-203, 2008.

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, 2003.

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.

TASMA, I.M.; LORENZEN, L.L.; GREEN, D.E.; SHOEMAKER, R.C. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. **Molecular Breeding**, New York, v. 8, p. 25-35, 2001.

ZENG, Z.B. Theoretical basis of precision mapping of quantitative trait loci, **Proceedings of the National Academy of Sciences 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.

ZHAO, M.; ZHANG, Z.; ZHANG, S.; LI, W.; JEFFERS, D.P.; RONG, T.; PAN, G. Quantitative Trait Loci for Resistance to Banded Leaf and Sheath Blight in Maize, **Crop Science**, Madison, v. 46, p. 1039–1045, 2006.

WASSOM, J.J.; MIKKELINENI, V.; BOHN, M.O.; ROCHEFORD, T.R. QTL for Fatty Acid Composition of Maize Kernel Oil in Illinois High Oil × B73 Backcross-Derived Lines, **Crop Science**, Madison, v. 48, p. 69-78, 2008a

WASSOM, J.J.; JEFFREY, C.W.; MARTINEZ, E.; KING, J.J.; DeBAENE, J. QTL Associated with Maize Kernel Oil, Protein, and Starch Concentrations; Kernel Mass; and Grain Yield in Illinois High Oil × B73 Backcross-Derived Lines, **Crop Science**, Madison, v. 48, p. 243–252, 2008b.

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

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, 2002a.

WU, R.; MA, C.X.; WU, S.S.; ZENG, Z.B. Linkage mapping of sex-specific differences, **Genet. Res.**, v. 79, p. 85-96, 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.

3 DESENVOLVIMENTO DE MODELO GENÉTICO-STATÍSTICO PARA MAPEAMENTO DE QTLs EM UMA POPULAÇÃO DE IRMÃOS COMPLETOS

Resumo

A maioria dos métodos genéticos-estatísticos para mapeamento de QTLs disponíveis na literatura foram basicamente desenvolvidos para populações experimentais, derivadas do cruzamento entre linhagens endogâmicas, como as populações F_2 e de retrocruzamentos. Para muitas espécies, como cana-de-açúcar, maracujá e eucalipto, tais populações não podem ser obtidas devido a características genéticas e/ou fisiológicas. Neste caso, a construção de mapas de ligação e o mapeamento de QTLs é conduzido a partir do cruzamento entre indivíduos não endogâmicos. Conseqüentemente, o número de alelos, os tipos de segregações e as fases de ligação entre QTLs e marcadores são variáveis entre locos. O objetivo do presente trabalho foi desenvolver uma abordagem para mapeamento de QTLs nesse tipo de espécie, baseada no mapeamento por intervalo composto (CIM), considerando um mapa genético integrado construído através de abordagem multiponto e com marcadores com diferentes padrões de segregação. A população segregante é obtida com cruzamentos biparentais de genitores não endogâmicos (progênie de irmãos completos). O modelo baseia-se em três contrastes ortogonais para estimar os efeitos aditivos (um em cada genitor) e de dominância para os alelos dos QTLs. São incluídos cofatores para controle da variação causada por QTLs fora do intervalo mapeado, o que aumenta o poder estatístico do mapeamento. Para a validação da metodologia foi realizado um estudo com simulação, sendo simulados 8 QTLs com diferentes efeitos e fases de ligação, posicionados num mapa genético com 60 marcadores posicionados com distância de 10 cM em 4 grupos de ligação (15 marcadores cada). Foi possível detectar de forma eficiente todos os QTLs simulados, bem como seus respectivos efeitos, suas fases de ligação com os marcadores e seus padrões de segregação. De forma geral, o modelo CIM mostrou-se superior a outras abordagens existentes, permitindo interpretar os resultados dos QTLs mapeados de uma forma bastante informativa para estudos de arquitetura genética dos caracteres quantitativos.

Palavras-chave: CIM; Modelos de mistura; Verossimilhança; Segregação; Arquitetura genética; Fase de ligação

DEVELOPMENT OF STATISTICAL-GENETICS MODEL FOR QTL MAPPING IN A FULL-SIB FAMILY

Abstract

Most statistical methods for QTL mapping available in the literature were primarily developed for experimental populations derived from the inbred lines, such as F_2 and backcrosses. For many species, like sugarcane, citrus and eucalyptus, such populations can not be obtained due to genetical or physiological limitations. For them, the construction of linkage maps and QTL mapping is commonly based on full-sib families. Consequently, the number of alleles, the patterns of segregation and the linkage phases between QTL and markers vary between loci. The objective of this study was to develop an approach for QTL mapping in full-sib families, based on mapping by composite interval (CIM) and using a multipoint genetic map with markers with different patterns of segregation. The model is based on three orthogonal contrasts to estimate additive (one in each parent) and dominant effects for the QTL. Cofactors are included to control the variation caused by QTL located outside the mapping interval, increasing the statistical power. To validate the model a simulation was carried out, having 8 simulated QTL with different effects and linkages phases, randomly positioned on a genetic map with 60 markers placed with a distance of 10 cM in 4 linkage groups (15 markers each). It was possible to detect all QTL simulated, their effects, linkage phases and patterns of segregation. In general, the CIM model was better than to other existing approaches, allowing to interpret the results of mapping in a very informative way, helping to understand the genetic architecture of quantitative traits.

Keywords: CIM; Mixture models; Likelihood; Segregation; Genetic architecture; Linkage phases

3.1 Introdução

O mapeamento de QTLs permite estudar a arquitetura genética dos caracteres quantitativos. Dentre as principais aplicações destes estudos, incluem-se a possibilidade de incorporar tais informações em programas de melhoramento genético, com intuito de torná-los mais eficientes. É possível, por exemplo, um melhor entendimento da base genética da correlação entre caracteres (JIANG; ZENG, 1995; MACKAY, 2001), da interação entre genótipos e ambientes (BOER et al. 2007; EEUWIJK et al., 2005; EEUWIJK; MALOSETTI; BOER, 2007; MALOSETTI et al., 2004), da base genética da heterose (GARCIA et al., 2008), bem como a determinação do valor genético dos indivíduos para realização de seleção assistida por marcadores (DEKKERS; HOSPITAL, 2002; KAO; ZENG; TEASDALE, 1999; ZENG; KAO; BASTEN, 1999).

Vários métodos estatísticos para mapeamento foram propostos, destacando-se o mapeamento por intervalo (LANDER; BOTSTEIN, 1989) e o mapeamento por intervalo composto, ou CIM (JANSEN; STAM, 1994; ZENG, 1993, 1994), além do mapeamento por múltiplos intervalos (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999). O primeiro propôs considerar o genótipo do QTL como variável latente, usando para as análises modelos de misturas, a partir de intervalos definidos por marcadores adjacentes. No CIM, buscou-se remover os efeitos de QTLs localizados fora da região mapeada, o que permitiu uma melhora significativa no poder do mapeamento. Modelos CIM são muito utilizados para mapeamento de QTLs em espécies vegetais com importância econômica, como milho (CARDINAL et al., 2001; MANGOLIN et al., 2004; SABADIN et al., 2008; SIBOV et al., 2003; WASSOM, 2008a; WASSOM, 2008b; ZHAO et al., 2006), trigo (ABATE; LIU; McKENDRY, 2008; MACCAFERRI et al., 2008), arroz (CHO et al., 2007; SEMAGN et al., 2007) e soja (LI; PFEIFFER; CORNELIUS, 2008). Entretanto, a maioria dos modelos estatísticos desenvolvidos até o momento consideram populações baseadas no uso de linhagens endogâmicas, tais como as populações F_2 , retrocruzamentos e linhagens endogâmicas recombinantes.

Para espécies que não possuem este tipo de material, tais como eucalipto, citros e cana-de-açúcar, a população para mapeamento pode ser obtida com cruzamento biparental entre indivíduos não-endogâmicos. Isso implica na presença de QTLs com diferentes padrões de segregação simultaneamente (1:1:1:1, 1:2:1, 3:1 e 1:1). Para realizar as análises genético-

estatístico, uma abordagem geralmente utilizada neste contexto é a construção do mapa de ligação e o mapeamento de QTLs usando os chamados *duplos pseudo-testcrosses* (GRATTA-PAGLIA; SEDEROFF, 1994). Para tanto, consideram-se apenas os locos marcadores com segregação 1:1, o que permite a obtenção de mapas de ligação separados para cada genitor, possibilitando a aplicação dos métodos de mapeamento de QTLs desenvolvidos para retrocruzamentos em cada um desses mapas. No entanto, tal abordagem não pode ser usada diretamente para mapas integrados usando marcadores com outros padrões de segregação (1:1:1:1, 1:2:1 e 3:1) e assim o mapeamento de QTLs nessas situações baseia-se em técnicas simples, como análise de variância ou regressão linear para cada marca individualmente. Isso leva à redução do poder estatístico de mapeamento, não explorando as vantagens dos métodos de mapeamento por intervalo.

Diversos trabalhos propõem a construção de mapas de ligação combinando marcadores com diferentes padrões de segregação. Ritter, Gebhardt e Salamini (1990), Ritter e Salamini (1996) e Maliepaard, Jansen e Ooijen (1997) propuseram métodos para determinação da fração de recombinação utilizando estimativas de dois pontos. Ridout et al. (1998) apresentaram a estimação das frações de recombinações com base em testes de três pontos. Wu et al. (2002a) e Lu, Cui e Wu (2004) desenvolveram abordagens baseadas em máxima verossimilhança para a estimação simultânea das frações de recombinações e fases de ligação entre marcadores. Butcher et al. (2002) e Wu et al. (2002b) apresentam métodos para construção de mapas de ligação através de modelos de máxima verossimilhança multiponto, utilizando cadeias de markov ocultas (HMM). A principal vantagem destes métodos em relação ao *duplo pseudo testcross* é a possibilidade de se construir mapas de ligação mais saturados e com maior representatividade do polimorfismo genético do cruzamento, já que marcadores com qualquer padrão de segregação podem ser usados. Conseqüentemente, melhores mapas são obtidos e há maior poder estatístico para detecção de QTLs.

Um número reduzido de métodos estatísticos foi desenvolvido para o mapeamento de QTLs neste contexto. Song, Soller e Genizi (1999) propuseram o mapeamento em populações de cruzamento aberto, com uso de delineamento análogo ao F_2 em populações endogâmicas. Johnson, Jansen e Arendonk (1999) estudaram o mapeamento de QTLs em progênies de meios-irmãos, com genotipagem seletiva, através de modelos de misturas. Lin et al. (2003) desenvolveram um procedimento de mapeamento por intervalo, considerando como parâme-

tros os efeitos e a posição dos QTLs, além da fase de ligação entre eles e com os marcadores. As principais limitações da abordagem de Lin et al. (2003) são: i) probabilidades a posteriori para os genótipos dos QTLs não baseadas em estimativas multiponto; ii) dificuldade de expansão do modelo para o contexto do CIM, que possui muitas vantagens em relação aos modelos sem inclusão de cofatores; iii) dificuldades computacionais em estimar as fases de ligação entre QTLs e marcadores usando o algoritmo *EM* com base no modelo de misturas.

O objetivo do presente trabalho foi desenvolver um modelo de mapeamento de QTLs baseado na abordagem CIM, considerando progênie de irmãos completos e mapa genético multiponto obtido usando-se marcadores moleculares com diferentes padrões de segregação. O modelo proposto permite localizar os QTLs, determinar seu padrão de segregação e também inferir as fases de ligação entre QTLs e marcadores, permitindo análise simultânea de marcadores e QTLs com qualquer padrão de segregação (1:1:1:1, 1:2:1, 3:1 e 1:1).

3.2 Metodologia

3.2.1 Modelo estatístico

Considere uma progênie de irmãos completos, resultante do cruzamento entre dois genitores P e Q não aparentados, diplóides e com mapa genético conhecido (Figura 2). Para um intervalo definido pelos marcadores adjacentes m e $m+1$, com alelos 1 ou 2, tem-se que os genótipos destes locos podem ser representados genericamente por $P_m^{\{1,2\}}$, $P_{m+1}^{\{1,2\}}$, $Q_m^{\{1,2\}}$ e $Q_{m+1}^{\{1,2\}}$, em que $\{1, 2\}$ indicam os possíveis alelos de cada loco. Assumiu-se a presença de um QTL neste intervalo, cujos alelos são representados por P^1 e P^2 , para o genitor P e Q^1 e Q^2 , para o genitor Q . Assumiu-se que os alelos P^1 e Q^1 tem efeito positivo sobre o fenótipo. Tal cruzamento pode então ser representado por $P_m^1 P_m^1 P_{m+1}^1 / P_m^2 P_m^2 P_{m+1}^2 \times Q_m^1 Q_m^1 Q_{m+1}^1 / Q_m^2 Q_m^2 Q_{m+1}^2$.

A segregação dos QTLs na progênie desse cruzamento resulta na formação de quatro classes genotípicas ($P^1 Q^1$, $P^1 Q^2$, $P^2 Q^1$ e $P^2 Q^2$) com proporção 1:1:1:1. Portanto, é possível definir três contrastes ortogonais entre as médias genotípicas destas classes, de forma similar

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

Figura 2 – Representação do tipo de cruzamento considerado. P e Q são os genitores, $P_m^{\{1,2\}}$, $Q_m^{\{1,2\}}$, $P_{m+1}^{\{1,2\}}$ e $Q_{m+1}^{\{1,2\}}$ os alelos dos marcadores dos locos m e $m+1$, e P^1 , P^2 , Q^1 e Q^2 são os alelos dos QTLs

ao apresentado por Lin et al. (2003):

$$\begin{array}{c}
 P^1Q^1 + P^1Q^2 - P^2Q^1 - P^2Q^2 \\
 P^1Q^1 - P^1Q^2 + P^2Q^1 - P^2Q^2 \\
 P^1Q^1 - P^1Q^2 - P^2Q^1 + P^2Q^2
 \end{array}$$

Os dois primeiros contrastes referem-se aos efeitos aditivos dos QTLs nos genitores P e Q , respectivamente, e o terceiro à interação intraloco (dominância) entre os efeitos aditivos de cada genitor. Os coeficientes dos contrastes podem ser representados nas colunas da matriz \mathbf{D} (matriz de delineamento genético, de semelhante ao apresentado por Kao e Zeng, 1997):

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \quad (7)$$

Vale ressaltar que o genótipo do QTL não é diretamente observável, sendo inferido a partir do genótipo dos marcadores flaqueadores. No presente caso, assumiu-se que as probabilidades condicionais para os genótipos do QTL são multiponto e foram obtidas através do mapa genético reconstruído via HMM, usando uma versão modificada do software OneMap (MARGARIDO; GARCIA; SOUZA, 2007).

Com base nos contrastes, é possível definir um modelo estatístico para mapeamento:

$$y_j = \mathbf{Z}_j \boldsymbol{\gamma} + \alpha_p^* x_{pj}^* + \alpha_q^* x_{qj}^* + \delta_{pq}^* x_{pj}^* x_{qj}^* + \epsilon_j \quad (8)$$

em que, y_j : fenótipo do j -ésimo indivíduo ($j = 1, \dots, n$); \mathbf{Z}_j : j -ésima linha da matriz indicadora \mathbf{Z} , com dimensão $n \times (1+3c)$, que contém uma coluna de 1's e variáveis indicadoras dos genótipos dos c cofatores (marcadores moleculares fora da região sendo mapeada), de acordo com os contrastes em (7); $\boldsymbol{\gamma}$: vetor de parâmetros $(1+3c) \times 1$ contendo o intercepto (μ) e coeficientes de regressão linear múltipla (α_p , α_q e δ_{pq}) para os cofatores; α_p^* e α_q^* : efeitos aditivos para o QTL nos genitores P e Q , respectivamente; δ_{pq}^* : efeito da interação intra-loco (dominância) entre os efeitos aditivos; ϵ_j erro. Assumiu-se $\epsilon_j \sim N(0, \sigma^2)$. As variáveis x_{pj}^* e x_{qj}^* indicam os contrastes em função dos genótipos dos QTLs:

$$x_{pj}^* = \begin{cases} 1 & \text{se } P^1Q^1 \\ 1 & \text{se } P^1Q^2 \\ -1 & \text{se } P^2Q^1 \\ -1 & \text{se } P^2Q^2 \end{cases}; \quad x_{qj}^* = \begin{cases} 1 & \text{se } P^1Q^1 \\ -1 & \text{se } P^1Q^2 \\ 1 & \text{se } P^2Q^1 \\ -1 & \text{se } P^2Q^2 \end{cases}$$

Os elementos de \mathbf{Z} são definidos de forma análoga, referindo-se contudo aos marcadores, seguindo o que é preconizado para modelos CIM. Para a seleção dos cofatores podem ser empregados por exemplo, métodos de regressão múltipla entre fenótipos e marcadores, com a seleção de modelos utilizando o critério de informação bayesiano, ou BIC (SCHWARZ, 1978), até a seleção do modelo final com no máximo $2\sqrt{n}$ parâmetros para evitar superparametrizações (SAKAMOTO; KITAGAWA; ISHIGURO, 1986; WANG; BASTEN; ZENG, 2007). Para cada etapa pode ser feita a inclusão simultânea de três efeitos (α_p , α_q e δ_{pq}) para cada marcador incluído como cofator; posteriormente, podem ser removidos os efeitos não significativos, de forma a reduzir o número de parâmetros.

3.2.2 Verossimilhança e estimadores

A função de verossimilhança para o modelo em (8) é:

$$L(\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma) = \prod_{j=1}^n \left[\sum_{k=1}^2 \sum_{l=1}^2 p_{klj} \phi \left(\frac{y_j - \mu_{klj}}{\sigma} \right) \right] \quad (9)$$

em que, $\boldsymbol{\theta}$: vetor de parâmetros para os QTLs (α_p^* , α_q^* e δ_{pq}^*), p_{klj} : probabilidade *a priori* do QTL apresentar genótipo P^kQ^l para o j -ésimo indivíduo, $\phi(\cdot)$: função densidade de probabilidades da distribuição normal padrão; $\mu_{11j} = \mathbf{Z}_j\boldsymbol{\gamma} + \alpha_p^* + \alpha_q^* + \delta_{pq}^*$; $\mu_{12j} = \mathbf{Z}_j\boldsymbol{\gamma} + \alpha_p^* - \alpha_q^* - \delta_{pq}^*$; $\mu_{21j} = \mathbf{Z}_j\boldsymbol{\gamma} - \alpha_p^* + \alpha_q^* - \delta_{pq}^*$; $\mu_{22j} = \mathbf{Z}_j\boldsymbol{\gamma} - \alpha_p^* - \alpha_q^* + \delta_{pq}^*$.

De forma análoga ao proposto por Zeng (1994) e utilizando a notação semelhante à apresentada por Kao e Zeng (1997), para a obtenção das estimativas de máxima verossimilhança foi utilizado o algoritmo *EM* (DEMPSTER; LAIRD; RUBIN, 1977), em duas etapas:

Passo E: Cálculo das probabilidades *a posteriori* (π_{klj}) para o genótipo do QTL, de acordo com o Teorema de Bayes:

$$\pi_{klj} = \frac{p_{klj} \phi\left(\frac{y_j - \mu_{klj}}{\sigma}\right)}{\sum_{k=1}^2 \sum_{l=1}^2 p_{klj} \phi\left(\frac{y_j - \mu_{klj}}{\sigma}\right)}$$

Passo M: Obtenção das estimativas de máxima verossimilhança:

$$\begin{aligned} \hat{\alpha}_p^{*t} &= \frac{1}{\mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_1\#\mathbf{D}_1)} \left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^t)' \boldsymbol{\Pi}^t \mathbf{D}_1 - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_1\#\mathbf{D}_2) \alpha_q^{*t} - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_1\#\mathbf{D}_3) \delta_{pq}^{*t} \right] \\ \hat{\alpha}_q^{*(t+1)} &= \frac{1}{\mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_2\#\mathbf{D}_2)} \left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^t)' \boldsymbol{\Pi}^t \mathbf{D}_2 - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_2\#\mathbf{D}_1) \alpha_p^{*t} - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_2\#\mathbf{D}_3) \delta_{pq}^{*t} \right] \\ \hat{\delta}_{pq}^{*(t+1)} &= \frac{1}{\mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_3\#\mathbf{D}_3)} \left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^t)' \boldsymbol{\Pi}^t \mathbf{D}_3 - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_3\#\mathbf{D}_1) \alpha_p^{*t} - \mathbf{1}' \boldsymbol{\Pi}^t (\mathbf{D}_3\#\mathbf{D}_2) \alpha_q^{*t} \right] \\ \hat{\boldsymbol{\gamma}}^{(t+1)} &= (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' (\mathbf{y} - \boldsymbol{\Pi}^t \mathbf{D} \boldsymbol{\theta}^{(t+1)}) \\ \hat{\sigma}^{2(t+1)} &= (1/n) \left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^{(t+1)})' (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^{(t+1)}) - 2 (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}^{(t+1)})' \boldsymbol{\Pi}^t \mathbf{D} \boldsymbol{\theta}^{(t+1)} + \boldsymbol{\theta}'^{(t+1)} \mathbf{V}^t \boldsymbol{\theta}^{(t+1)} \right] \end{aligned}$$

em que:

$$\mathbf{V} = \begin{bmatrix} \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_1\#\mathbf{D}_1) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_1\#\mathbf{D}_2) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_1\#\mathbf{D}_3) \\ \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_2\#\mathbf{D}_1) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_2\#\mathbf{D}_2) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_2\#\mathbf{D}_3) \\ \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_3\#\mathbf{D}_1) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_3\#\mathbf{D}_2) & \mathbf{1}'\boldsymbol{\Pi}(\mathbf{D}_3\#\mathbf{D}_3) \end{bmatrix}$$

\mathbf{D}_1 , \mathbf{D}_2 e \mathbf{D}_3 são as colunas da matriz \mathbf{D} (7), $\boldsymbol{\Pi} = \{\pi_{klj}\}$, $n \times 4$: matriz de probabilidades

a posteriori dos genótipos dos QTLs; # representa o produto de Hadamard.

O algoritmo inicia-se com a atribuição de valores arbitrários na iteração t para parâmetros contidos em θ , o que permite o cálculo das probabilidades *a posteriori* no passo E ($t + 1$); as novas estimativas são então utilizadas para a re-estimação dos parâmetros do modelo de acordo com os estimadores obtidos no passo M . Tal procedimento é repetido até a convergência e obtenção das estimativas de máxima verossimilhança.

3.2.3 Testes de Hipóteses

O procedimento de busca por QTLs é realizado através da comparação da verossimilhança dos modelos considerando presença de possível QTL (H_a) versus modelo sem QTL (H_0):

$$H_0 : \alpha_p^* = \alpha_q^* = \delta_{pq}^* = 0$$

$$H_a : \text{pelo menos um efeito difere de zero}$$

Os testes de hipóteses podem ser feitos em todas as posições do genoma, via LOD Score ou teste da razão de verossimilhança, de forma análoga ao apresentado por Zeng (1994). Em função da realização de testes múltiplos é necessário adotar uma estratégia para controlar o erro tipo I . Uma opção apresentada por Vieira et al. (2000) consiste em determinar o número de testes independentes no genoma, com a aplicação da correção de Bonferroni para a obtenção de um nível de significância global. Outra alternativa muito utilizada é o uso de permutações, como proposto por Churchill e Doerge (1994).

3.2.4 Fases de ligação e segregação dos QTLs

Após a identificação dos QTLs, é possível fazer inferências a respeito das fases de ligação dos mesmos com base nos sinais das estimativas de α_p^* e α_q^* . Na definição dos contrastes adotou-se a configuração $P_m^1 P_{m+1}^1 / P_m^2 P_{m+1}^2 \times Q_m^1 Q_{m+1}^1 / Q_m^2 Q_{m+1}^2$, assumindo que os alelos P^1 e Q^1 causem aumento no fenótipo. Defini-se tal configuração como sendo de associação em ambos os genitores, o que implica em estimativas positivas para α_p^* e α_q^* (Figura 2). Caso outras configurações ocorram, simplesmente surgirão mudanças nos sinais das estimativas. No entanto, independente da configuração real, os contrastes definidos para o desenvolvimento do modelo sempre testam os mesmos efeitos, exceto para os sinais. Neste

contexto, a determinação da fase de ligação entre marcadores e QTL pode ser inferida de forma simples com a identificação dos alelos que promovem o aumento ou redução do fenótipo, não sendo portanto necessário incluir as fases de ligação no modelo, como feito por Lin et al. (2003). Isto reduz sensivelmente a complexidade do processo de estimação via algoritmo *EM*.

A segregação dos QTLs depende das relações de dominância entre os alelos. Para inferir, diversas hipóteses estatísticas apropriadas foram definidas (Tabela 5). Tais hipóteses devem ser testadas apenas nas posições para as quais QTLs foram mapeados numa primeira etapa, evitando dessa forma os problemas de realização de múltiplos testes, uma vez que mesmo nos casos mais difíceis, poucos testes complementares serão realizados.

Assim, pode ser estabelecidos um procedimento em duas etapas para mapear os QTLs, identificando seus efeitos, fase de ligação e padrão de segregação. Inicialmente, faz-se uma busca pelo genoma por possíveis QTLs, testando-se simultaneamente a significância dos efeitos α_p^* , α_q^* e δ_{pq}^* . Nas posições com evidências da presença de QTLs, testes complementares são conduzidos (H_{01} , H_{02} e H_{03}). Se apenas um efeito for significativo, conclui-se que há segregação 1:1 e as fases de ligação são facilmente inferidas. Caso mais de uma estimativa seja significativa, testes complementares são necessários e são realizados considerando os sinais das estimativas. Para as situações em que há dois efeitos estatisticamente significativos, testes sobre igualdade desses efeitos são realizados, sendo possível inferir se a segregação é 1:2:1 ou 1:1:1:1. Quando o QTL apresenta todos efeitos significativos faz-se necessário testar as três hipóteses complementares H_{04} , H_{05} e H_{06} (Tabela 5). Neste caso, se todas hipóteses não forem rejeitadas, conclui-se que a segregação é 3:1; nas situações em que apenas uma hipótese não for rejeitada, conclui-se que a segregação é 1:2:1. Nos demais casos, conclui-se que a segregação é 1:1:1:1. Em todos os casos, as fases de ligação são facilmente deduzidas a partir dos sinais das estimativas.

Para a condução dos testes de hipóteses da Etapa 2 (Tabela 5) é necessário a obtenção dos novos estimadores sob as restrições de cada hipótese considerada. Define-se então a matriz \mathbf{T} , que pós-multiplica \mathbf{D} , de forma que uma nova matriz \mathbf{DT} de contrastes seja obtida. Tal matriz substitue \mathbf{D} nos passos do algoritmo *EM* e novas estimativas de verossimilhança podem ser facilmente obtidas (Tabela 6).

Tabela 5 – Hipóteses estatísticas para inferências sobre segregações e fases de ligação dos QTLs

Etapa 1		Sinais das estimativas ⁽¹⁾		Etapa 2		Conclusões	
Hipóteses Rejeitadas		α_p^* , α_q^* , δ_{pq}^*		Hipóteses	Segregação	Fase ⁽²⁾	
H_{01} : $\alpha_p^* = 0$		$+\phi\phi$ ou $-\phi\phi$		–	1 : 1	P^1P^2 (+ $\phi\phi$) ou P^2P^1 (- $\phi\phi$) (a)	
H_{02} : $\alpha_q^* = 0$		$\phi+\phi$ ou $\phi-\phi$		–	1 : 1	Q^1Q^2 ($\phi+\phi$) ou Q^2Q^1 ($\phi-\phi$) (b)	
H_{03} : $\delta_{pq}^* = 0$		$\phi\phi+$ ou $\phi\phi-$		–	1 : 1	não determinada	
H_{01} e H_{02}		$++\phi$		H_{04} : $\alpha_p^* = \alpha_q^*$, rejeitada	1 : 1 : 1 : 1	$P^1P^2 \times Q^1Q^2$ (c)	
		$+-\phi$		H_{04} : $\alpha_p^* = -\alpha_q^*$, rejeitada	1 : 1 : 1 : 1	$P^1P^2 \times Q^2Q^1$ (d)	
		$-\phi\phi$		H_{04} : $-\alpha_p^* = \alpha_q^*$, rejeitada	1 : 1 : 1 : 1	$P^2P^1 \times Q^1Q^2$ (e)	
		$-\phi\phi$		H_{04} : $-\alpha_p^* = -\alpha_q^*$, rejeitada	1 : 1 : 1 : 1	$P^2P^1 \times Q^2Q^1$ (f)	
		$++\phi$; $+-\phi$; $-\phi\phi$; $-\phi\phi$		H_{04} não rejeitada	1 : 2 : 1	idem (c), (d), (e) e (f)	
H_{01} e H_{03}		$+\phi+$ ou $-\phi-$		H_{05} : $\alpha_p^* = \delta_{pq}^*$ ou $-\alpha_p^* = -\delta_{pq}^*$, rejeitada	1 : 1 : 1 : 1	idem (a)	
		$+\phi-$ ou $-\phi+$		H_{05} : $\alpha_p^* = -\delta_{pq}^*$ ou $-\alpha_p^* = \delta_{pq}^*$, rejeitada	1 : 1 : 1 : 1	idem (a)	
		$+\phi\pm$; $-\phi\pm$		H_{05} não rejeitada	1 : 2 : 1	idem (a)	
H_{02} e H_{03}		$\phi++$ ou $\phi--$		H_{06} : $\alpha_q^* = \delta_{pq}^*$ ou $-\alpha_q^* = -\delta_{pq}^*$, rejeitada	1 : 1 : 1 : 1	idem (b)	
		$\phi+-$ ou $\phi-+$		H_{06} : $\alpha_q^* = -\delta_{pq}^*$ ou $-\alpha_q^* = \delta_{pq}^*$, rejeitada	1 : 1 : 1 : 1	idem (b)	
		$\phi+\pm$; $\phi-\pm$		H_{06} não rejeitada	1 : 2 : 1	idem (b)	
H_{01} , H_{02} e H_{03}		$++\pm$; $+-\pm$; $+-\pm$; $+-\pm$		uma ou três hipóteses dentre H_{04} , H_{05} e H_{06} rejeitadas	1 : 1 : 1 : 1	idem (c), (d), (e) e (f)	
H_{01} , H_{02} e H_{03}		$++\pm$; $+-\pm$; $+-\pm$; $+-\pm$		duas hipóteses dentre H_{04} , H_{05} e H_{06} rejeitadas	1 : 2 : 1	idem (c), (d), (e) e (f)	
H_{01} , H_{02} e H_{03}		$++\pm$; $+-\pm$; $+-\pm$; $+-\pm$		nenhuma dentre H_{04} , H_{05} e H_{06} rejeitada	3 : 1	idem (c), (d), (e) e (f)	

(1): Os sinais $+$, $-$ e ϕ indicam, respectivamente, QTL com efeito positivo, negativo e nulo (não significativo) sobre o fenótipo. O sinal \pm indica que há efeito significativo, cujo sinal não interfere na Etapa 2.

(2): Para simplificar são apresentadas apenas os alelos do QTL; P^1P^2 significa configuração $P_m^1P_{m+1}^1/P_m^2P_{m+1}^2$, e assim por diante.

Tabela 6 – Matrizes para realizar os testes da Etapa 2

Hipótese	\mathbf{T}
$H_{04}: \alpha_p^* = \alpha_q^* \text{ ou } -\alpha_p^* = -\alpha_q^*$	$\begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{bmatrix}$
$H_{04}: \alpha_p^* = -\alpha_q^* \text{ ou } -\alpha_p^* = \alpha_q^*$	$\begin{bmatrix} 1/2 & 0 \\ -1/2 & 0 \\ 0 & 1 \end{bmatrix}$
$H_{05}: \alpha_p^* = \delta_{pq}^* \text{ ou } -\alpha_p^* = -\delta_{pq}^*$	$\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ 1/2 & 0 \end{bmatrix}$
$H_{05}: \alpha_p^* = -\delta_{pq}^* \text{ ou } -\alpha_p^* = \delta_{pq}^*$	$\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \\ -1/2 & 0 \end{bmatrix}$
$H_{06}: \alpha_q^* = \delta_{pq}^* \text{ ou } -\alpha_q^* = -\delta_{pq}^*$	$\begin{bmatrix} 0 & 1 \\ 1/2 & 0 \\ 1/2 & 0 \end{bmatrix}$
$H_{06}: \alpha_q^* = -\delta_{pq}^* \text{ ou } -\alpha_q^* = \delta_{pq}^*$	$\begin{bmatrix} 0 & 1 \\ 1/2 & 0 \\ -1/2 & 0 \end{bmatrix}$

3.3 Simulação

Para exemplificar e validar o modelo apresentado uma simulação foi realizada, de forma análoga ao apresentado por Zeng (1994), Kao e Zeng (1997) e Lin et al. (2003). Considerou-se uma progênie de irmãos completos com 300 indivíduos e mapa genético com quatro cromossomos, com 15 marcadores moleculares igualmente distanciados a 10 cM, usando a função de Kosambi (KOSAMBI, 1944). Foram considerados marcadores com diferentes tipos de segregação, optou-se por utilizar a notação proposta por Wu et al. (2002a). Resumidamente, os marcadores são classificados em quatro tipos, em função de seu tipo de segregação: *A* (1:1:1:1), *B* (1:2:1, divididos em B_1 , B_2 e B_3 em função da presença do alelo nulo no genitor *P*, *Q* ou em nenhum deles, respectivamente), *C* (3:1) e *D* (1:1, sendo D_1 caso o pai heterozigoto seja *P* e D_2 se *Q* for heterozigoto). Do total simulado, 15 marcadores foram completamente informativos (tipo *A*); 15 marcadores tipo *B* (igualmente distribuídos entre os casos B_1 , B_2 , B_3); 10 marcadores tipo *C* e 20 marcadores do tipo *D*, sendo metade do tipo D_1 e metade do tipo D_2 . A distribuição dos marcadores entre os cromossomos foi feita de forma aleatória, resultando na seguinte distribuição: cromossomo um, 5 *A*, 1 B_1 , 0 B_2 , 2 B_3 , 4 *C*, 2 D_1 , 1 D_2 ; cromossomo dois, 3 *A*, 2 B_1 , 1 B_2 , 1 B_3 , 1 *C*, 2 D_1 , 5 D_2 ; cromossomo três, 4 *A*, 0 B_1 , 1 B_2 , 2 B_3 , 2 *C*, 4 D_1 , 2 D_2 ; cromossomo quatro, 3 *A*, 2 B_1 , 3 B_2 , 0 B_3 , 3 *C*, 2 D_1 , 2 D_2 . A ordem dos marcadores é apresentada nas Figuras 3 e 4 (eixo x).

O caráter simulado é controlado por oito QTLs posicionados ao longo dos quatro cromossomos, cujos efeitos genéticos foram distribuídos de forma a representar várias fases de ligação e padrões de segregações (Tabela 7). A herdabilidade do caráter foi de 0,70.

Para realizar o mapeamento de QTLs, foi necessário obter as probabilidades condicionais multiponto a priori (p_{klj}) para os genótipos P^1Q^1 , P^1Q^2 , P^2Q^1 e P^2Q^2 , através da abordagem multiponto desenvolvida por Mollinari; Margarido e Garcia¹ (em fase de elaboração) e incorporadas no software OneMap (MARGARIDO; GARCIA; SOUZA, 2007). Tais probabilidades foram calculadas ao longo dos cromossomos a cada 1 cM. Foi então realizado o mapeamento por intervalo, isto é, considerando o modelo em (8) sem a inclusão de cofatores. Em seguida, foi empregado também o mapeamento por intervalo composto. A realização do mapeamento sem os cofatores foi feita para verificar se o novo modelo CIM possui as

¹MOLLINARI, M.; MARGARIDO, G.R.A.; GARCIA, A.A.F. **A Multipoint Approach Based on Hidden Markov Chains for Construction of Linkage Maps in Outcrossing Species.**

Tabela 7 – Efeitos genéticos, segregação e fase de ligação dos QTLs simulados

QTL	Posição		Efeitos genéticos				Seg.	Fase
	Crom.	$cM^{(1)}$	μ	α_p^*	α_q^*	δ_{pq}^*		
I	1	15	0	1,70	0,00	0,00	1:1	P^1P^2
II	1	115	0	0,00	-1,70	0,00	1:1	Q^2Q^1
III	2	45	0	2,00	1,10	-2,00	1:2:1	$P^1P^2 \times Q^1Q^2$
IV	2	75	0	-1,60	-1,60	1,60	3:1	$P^2P^1 \times Q^2Q^1$
V	2	105	0	1,80	0,00	-1,80	1:2:1	P^1P^2
VI	3	25	0	1,10	-1,50	-0,60	1:1:1:1	$P^1P^2 \times Q^2Q^1$
VII	3	65	0	1,60	0,00	0,00	1:1	P^1P^2
VIII	4	55	0	1,50	1,50	0,00	1:2:1	$P^1P^2 \times Q^1Q^2$

⁽¹⁾: em relação ao primeiro marcador do cromossomo.

propriedades apresentadas por Zeng (1994). A seleção de cofatores foi realizada através de regressão linear múltipla via stepwise, utilizando o critério de informação bayesiano (BIC). Cada marcador incluído implicou na inclusão de três parâmetros (α_p , α_q e δ_{pq}), cuja significância foi posteriormente verificada, sendo os efeitos não significativos (5%) removidos do modelo. Ainda, de acordo com proposto por Zeng (1994) dentre os cofatores selecionados, considerou-se para a inclusão no modelo, apenas os marcadores que estivessem a uma distância maior que 10 cM (*window size*) dos marcadores flanqueadores do intervalo sendo mapeado.

Para a busca de QTLs ao longo do genoma foi usado o teste da razão de verossimilhança, com três graus de liberdade. Os demais testes realizados na Etapa 1 (H_{01} , H_{02} e H_{03}) e na Etapa 2 (H_{04} , H_{05} e H_{06}) foram feitos considerando um grau de liberdade; tais testes são feitos apenas nas posições com possíveis QTLs, não implicando nos problemas decorrentes da realização de múltiplos testes (JIANG; ZENG, 1995).

O limiar para considerar a presença de QTLs na etapa de busca foi obtido através da abordagem apresentada por Vieira et al. (2000). Neste caso, ajustou-se o número de testes independentes realizado ao longo do genoma e utilizou-se a correção de Boferroni para adequação do nível de significância em múltiplos testes. De maneira geral, no CIM o comprimento de região delimitada pelo intervalo do *window size* é considerado independente das demais regiões do cromossomo. De forma conservadora, essa região tinha 30 cM, no presente

caso (10 cM para o intervalo e 2×10 com a inclusão do window size). Portanto, o número estimado de testes independentes foi 23, obtidos a partir de $4 \times \left(\frac{T_w}{30} + 1\right)$, em que T_w é o comprimento total w -ésimo cromossomo (140 cM). Considerando-se um nível global de significância de 5%, três graus de liberdade e 23 testes múltiplos tem-se um nível de significância individual de $2,17 \times 10^{-4}$, correspondente a $LRT=14,61$ (LOD Score 3,2).

3.4 Resultados

Mapeamento por intervalo (IM):

Como esperado, o mapeamento por intervalo apresentou comportamento apenas razoável para a detecção de QTLs (Figuras 3 e 4), pois permitiu a correta localização dos QTLs no cromossomo 1, em que os QTLs se comportam de maneira independente, uma vez que estão distanciados a 100 cM. No cromossomo 2 foram detectados apenas dois QTLs dos três simulados. Para o cromossomo 3, há indícios para a presença de QTLs, pois uma grande região com cerca de 80 cM apresenta LOD acima do limiar. Entretanto, o resultado do mapeamento não permite concluir com precisão se há dois QTLs posicionados a 25 cM e 65 cM, conforme simulado, ou se foi detectado um QTL com efeito residual nos intervalos adjacentes. No cromossomo 4, houve a detecção do QTL posicionado 60 cM, mas também houve a sugestão de um falso QTL a 12 cM.

Seleção de cofatores:

Foram selecionados oito cofatores, sendo todos eles flanqueadores das regiões com os QTLs simulados. Não houve problemas com superparametrização do modelo CIM, devido ao tamanho amostral que permitiria um total $2\sqrt{300} = 34$ parâmetros, e foram incluídos 15 efeitos genéticos no modelo com oito cofatores (Tabela 8). Nota-se que nesta etapa 9 efeitos foram removidos, de forma que permaneceram no modelo apenas os efeitos dos cofatores de alguma forma associados aos QTLs. Vale ressaltar, que apesar de alguns cofatores selecionados apresentarem apenas informação em um dos genitores (D_1 ou D_2), os efeitos de dominância foram conservados em alguns casos. Isto ocorre porque a abordagem multiponto para obtenção das probabilidades permite a recuperação da informação dos genótipos mesmo com marcadores pouco informativos.

Mapeamento por intervalo composto (CIM):

Com a realização do mapeamento por intervalo composto (Figuras 3 e 4) é possível ob-

Tabela 8 – Localização dos cofatores em relação a posição dos QTLs simulados

Crom.	QTL		Cofatores			
	cM	Seg.	cM	Seg.	Efeitos Significativos	
1	15	1 : 1	10	A	α_p	
1	115	1 : 1	120	D_1	α_q	
2	45	1 : 2 : 1	40	D_1	α_p e δ_{pq}	
2	75	3 : 1	80	C	α_p, α_q e δ_{pq}	
2	105	1 : 2 : 1	110	D_2	α_p e δ_{pq}	
3	25	1 : 1 : 1 : 1	20	B_3	α_p e α_q	
3	65	1 : 1	60	D_1	α_p e δ_{pq}	
4	55	1 : 2 : 1	60	B_2	α_p e α_q	

servar melhora significativa dos resultados em relação ao IM, pois todos os QTLs simulados foram mapeados e houve, inclusive, a eliminação do provável falso positivo detectado no mapeamento por intervalo. Destaca-se também que praticamente todos os QTLs mapeados no modelo CIM apresentam LOD maior, o que evidencia maior poder estatístico. Isso comprova as vantagens do CIM em relação ao IM preconizados por Zeng (1994).

Ao comparar o modelo CIM com o IM para o cromossomo 1, é possível verificar que novamente os dois QTLs foram detectados, porém com melhor resultados, uma vez que as posições sugeridas foram mais próximas aos locais que os QTLs se encontravam, com maior LOD Score. Com isto, detectaram-se os QTLs nas posições 15 e 111 cM. A primeira posição refere-se a localização exata do QTL simulado e a segunda esta contida no mesmo intervalo de marcadores que o QTL simulado.

O modelo CIM apresentou resultados interessantes para o cromossomo 2, pois diferentemente da abordagem IM houve a detecção dos três QTLs simulados, além da melhor definição do QTL na posição 44 cM. Nota-se que a definição da posição dos QTLs é mais evidente.

Para o cromossomo 3 também houve uma melhoria substancial nos resultados, pois com o modelo CIM houve a clara definição de dois picos distintos na região em que o mapeamento por intervalo indicava uma grande região de 80 cM, sem precisar a localização dos QTLs. Novamente, as posições mapeadas estão contidas nos mesmos intervalos em que os QTLs foram simulados.

No cromossomo 4 o modelo CIM removeu o QTL falso positivo próximo a posição 10

cM e também posicionou um QTL a 61 cM. Contudo, o QTL simulado estava a 55 cM, fora do intervalo simulado, embora adjacente. A análise gráfica permite inferir que o intervalo de confiança para este QTL conteria a posição simulada, o que em situações reais não acarretaria consequências que comprometessem as aplicações práticas destes resultados.

Fase de ligação e segregação:

De forma geral, os resultados sobre a inferência dos padrões de segregação dos QTLs foram satisfatórios para todos os casos (Tabela 7 e 9), uma vez que os efeitos estimados estavam relativamente próximos dos valores reais. Nota-se algum viés nas estimativas da média, que mostrou tendência de estimativas negativas, o que precisa ser melhor investigado. No caso dos testes marginais (Etapa 1), $LOD = 1,2$ pode ser usado para inferir sobre a significância dos efeitos.

Os QTL que segregavam na proporção 1:1 (i, ii e vii) (Tabela 9) foram corretamente caracterizados, sendo os efeitos estimados muito próximos aos simulados. O QTL VI, com segregação 1:1:1:1 e com efeitos aditivos maiores que o efeito de dominância, foi mapeado (QTL vi) como tendo segregação 1:2:1 e com efeitos $\alpha_p^* = -\alpha_q^*$, mas δ_{pq}^* não significativo. Neste caso, a segregação inferida foi diferente da segregação simulada, provavelmente pela pequena magnitude dos efeitos, dificultando a correta identificação.

Os QTLs que segregavam na proporção 1:2:1 (iii, v, viii) tiveram a segregação corretamente inferida. Para os três casos houve sempre dois efeitos significativos e estatisticamente idênticos. Por isto, apesar do QTL iii ter tido a segregação corretamente inferida, indicando que $\alpha_p^* = -\delta_{pq}^*$, o efeito α_q^* foi considerado erroneamente como não significativo.

Para a situação em que o QTL simulado apresentava segregação 3:1 (IV) houve a identificação de segregação como sendo 1:2:1 (iv), pois a estimativa para α_p^* nessa posição não foi significativa. Uma provável explicação seria sua localização entre dois outros QTLs com efeitos maiores, dificultando sua identificação.

Com relação as estimativas de fase de ligação dos QTLs, o modelo CIM teve comportamento muito satisfatório, pois em todos os casos em que os efeitos estimados foram significativos, as fases de ligação foram corretamente estimadas. As diferenças que ocorreram para os QTLs iii e iv devem ao fato de alguns efeitos não foram significativos.

Tabela 9 – QTTLs mapeados usando modelo CIM

		Etapa 1										Etapa 2			Conclusão	
		Posição					Efeitos					p-valores				
QTTL	Crom.	cM	LOD	$\hat{\mu}$	$\hat{\alpha}_p^*$	LOD	$\hat{\alpha}_q^*$	LOD	$\hat{\delta}_{pq}^*$	LOD	$\hat{\sigma}^{2*}$	H_{04}	H_{05}	H_{06}	Seg.	Fase
i	1	15	6,94	-0,20	1,53	6,40	-0,22	0,17	-0,40	0,40	20,57	-	-	-	1:1	P^1P^2
ii	1	111	6,49	-0,24	-0,03	0,00	-1,50	5,88	0,28	0,74	20,70	-	-	-	1:1	Q^2Q^1
iii	2	44	11,99	-0,21	1,26	5,84	0,45	0,71	-1,63	5,35	17,13	-	1	-	1:2:1	P^1P^2
iv	2	82	7,13	-0,23	-0,47	0,95	-1,11	2,95	0,98	2,45	20,22	-	-	1	1:2:1	Q^2Q^1
v	2	110	8,38	-0,24	1,26	4,78	0,16	0,04	-1,27	3,72	19,57	-	1	-	1:2:1	P^1P^2
vi	3	20	5,04	-0,19	0,73	1,78	-1,06	3,15	-0,48	0,76	20,39	0,61	-	-	1:2:1	$P^1P^2 \times Q^2Q^1$
vii	3	68	8,39	-0,17	1,60	7,94	0,13	0,03	0,40	0,38	20,72	-	-	-	1:1	P^1P^2
viii	4	61	10,69	-0,19	1,26	4,53	1,49	6,68	0,31	0,64	20,91	0,49	-	-	1:2:1	$P^1P^2 \times Q^1Q^2$

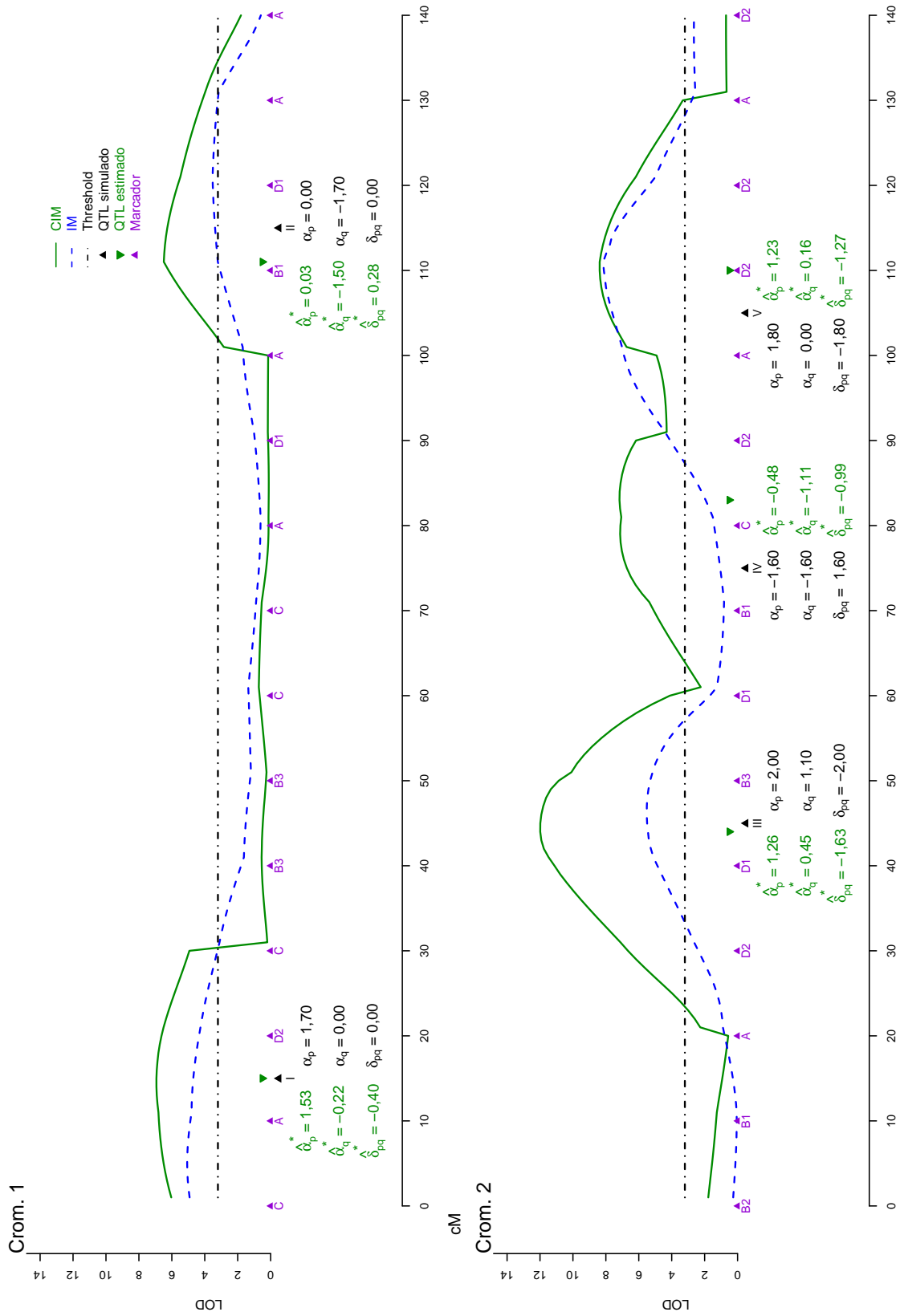


Figura 3 – Mapeamento de QTLs para população simulada, cromossomos 1 e 2. Os efeitos genéticos são indicados por α_p^* (aditivo, genitor P), α_q^* (aditivo, genitor Q), δ_{pq}^* (dominância)

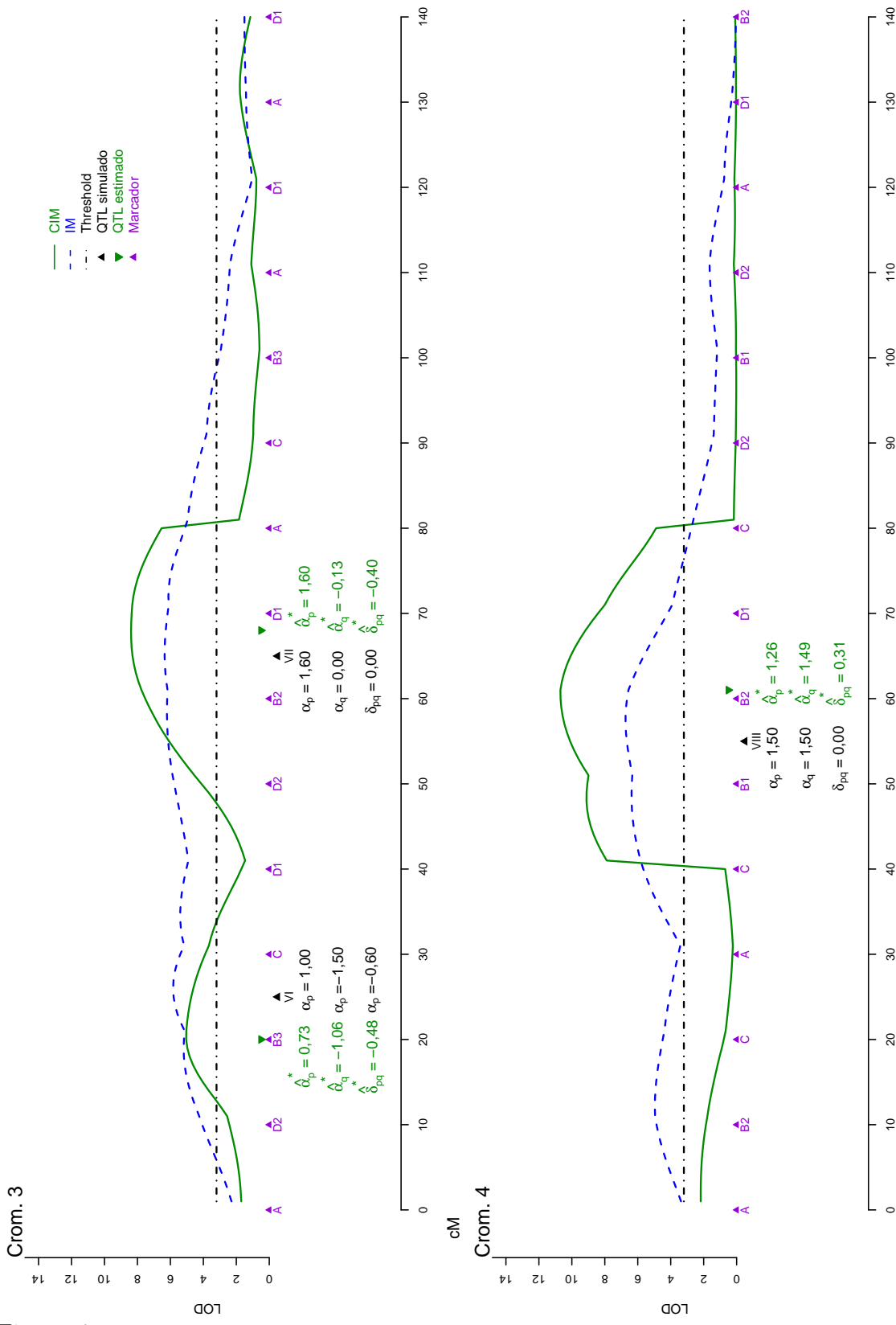


Figura 4 – Mapeamento de QTLs para população simulada, cromossomos três e quatro. Os efeitos genéticos são indicados por α_p^* (aditivo, genitor P), α_q^* (aditivo, genitor Q), δ_{pq}^* (dominância)

3.5 Discussão

No presente trabalho foi apresentado um modelo para mapear QTLs em espécies diploides de polinização aberta, usando uma progênie de irmãos completos. O modelo considera os diferentes tipos de segregação que os marcadores moleculares e os QTLs podem assumir neste contexto. A abordagem utilizada foi baseada no mapeamento por intervalo composto (ZENG, 1993, 1994), que foi desenvolvido para populações de mapeamento derivadas de linhagens endogâmicas. Para a validação do modelo foi feita uma simulação que considerava um caráter quantitativo com herdabilidade 0,70 e controlado por oito QTLs, com diferentes efeitos, segregações, fases de ligação.

No geral, o uso do modelo permitiu mapear os QTLs simulados, com sua correta caracterização. Sempre que os efeitos dos QTLs foram significativos, o modelo forneceu estimativas corretas das fase de ligação, isto é, foi possível identificar a origem dos alelos que aumentam ou reduzem o fenótipo. A principal vantagem disso é que os resultados do mapeamento podem ser úteis em programas de seleção assistida, mesmo que ocasionalmente ocorram falhas na inferência da segregação e/ou das estimativas dos efeitos dos QTLs. Etapas seguintes, como por exemplo, clonagem de QTLs de interesse, também podem se beneficiar de resultados como os apresentado.

O modelo apresenta vantagens sobre a proposta de Lin et al. (2003). Diferentemente do que esses autores apresentam, as fases de ligação não foram consideradas como parâmetros a serem estimados no modelo, mas sendo obtidas através da interpretação dos sinais das estimativas α_p^* e α_q^* . Isto reduz sensivelmente a complexidade do algoritmo *EM*, o que permite que o modelo apresentado pudesse ser facilmente expandido para o contexto do CIM. Situações mais complexas, como as encontradas no mapeamento por múltiplos intervalos (KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999) e mapeamento para múltiplas características e ambientes (JIANG; ZENG, 1995), também poderão ser feitas usando o modelo como ponto de partida em estudos futuros podem incluir investigações sobre interações epistáticas, interações entre QTLs-ambientas e correlação entre caracteres. A expansão destes modelo para abordagem de modelos mistos (BOER et al. 2007; EEUWIJK et al., 2005; EEUWIJK; MALOSETTI; BOER, 2007; MALOSETTI et al., 2004) também é possível, uma vez que o modelo de mapeamento está desenvolvido a partir de contrastes ortogonais entre genótipos

dos QTLs.

Lin et al. (2003) comentam que segregação do tipo 1:1 é testada supondo que um dos efeitos marginais é igual a zero; o padrão 1:2:1, análogo o que ocorre em F_2 , ocorre quando os efeitos marginais são estatisticamente iguais. Contudo, o modelo apresentado permite identificar as possíveis segregações incluindo a segregação 3:1, através de um procedimento para a identificação destas situações e contornando os problemas dos múltiplos testes. Sua validade foi verificada para o caso simulado. O fato do modelo basear-se em probabilidades condicionais multiponto é uma grande vantagem, já que a presença de marcadores informativos ao longo dos grupos de ligação permite a detecção de QTLs de diferentes padrões de segregação (1:1:1:1, 1:2:1, 3:1), mesmo em regiões com marcadores menos informativos. Por exemplo, Lin et al. (2003) não apresentaram maneiras eficientes de estimar as probabilidades condicionais entre marcadores menos informativos, como 1:1 e 3:1. No presente caso, a informação dos marcadores adjacentes é recuperada e as probabilidades são melhor estimadas. Jiang e Zeng (1997) comprovam isso para populações baseadas em linhagens endogâmicas, mas no nosso conhecimento esse é o primeiro trabalho a usar probabilidades condicionais multiponto para os genótipos dos QTLs em progênies de irmãos completos.

A utilização da estratégia para identificar a segregação e fases dos QTLs dependem de sua correta localização e assim modelos que permitam um maior controle da variância residual são vantajosos. Zeng (1994) aponta que a utilização de regressão linear múltipla combinado com o mapeamento por intervalo (LANDER; BOTSTEIN, 1989) fornecem estimativas para efeitos dos QTLs mais fidedignos se comparados a utilização destas abordagens. No presente trabalho, observou que o modelo CIM, posicionou os QTLs de forma mais precisa que a abordagem IM, além de apresentar maior poder de mapeamento. O modelo de Lin et al. (2003) não permite a inclusão de cofatores de maneira simples como aqui apresentada.

Pelas razões expostas, espera-se que o trabalho seja útil para mapeamento de QTLs em espécies que não dispõem de linhagens endogâmicas e que portanto não podem realizar o mapeamento com maior precisão. Acredita-se que o uso de mapas genéticos integrados, associados ao mapeamento de QTLs com maior precisão possa permitir um melhor entendimento da base genética dos caracteres quantitativos nessas espécies.

3.6 Conclusão

O modelo CIM permitiu o mapeamento de QTLs em progênies de irmãos completos, com maior precisão que outras abordagens presentes na literatura. O modelo permitiu a identificação da posição dos QTLs, bem como seus efeitos e fases de ligação. Tal modelo pode ser estendido para o contexto multivariado, o que seria de grande valia para estudos de interação entre QTLs e ambientes e estudos de correlação entre caracteres.

Referências

- ABATE, Z.A.; LIU, S.; McKENDRY, A.L. Quantitative Trait Loci Associated with Deoxynivalenol Content and Kernel Quality in the Soft Red Winter Wheat 'Ernie', **Crop Science**, Madison, v. 48, p. 1408-1418, 2008.
- 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, 2008.
- CARDINAL, A.J.; LEE, M.; SHAROPOVA, N. WOODMAN-CLIKEMAN, W.L.; LONG, M.L. Genetic Mapping and Analysis of Quantitative Trait Loci for Resistance to Stalk Tunneling by the European Corn Borer in Maize, **Crop Science**, Madison, v. 41, p. 835-845, 2001.
- CHO, Y.G.; KANG, H.J.; LEE, J.S.; LEE, Y.T.; LIM, S.J.; GAUCH, H.; EUN, M.Y.; McCouch, S.R. Identification of Quantitative Trait Loci in Rice for Yield, Yield Components, and Agronomic Traits across Years and Locations, **Crop Science**, Madison, v. 47 p. 2403–2417, 2007.
- CHURCHILL, G.A.; DOERGE, R.W. Empirical Threshold Values for Quantitative Trait Loci, **Genetics**, Bethesda, 138, 963-971, 1994.
- 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; RUBIN, D.B. Maximum Likelihood from Incomplete Data via EM algorithm, **Journal of the Royal Statistical Society. Series B. Methodological**, London, v. 39, 1-38, 1977.

EEUWIJK, F.A. van; MALOSETTI, M. BOER; M.P. Modelling the genetic basis of response curves underlying genotype x environment interaction. In: SPIERTZ, J.H.J.; STRUIK, P.C., LAAR, H.H. van. **Scale and Complexity in Plant Systems Research. Gene-Plant-Crop Relations**. Dordrecht: Springer, 2007. Chap. 10, p. 115-126.

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

GARCIA, A.A.F.; WANG, S.; MELCHINGER, A.E.; ZENG, Z.B. Quantitative Trait Loci Mapping and The Genetic Basis of Heterosis in Maize and Rice, **Genetics**, Bethesda, v. 180, p. 1707-1724, 2008.

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.

JANSEN, R.C.; STAM, P. 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.

JIANG, C.; ZENG, Z. B. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. **Genetica**, Gravenhage, v. 101, p. 47-58, 1997.

JOHNSON, D.L.; JANSEN, R.C.; ARENDONK, J.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.

KAO, C.H.; ZENG, Z.B. General formulae 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.

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

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

LI, D.; PFEIFFER, T.W.; CORNELIUS, P.L. Soybean QTL for Yield and Yield Components Associated with Glycine soja Alleles, **Crop Science**, Madison, v. 48, p. 571–581, 2008.

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.

LU, Q.; CUI, Y; WU, R. A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. **BMC Genetics**, London, v. 5, p. 5-20, 2004.

MACCAFERRI, M.; SANGUINETI, M.C.; CORNETI, S.; ORTEGA, J.L.A.; SALEM, M.B.; BORT, J.; DeAMBROGIO, E.; MORAL, L.F.G. del; DEMONTIS, A.; EL-AHMED, A.; MAALOUF, F.; MACHLAB, H.; MARTOS, V.; MORAGUES, M.; MOTAWAJ, J.; NACHIT, M.; NSERALLAH, N.; OUABBOU, H.; ROYO, C.; SLAMA, A.; TUBEROSA, R. Quantitative Trait Loci for Grain Yield and Adaptation of Durum Wheat (*Triticum durum* Desf.) Across a Wide Range of Water Availability, **Genetics**, Bethesda, v. 178, p. 489–511, 2008.

MACKAY, T.F.C. The genetic Architecture of Quantitative Traits, **Annual Reviews of Genetics**, Palo Alto, v. 35, p. 303–39, 2001.

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

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.

MANGOLIN, C.A.; SOUZA JÚNIOR, C.L. de; GARCIA, A.A.F.; GARCIA, A.F.; SIBOV, S.T.; SOUZA, A.P. de. Mapping QTLs for kernel oil content in a tropical maize population, **Euphytica**, Wageningen, v. 137, p. 251–259, 2004.

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.

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.

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 RFLP Linkage Maps in Plants From Crosses Between Heterozygous Parents, **Genetics**, Bethesda, v. 125, p. 645-654, 1990.

SABADIN, P.K.; SOUZA JR., C.L.; SOUZA, A.P.; GARCIA, A.A.F. QTL mapping for yield components in a tropical maize population using microsatellite markers, **Hereditas**, Lund, v. 145, p. 194-203, 2008.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA. **Akaike information criterion statistics**. Tokyo: KTK Scientific Publishers, 1986, 290 p.

SCHWARZ, G. Estimating The Dimension of a model, **Annals of statistics**, Hayward, v. 6, p. 461-464, 1978.

SEMAGN, K.; SKINNES, H.; BJØRNSTAD, A.; MARØY, A.G.; TARKEGNE, Y. Quantitative Trait Loci Controlling Fusarium Head Blight Resistance and Low Deoxynivalenol Content in Hexaploid Wheat Population from ‘Arina’ and NK93604. **Crop Science**, Madison, v. 47, p. 294–303, 2007.

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, 2003.

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, 1998.

VIEIRA, C.; PASYUKOVA, E.G.; ZENG, Z.B.; HACKETT, J.B.; LYMAN, R.F.; MACKAY, T.F. Genotype-Environment Interaction for Quantitative Trait Loci Affecting Life Span in *Drosophila melanogaster*, **Genetics**, Bethesda, v. 154, p. 213–227, 2000.

WANG, S., Basten, C.J.; ZENG, Z.B. **Windows QTL Cartographer**: Version 2.5. Department of Statistics, North Carolina State University, North Carolina: Raleigh, 2007.

WASSOM, J.J.; MIKKELINENI, V.; BOHN, M.O.; ROCHEFORD, T.R. QTL for Fatty Acid Composition of Maize Kernel Oil in Illinois High Oil × B73 Backcross-Derived Lines, **Crop Science**, Madison, v. 48, p. 69-78, 2008a

WASSOM, J.J.; JeFFREY, C.W.; MARTINEZ, E.; KING, J.J.; DeBAENE, J. QTL Associated with Maize Kernel Oil, Protein, and Starch Concentrations; Kernel Mass; and Grain Yield in Illinois High Oil × B73 Backcross-Derived Lines, **Crop Science**, Madison, v. 48, p. 243–252, 2008b.

WU, R.L.; 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, 2002a.

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, 2002b.

ZENG, Z.B. Theoretical basis of precision mapping of quantitative trait loci, **Proceedings of the National Academy of Sciences 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., KAO, C.H.; BASTEN, C.J. Estimating the genetic architecture of quantitative traits. **Genetical Research**, London, v. 74, p. 279-289, 1999.

ZHAO, M.; ZHANG, Z.; ZHANG, S.; LI, W.; JEFFERS, D.P.; RONG, T.; PAN, G. Quantitative Trait Loci for Resistance to Banded Leaf and Sheath Blight in Maize, **Crop Science**, Madison, v. 46, p. 1039–1045, 2006.

4 MAPEAMENTO DE QTLs EM CANA-DE-AÇÚCAR UTILIZANDO MAPEAMENTO POR INTERVALO COMPOSTO E MAPA GENÉTICO INTEGRADO

Resumo

O mapeamento de QTLs em cana-de-açúcar é dificultado pela não disponibilidade de linhagens endogâmicas e pelo alto nível de ploidia dos cultivares modernos. Nesse caso, a população segregante para mapeamento é obtida a partir do cruzamento entre dois indivíduos não endogâmicos. Assim, os locos podem apresentar diversos padrões de segregação. A construção de mapas de ligação e mapeamento de QTLs são realizados com a utilização de marcadores que apresentam padrões de segregação 1:1 e 3:1, sendo muito utilizada a abordagem de duplo pseudo testcross, que permite a utilização de modelos elaborados para populações experimentais. Tal abordagem é baseada na construção de dois mapas genéticos, um para cada genitor. Diversos autores apresentam metodologias que permitem a construção de mapas integrados, porém metodologias para mapeamento de QTLs nesse contexto ainda são escassas. Gazaffi² desenvolveu um procedimento baseado no mapeamento por intervalo composto (CIM), utilizando marcadores com diferentes padrões de segregação, destacando a possibilidade de caracterizar as fases de ligação e a segregação dos QTLs. O objetivo do presente trabalho foi aplicar tal metodologia para mapeamento de QTLs de caracteres relacionados à produção de cana-de-açúcar, tais como toneladas de açúcar por hectare (TPH), tonelada de cana por hectare (TCH), porcentagem de açúcar na cana (PCC) e teor de Fibra, considerando dois cortes. O mapeamento de QTLs com modelo CIM detectou 41 QTLs, sendo 18 para primeiro corte e 23 para o segundo corte. Foram obtidos 14 QTLs para TPH (8 primeiro corte e 6 no segundo corte), 8 QTLs para TCH (4 em cada corte), 12 para PCC (5 no primeiro e 7 no segundo corte) e 7 QTLs para teor de Fibra (1 no primeiro corte e seis no segundo). Dentre os QTLs mapeados 33 QTLs apresentavam segregação 1:1, 2 apresentaram segregação 3:1, 5 segregavam 1:2:1 e um QTL tinha segregação 1:1:1:1. Os QTLs tiveram R^2 variando de 0,5% (TCH, primeiro corte) a 35% (TCH, segundo corte), e tiveram predomínio de ação gênica aditiva. No geral, a metodologia foi superior as metodologias comumente aplicadas para mapeamento de QTLs neste contexto, pois apresentou maior poder estatístico, além de identificar outros padrões de segregação que não seria possíveis de serem obtidos em outros abordagens.

Palavras-chave: QTL; Caracteres quantitativos; Poliploide; Análise multiponto; EST-SSR; EST-RFLP

QTL MAPPING IN SUGARCANE USING COMPOSITE INTERVAL MAPPING AND INTEGRATED GENETIC MAP

Abstract

QTL mapping in sugarcane is normally done using segregating population developed from crosses between two non-inbred individuals. Thus, QTL and marker loci may show different patterns of segregation. QTL mapping is generally made using markers that segregate in a 1:1 and 3:1 fashion, using the strategy known as *double pseudo testcross*, which allows the use of statistical models designed for inbred-based populations, for each of the two genetic maps (one for each parent). Several authors developed methodologies that allow the construction of integrated genetic maps, but methodologies for QTL mapping in this context are still scarce. Gazaffi² developed a procedure based on composite interval mapping (CIM), using markers with different patterns of segregation and allowing the possibility of estimate linkages phases between QTL and markers, as well as their patterns of segregation.

The objective of this study was to apply this methodology for QTL mapping for traits related to yield in sugarcane, such as tonnes of sugar per hectare (TPH), ton of cane per hectare (TCH), percentage of sugar cane (PCC) and fiber content, considering two harvests. QTL mapping with CIM model detected 41 QTL, 18 for the plant crop to 23 for ratoon crop; 14 QTL were obtained for TPH (8 for plant crop and 6 in ratoon crop), 8 QTL for TCH (4 in each ratoon), 12 for PCC (5 in plant crop and 7 in the ratoon crop) and 7 QTL for fiber content (1 the field plant and six in the ratoon). 33 QTL had segregation 1:1, 2 showed 3:1 segregation, 5 had segregation 1:2:1 and a QTL had 1:1:1:1 segregation. Mapped QTL explained from 0.5% (TCH, plant crop) to 35% (TCH, ratoon crop) of the phenotic variation and had a predominance of additive gene action. In general, the CIM model provided better results than others approaches, greater statistical power and identifying others patterns not considered in other models.

Keywords: QTL; Quantitative trait; Polyploid; Multipoint analysis; EST-SSR; EST-RFLP

4.1 Introdução

A cana-de-açúcar (*Saccharum* spp) apresenta grande importância econômica, pois é cultivada em diversos países de clima tropical e subtropical para a produção de açúcar e álcool. O atual interesse econômico-ambiental despertado por seu principal produto, o etanol, faz crescer o potencial de expansão desta cultura nos próximos anos. Para atender esta demanda é essencial a obtenção de novos cultivares mais aptos e resistentes às diversas condições de cultivo. No entanto, o melhoramento genético da cana-de-açúcar é oneroso, baseando-se em cruzamentos interespecíficos entre indivíduos do gênero *Saccharum*, o que resultou na produção de cultivares com elevado nível de ploidia e com diferenças no número de cromossomos (HEINZ; TEW, 1987; IRVINE, 1999).

A obtenção de novos cultivares de cana-de-açúcar é realizada através de diversas etapas de seleção e propagação vegetativa dos melhores indivíduos, os quais são comumente originários de uma progênie segregante formada pelo cruzamento entre genitores com características de interesse. Esta seleção é complexa, pois os principais caracteres de interesse econômico têm herança quantitativa e podem ser influenciados pelo ambiente. O uso de marcadores moleculares poderia tornar o processo de seleção mais rápido e eficiente, via *seleção assistida por marcadores* (MAS) (CHARCOSSET; MOREAU, 2004; MOHAN et al., 1997; MORGANTE; SALAMINI, 2003; TAKEDA; MATSUOKA, 2008). No entanto, para que MAS possa ser implementada de forma eficiente, estudos preliminares referentes a construção de mapas de ligação e mapeamento de QTLs (quantitative trait loci) são necessários para o correto entendimento da arquitetura genética dos caracteres de interesse comercial. O mapeamento de QTLs consiste no procedimento de se estimar o número de regiões que controlam a variação dos caracteres quantitativos, assim como suas respectivas posições, efeitos e interações (ZENG; KAO; BASTEN, 1999).

Vários métodos estatísticos para construção de mapas de ligação e mapeamento de QTLs em espécies diploides foram desenvolvidos para populações experimentais derivadas de cruzamentos entre linhagens endogâmicas (tais como F_2 , retrocruzamentos e linhagens endogâmicas recombinantes), o que permite o controle do número de alelos na população, conhecimento das fases de ligação e maximização do desequilíbrio de ligação (DOERGE, 2002).

Para o mapeamento de QTLs nessas populações há diversas abordagens disponíveis, desde as mais simples que envolvem análises de regressão e variância (DOERGE; ZENG; WEIR, 1997; LIU, 1998), até modelos mais complexos baseados em modelos de misturas e verossimilhança, considerando uma ou várias regiões do genoma simultaneamente (JANSEN; STAM, 1994; KAO; ZENG, 1997; KAO; ZENG; TEASDALE, 1999; LANDER; BOTSTEIN, 1989; ZENG, 1993, 1994).

No caso de espécies que não possuem linhagens endogâmicas, como a cana-de-açúcar, uma abordagem para construção de mapas baseia-se na estratégia de pseudo-testcross (GRATTA-PAGLIA; SEDEROFF, 1994), com marcadores que segregam 1:1 (AITKEN; JACKSON; McINTYRE, 2005; AL-JANABI et al. 2007; HOARAU et al., 2001; MING et al. 2002a; RABOIN et al. 2006). Há também propostas para construção mapas integrados (BUTCHER et al., 2002; ; LU; CUI; WU, 2004; WU et al., 2002a; WU et al., 2002b), o que tem permitido a construção de mapas de cana-de-açúcar considerando marcadores 3:1 e 1:1 (GARCIA et al., 2006; OLIVEIRA et al., 2007).

Apesar desses avanços para a construção de mapas genéticos em cana-de-açúcar, as muitas das estratégias usadas no mapeamento de QTLs ainda baseiam-se em análises de mapas separados para cada genitor e análise de marcas individuais (AITKEN; JACKSON; McINTYRE, 2006; AITKEN et al. 2008; AL-JANABI et al. 2007; HOARAU et al., 2002; McINTYRE et al., 2005a; MING et al., 2002b; RABOIN et al., 2006). A análise de marcas individuais pode ser utilizada para cana-de-açúcar, pois permite usar todos os marcadores disponíveis, independente da segregação e de sua alocação no mapa genético. No entanto, possui baixo poder estatístico, pois há confundimento entre os efeitos do QTL e a sua posição em relação ao marcador, o que pode limitar sua aplicação.

O mapeamento conduzido através da abordagem *duplo pseudo testcross* permite o emprego de metodologias estatísticas desenvolvidas para populações baseadas em linhagens endogâmicas, com maior poder estatístico, como o mapeamento por intervalo (IM) e o mapeamento por intervalo composto (CIM). Para mapas integrados, essa abordagem não pode ser usada diretamente, não usando toda informação disponível. Para tanto, Gazaffi² (em fase de elaboração) desenvolveu um modelo genético-estatístico baseado no mapeamento por

²GAZAFFI, R. (Escola Superior de Agricultura Luiz de Queiroz) **Desenvolvimento de modelo genético-estatístico para o mapeamento de QTLs em progênie de irmãos completos.**

intervalo composto, em que é possível detectar QTLs a partir de um mapa genético integrado formado por marcadores com diversos tipos de segregação. Dentre suas principais vantagens, destacam-se a possibilidade de detectar os efeitos dos QTLs, sua fase de ligação com os marcadores e seu padrão de segregação, explorando as vantagens obtidas com a construção de mapas de ligação integrados baseados em abordagem multiponto e a inclusão de cofatores no modelo.

O objetivo do presente trabalho foi realizar o mapeamento de QTLs em uma progênie de irmãos completos de cana-de-açúcar obtida a partir do cruzamento biparental de cultivares pré comerciais. As metodologias utilizadas foram mapeamento por intervalo composto considerando um mapa genético integrado multiponto, obtido para marcadores com diferentes padrões de segregação. Os resultados foram comparados com a análise de marcas individuais, que é a abordagem mais utilizada para esta cultura nos estudos de mapeamento.

4.2 Materiais e Métodos

4.2.1 População de mapeamento e medidas fenotípicas

Para a obtenção da população de mapeamento foram amostrados 100 indivíduos originários do cruzamento entre os cultivares pré-comerciais SP80-180 (genitor feminino) e SP80-4966 (genitor masculino), realizado na Estação Experimental do CTC (antiga Copersucar) em Camamu, BA. Estes materiais são contrastantes e complementares, sendo que SP80-180 apresenta alta produção de colmos e baixo teor de sacarose e SP80-4966 apresenta baixa produção de colmos e alto teor de sacarose. Tal cruzamento tem sido muito estudado para finalidades de mapeamento genético (GARCIA et al., 2006; OLIVEIRA et al., 2007; PINTO et al., 2004; PINTO et al., 2006). O delineamento estatístico para avaliação dos genótipos foi blocos casualizados com duas repetições, em parcelas formadas por cinco linhas de quatro metros de comprimento. O experimento foi instalado no campo experimental do CTC em Piracicaba, SP, em agosto de 2002.

Os caracteres avaliados foram produtividade (TCH - toneladas de cana por hectare), porcentagem de Fibra, porcentagem de açúcar na cana (PCC) e tonelada de açúcar por hectare (TPH), todos referentes a duas épocas de colheita: primeiro corte em agosto de 2003 (cana planta) e segundo corte (cana soca) em agosto de 2004. As análises estatísticas

foram conduzidas de forma separada para cada corte, considerando o efeito de blocos e progênies como aleatório. Médias ajustadas para cada genótipo foram obtidas para realizar o mapeamento em etapa posterior.

4.2.2 Marcadores moleculares e construção do mapa de ligação

Para a construção do mapa de ligação foram utilizados marcadores RFLP, descritos por Garcia et al. (2006), EST-RFLP e EST-SSR, descritos por Oliveira et al. (2007). O mapa de ligação foi construído usando abordagem multiponto (BUTCHER et al., 2002; LU; CUI; WU, 2004; WU et al., 2002b), a qual permite estimar a ordem e as possíveis fases de ligação entre os marcadores via cadeias de markov ocultas (HMM). Em síntese, 740 marcadores moleculares foram utilizados, sendo que 459 locos tinham segregação 1:1 (100 RFLP, 27 EST-RFLP, 332 EST-SSR) e 281 locos apresentavam segregação 3:1 (88 RFLP, 10 EST-RFLP, 183 EST-SSR). Seguindo a notação proposta por Wu et al. (2002a), os marcadores informativos apenas para o genitor SP80-180 foram denominados D_1 , com configuração ' $ao \times oo$ ', sendo o alelo a dominante em relação ao alelo o , nulo. Os locos informativos apenas para SP80-4966 foram denominados D_2 , com configuração ' $oo \times ao$ ', e os marcadores informativos para os dois genitores foram denominados C , com configuração ' $ao \times ao$ '. Os grupos de ligação foram estabelecidos com base em testes de dois pontos, utilizando para tanto LOD mínimo de 5,9 e fração de recombinação máxima de 0,5. Marcadores D_1 e D_2 foram considerados ligados quando apresentavam ligação em comum com marcadores do tipo C . O estabelecimento da ordem e das fases de ligação entre os marcadores em grupos de ligação com até cinco marcadores foi estimada comparando-se todas as ordens, de forma análoga ao realizado com o comando *compare* do software MAPMAKER/EXP (LANDER et al., 1987). Para grupos de ligação com seis ou mais marcas foram tomados cinco marcadores, cuja ordem foi determinada comparando todas as possíveis ordens, sendo os demais marcadores posicionados sequencialmente, a partir da ordem inicial, procedimento este análogo ao obtido com o comando *try* do MAPMAKER/EXP (LANDER et al., 1987). As distâncias em cM foram obtidas com a função de Kosambi (KOSAMBI, 1944). Com isto, obteve-se um mapa genético de 2468,14 cM, formado por 96 grupos de ligação, cuja densidade média de marcadores era 5,57 cM. Tais análises foram feitas com software Onemap (MARGARIDO; GARCIA; SOUZA, 2007) em sua nova versão, adaptada para também realizar análises multiponto, bem como estimar

as probabilidades condicionais de cada genótipo dos QTLs, necessários para o mapeamento.

4.2.3 Análise de marcas individuais

Com o intuito de comparar os resultados obtidos com o novo modelo, com a abordagem usualmente utilizada em cana-de-açúcar, também foi feito o mapeamento usando modelo de regressão linear para cada marcador:

$$y_j = \theta_0 + \theta x_j + \epsilon_j \quad (10)$$

em que, y_j : fenótipo do j -ésimo indivíduo; θ_0 : intercepto do modelo; θ : efeito genético do marcador; x_j : variável indicadora, que pode assumir valores 0 (ausência do alelo) e 1 (presença do alelo), para um dado marcador. A significância do efeito θ foi testada utilizando a estatística F com um grau de liberdade e para a apresentação dos resultados os p -valores foram convertidos em LOD Score. O nível de significância utilizado, para considerar presença de QTL foi LOD Score 3.

4.2.4 Mapeamento por intervalo composto

O mapeamento de QTLs foi conduzido usando modelo proposto por Gazaffi², para cada caráter e corte separadamente. Tal abordagem permite aplicar o mapeamento por intervalo composto considerando mapa genético formado por marcadores com diferentes padrões de segregações. Essencialmente, o mapeamento por intervalo composto (JANSEN; STAM, 1994; ZENG, 1994), desenvolvido para uso em populações originadas a partir do uso de linhagens endogâmicas, tem por objetivo estabelecer associações entre fenótipos e regiões do genoma, definida por um par de marcadores adjacentes. Marcadores moleculares localizados em outros grupos de ligação ou suficientemente distanciados do intervalo em estudo são incluídos no modelo para aumentar a precisão do mapeamento. Em progênies de irmãos completos, o número e a segregação dos alelos por loco pode variar; além disso, as fases de ligação entre QTLs e marcadores são desconhecidas. O modelo de Gazaffi² permite considerar os diferentes padrões de segregação que os QTLs podem assumir, bem como suas respectivas fases de ligação com os marcadores. A representação do cruzamento biparental considerado (Figura 5)

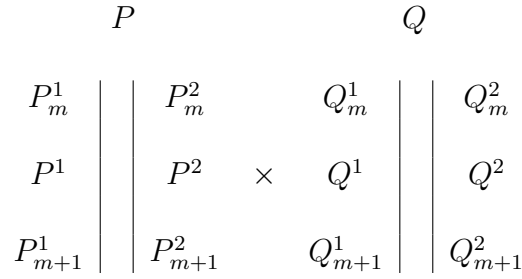


Figura 5 – Representação do tipo de cruzamento considerado. P e Q são os genitores, $P_m^{\{1,2\}}$, $Q_m^{\{1,2\}}$, $P_{m+1}^{\{1,2\}}$ e $Q_{m+1}^{\{1,2\}}$ os alelos dos marcadores dos locos m e $m+1$, e P^1 , P^2 , Q^1 e Q^2 são os alelos dos QTLs

permite definir o seguinte modelo genético-estatístico para mapeamento:

$$y_j = \mathbf{Z}_j \boldsymbol{\gamma} + \alpha_p^* x_{pj}^* + \alpha_q^* x_{qj}^* + \delta_{pq}^* x_{pj}^* x_{qj}^* + \epsilon_j \quad (11)$$

em que, y_j : fenótipo do j -ésimo indivíduo ($j = 1, \dots, n$); x_{pj}^* e x_{qj}^* : variáveis indicadores para o genótipo do QTL na progênie; α_p^* , α_q^* , δ_{pq}^* : efeitos aditivos para os genitores P , Q e interação intra loco, respectivamente; \mathbf{Z}_j : j -ésima linha da matriz indicadora \mathbf{Z} , com dimensão $n \times (1+3c)$, que contém uma coluna de 1's e variáveis indicadoras dos genótipos dos c cofatores (marcadores moleculares fora da região sendo mapeada); $\boldsymbol{\gamma}$: vetor de parâmetros $(1+3c) \times 1$ contendo intercepto (μ) e coeficientes de regressão linear múltipla (α_p , α_q e δ_{pq}) para os cofatores. No presente estudo, o genitor P refere-se à SP80-180 e o genitor Q refere-se à SP80-4966. Os valores que x_{pj}^* e x_{qj}^* podem assumir são:

$$x_{pj}^* = \begin{cases} 1 & \text{se } P^1 Q^1 \\ 1 & \text{se } P^1 Q^2 \\ -1 & \text{se } P^2 Q^1 \\ -1 & \text{se } P^2 Q^2 \end{cases}; \quad x_{qj}^* = \begin{cases} 1 & \text{se } P^1 Q^1 \\ -1 & \text{se } P^1 Q^2 \\ 1 & \text{se } P^2 Q^1 \\ -1 & \text{se } P^2 Q^2 \end{cases}$$

A presença de QTLs é testada ao longo do genoma, com a utilização do teste da razão de verossimilhança (LRT), com três graus de liberdade (busca por QTLs). Após a localização dos QTLs, hipóteses adicionais são consideradas para inferir a segregação do QTL detectado. Inicialmente, verifica-se a significância dos efeitos α_p^* , α_q^* e δ_{pq}^* através das hipóteses, $H_{01} : \alpha_p^* = 0$, $H_{02} : \alpha_q^* = 0$ e $H_{03} : \delta_{pq}^* = 0$, respectivamente. Para as situações em que há mais

de um efeito significativo, são testadas hipóteses para verificar a igualdade das estimativas (considerando seu sinal): $H_{04} : \alpha_p^* = \alpha_q^*$ (ou $\alpha_p^* = -\alpha_q^*$, ou $-\alpha_p^* = \alpha_q^*$, ou $-\alpha_p^* = -\alpha_q^*$), $H_{05} : \alpha_p^* = \delta_{pq}^*$ (ou $\alpha_p^* = -\delta_{pq}^*$, ou $-\alpha_p^* = \delta_{pq}^*$, ou $-\alpha_p^* = -\delta_{pq}^*$) e $H_{06} : \alpha_q^* = \delta_{pq}^*$ ou ($\alpha_q^* = -\delta_{pq}^*$, ou $-\alpha_q^* = \delta_{pq}^*$, ou $-\alpha_q^* = -\delta_{pq}^*$). A segregação dos QTLs pode ser inferida a partir da interpretação dos testes realizados, e a fase de ligação dos mesmos com os marcadores é identificada com a interpretação dos sinais de α_p^* e α_q^* . Para os testes complementares foi considerado 5% como nível de significância (*LOD Score* 1,2), uma vez que não há os problemas decorrentes dos múltiplos testes, já que apenas as posições com evidências de QTLs são considerados (JIANG; ZENG, 1995)

Para a seleção dos cofatores, foi realizada uma seleção de modelos utilizando critério de informação bayesiano, *BIC* (SCHWARZ, 1978). Como os marcadores utilizados no presente trabalho (*C*, *D*₁ ou *D*₂) são menos informativos que marcadores com segregação 1:1:1:1 e 1:2:1 foi necessário fazer uma modificação no procedimento sugerido por Gazaffi². Ao invés de considerar a entrada simultânea dos três efeitos (α_p , α_q e δ_{pq}) para cada marcador incluído no modelo, com posterior retirada dos efeitos não significativos, realizou-se a seleção considerando separadamente os efeitos de cada marcador. Para evitar problemas de superparametrização do modelo de mapeamento, procurou-se obter os modelos CIM com no máximo 20 parâmetros, limite recomendado em função tamanho da progênie, $2 \times \sqrt{n}$ (WANG; BASTEN; ZENG, 2007). Por fim, a proporção da variância fenotípica (R^2) que é explicada pelos QTLs mapeados (individualmente e simultaneamente) foi determinada ajustando um modelo com todos os QTLs simultaneamente. Para a realização das análises foi desenvolvido um software usando ambiente de programação R (R Development Core Team, 2008), que deve ser disponibilizada em breve.

4.3 Resultados

4.3.1 Dados fenotípicos

Antes de proceder o mapeamento de QTLs, algumas análises foram conduzidas para a caracterização dos fenótipos na progênie segregante. Observou-se que para o primeiro corte as médias para TPH e TCH foram 8,59 ton ha⁻¹ e 54,39 ton ha⁻¹, respectivamente, ambos com valores de coeficiente de variação (C.V.) próximos de 27%. Para o segundo corte, estas

médias subiram para 10,49 ton ha⁻¹ (TPH) e 71,50 ton ha⁻¹ (TCH), com C.V. de 30%. O caráter PCC apresentou média no primeiro corte de 15,88% e C.V. 7,67%; no segundo corte estes valores foram 14,71% e 7,27%. Ao avaliar o caráter teor de fibra, obteve-se média 11,46% e C.V. 13,13% no primeiro corte e média 10,17%, com C.V. 11,98% no segundo corte.

Foi verificado se havia a necessidade de transformação de dados para todos os fenótipos, utilizando a função *boxcox* do pacote estatístico R (R Development Core Team, 2008). Neste caso, observou-se que os caracteres para TPH, TCH, PCC não requeriam nenhuma transformação; para teor de fibra foi necessário obter o recíproco dos valores, para em seguida proceder o mapeamento.

4.3.2 Marcadores analisados individualmente

O mapeamento de QTLs realizado através do modelo de regressão linear identificou QTLs para todos caracteres, exceto teor de fibra no segundo corte. No total foram detectadas 11 associações entre fenótipos e marcadores (Tabela 10), sendo oito QTLs identificados em marcadores posicionados no mapa genético. De forma geral, nota-se que os *LOD* Scores foram baixos variando de 3,17 (PCC primeiro corte) a 4,25 (TPH primeiro corte). Os *R*² variaram de 3,86% a 5,69%. Isto mostra que a os marcadores analisados individualmente tem baixo poder de detecção de QTLs, como esperado.

Para o caráter TPH foram encontrados quatro QTLs, três no primeiro corte e um no segundo corte. Os QTLs em cana planta estavam localizados nos grupos de ligação 4 e 32, sendo um terceiro QTL identificado com marcador que não estava contido no mapa genético (SG61B). O QTL referente ao segundo corte foi posicionado no segundo grupo de ligação. Dentre os QTLs encontrados, apenas um apresentava efeito positivo (grupo de ligação 32), isto é, a presença do alelo originário do genitor SP80-4966 favorecia o aumento do caráter.

Para TCH foram encontrados três QTLs, dois no primeiro corte localizados no grupo de ligação 32 e no marcador SG61B, e um QTL para o segundo corte, no grupo de ligação 2. Vale destacar que estes três marcadores apresentam também associação com QTLs para TPH, o que deve ser ocorrido porque as características são correlacionadas. Ainda, os marcadores que apresentaram associação com ambas características tinham estimativas com mesmos sinais, o que é esperado, já que há correlação fenotípica positiva entre os fenótipos (0,95).

O caráter PCC apresentou associação com apenas três marcadores, sendo apenas um QTL

Tabela 10 – Mapeamento de QTLs com marcas analisadas individualmente

Corte	Caráter ⁽¹⁾	marcador	Grupo Lig.	<i>LOD</i>	$\hat{\theta}_0$	$\hat{\theta}$	$\hat{\sigma}^2$	R^2
cana planta	TPH	ESTC109m2D2	4	3,31	11,45	-1,76	4,49	4,30%
	TPH	ESTA48m2-D2	32	3,94	5,54	1,97	4,80	4,27%
	TPH	SG61B-C	-	4,25	11,98	-2,17	4,49	5,06%
	TCH	ESTA48m2-D2	32	4,04	34,41	12,84	198,01	4,27%
	TCH	SG61B-C	-	3,75	74,21	-12,67	174,38	5,06%
	PCC	ESTB111m2-C	6	3,17	17,18	-1,07	1,31	5,69%
	Fibra	ESTB94m6-D1	73	3,43	0,08	0,01	1,04 ⁽²⁾	4,00%
cana soca	TPH	ESTC134m2D2	2	3,62	14,27	-2,47	9,10	3,86%
	TCH	ESTC134m2-D2	2	3,29	95,65	-15,79	409,44	3,86%
	PCC	ESTC120m2-D2	3	3,43	13,56	0,80	0,99	3,95%
	PCC	ESTA69m1-D2	-	4,16	16,06	-0,86	0,91	4,08%

(1)TPH: Tonelada de POL por hectare; TCH: tonelada de cana por hectare; PCC: teor de açúcar na cana; Fibra: Teor de Fibra (%)

(2) Valor deve ser multiplicado por 10^{-4} .

associado ao primeiro corte (grupo de ligação 8) e dois (grupo de ligação 3 e ESTA69m2-D2) ao segundo corte. Dentre eles, o marcador encontrado no grupo de ligação oito apresenta efeito no qual a presença do alelo favorece o aumento do teor de açúcar no genitor SP80-4966. Para o teor de Fibra foi encontrado apenas um QTL posicionado no grupo de ligação 73, associado ao genitor SP80-180. Neste caso, a presença do alelo propiciava a redução do teor de Fibra.

De forma geral, nota-se que dentre os 11 marcadores significativos, apenas um é polimórfico apenas para o genitor SP80-180 (D_1), justamente aquele associado ao teor de Fibra; dois marcadores eram do tipo C , e todos os demais eram informativos somente ao genitor SP80-4966. Também notou-se que nas situações em havia mais de um marcador associado aos fenótipos, os efeitos genéticos eram relativamente próximos, o que não indicaria a presença de um QTL de efeito maior em relação aos demais. O mesmo pode ser estendido ao comparar mapeamento entre características e entre cortes, pois todos os QTLs apresentaram *LODs* variando entre 3,2 e 4,2.

Comparando-se o número de associações significativas entre marcadores e C.V. ao longo dos cortes pode se observar que os caracteres TPH e TCH apresentaram aumento nos valores de C.V. (cana planta: 27% e cana soca: 30%). Isto poderia explicar a redução no número de associações estabelecidas entre marcadores e fenótipos, pois houve uma redução na precisão

das estimativas. No entanto, esta tendência não foi verificada para PCC, pois o C.V. manteve-se estável (cana planta: 7,6% e cana soca: 7,2%) e houve o aumento de um QTL entre cana planta e cana soca. Para teor de fibra houve a redução do C.V. de cana planta (13%) para cana soca (12%), mas não houve a detecção de QTLs no segundo corte.

4.3.3 Mapeamento por intervalo composto

Ao utilizar *LOD* Score 3 como limiar foram encontrados no total 41 QTLs, sendo 18 detectados no primeiro corte (Tabela 11, Figuras 6 e 7) e 23 no segundo corte (Tabela 12, Figuras 8 a 10). De forma geral, as características com mais QTLs identificados foram TPH (12 QTLs, sendo 8 no primeiro e 6 no segundo corte) e PCC (12 QTLs, sendo 5 no primeiro e 7 no segundo corte). O caráter com menos QTLs mapeados foi teor de Fibra, com apenas um QTL no primeiro corte e com 6 no segundo. Os valores de *LOD* Score variaram de 3,1 a 10,4 e os R^2 variaram de 0,53% a 35,06%. A maioria dos QTLs mapeados apresentavam segregação 1:1, muito possivelmente porque a maioria dos marcadores também possui esse padrão de segregação.

Primeiro Corte

Para TPH houve a identificação de oito QTLs, nos grupos de ligação 8 (44 cM), 16 (14,5 cM), 20 (0,0 cM), 21 (0,0 cM), 26 (0,0 cM), 32 (26,3 cM), 69 (11,3 cM) e 90 (4,1 cM), explicando conjuntamente 54,9% da variação fenotípica. Nota-se o QTL localizado no grupo de ligação 32 apresentou maior valor de *LOD* (10,4) dentre todos os QTLs mapeados no presente estudo, explicando 8,89% da variação fenotípica. Porém, o gráfico não detalha se a região com o QTL está localizada no marcador ESTA48m2-D2 (23,68 cM), pois o comportamento da curva ainda é ascendente e não há mais marcadores à direita. QTLs mapeados em regiões terminais também foram encontrados nos grupos de ligação 20, 21 e 26, porém com valores de *LOD* Score menores (4; 6,1 e 5,5; respectivamente).

Com relação à segregação dos QTLs, observa-se a predominância do padrão 1:1 (grupos de ligação 8, 16, 20, 21, 26 e 32), destacando o QTL posicionado no grupo de ligação 16, cuja fase de ligação é não identificável, pois somente o efeito de interação é significativo. O único QTL mapeado para o genitor SP80-180 foi posicionado no grupo de ligação 21 e estava em repulsão. Os outros quatro QTLs com este mesmo padrão de segregação eram referentes ao genitor SP80-4966, sendo dois grupos de ligação 8 e 26 com fase de ligação em associação e

outros dois em repulsão (grupos de ligação 20 e 32). Já o QTL mapeado no grupo de ligação 69 (com *LOD* Score 3,4) apresentou segregação 3:1, uma vez que os três efeitos genéticos do QTL foram significativos e estatisticamente idênticos, com alelos dos QTLs em repulsão para ambos genitores.

Para o caráter TCH, foram localizados quatro QTLs nos grupos de ligação 23 (0,0 cM), 32 (26,0 cM), 78 (0,0 cM) e 90 (0,0 cM), explicando conjuntamente 54,89% da variação fenotípica. O QTL com maior *LOD* Score (9,3) foi detectado no grupo de ligação 90, o qual explicava 29,19% da variação fenotípica do caráter. Esse também foi o único QTL que apresentou segregação 1:2:1, com efeitos α_q^* e δ_{pq}^* significativos. Para os QTLs mapeados no grupo de ligação 23 e 32 o alelo que favorecia o caráter era originário de SP80-4966, sendo o primeiro em associação e o segundo em repulsão. O QTL detectado para grupo de ligação 78 apresentou alelos originários do genitor SP80-180 em associação. Destaca-se ainda que os QTLs identificados para os grupos de ligação 32 e 90 também foram encontrados para o caráter TPH, em que as segregações e fases de ligação foram análogas.

Os resultados para teor de Fibra apontaram apenas um QTL localizado no grupo de ligação 73, com segregação 1:1 e alelos favoráveis presentes em associação no genitor SP80-180 e explicando apenas 1,1% da variação fenotípica. Novamente, por se tratar de um grupo de ligação pequeno, o posicionamento correto do QTL é dificultado.

Para PCC houve a detecção cinco regiões localizadas nos grupos de ligação 1, 3, 6, 11 e 81, explicando 45,5% da variação fenotípica do caráter. Todos os QTLs apresentavam segregação 1:1, sendo três para o genitor SP80-4966 em repulsão (1, 11 e 81) e dois cujas fases não podiam ser determinadas pela ausência de efeitos marginais significativos (3 e 6). Todos os picos de *LOD* Score ao longo dos grupos de ligação demonstram que os marcadores utilizados para a construção do mapa conseguiram flanquear satisfatoriamente as regiões controladoras deste caráter.

Segundo Corte

O caráter TPH foi o único que demonstrou redução no número de QTLs mapeados no segundo corte, pois foram detectados seis QTLs, nos grupos de ligação 2 (118,6 cM), 8 (45,0 cM), 20 (19,0 cM), 32 (26,3 cM), 48 (18,1 cM) e 90 (5,0 cM) (Tabela 12 e Figura 8). No entanto, foram detectada a presença de três outras regiões responsáveis pelo controle do fenótipo TPH, as quais estão posicionadas nos grupos de ligação 2, 20 e 48. Estes seis QTLs

controlam 26,5% da variação fenotípica do carácter. Para os QTLs posicionados nos grupos de ligação 8, 32 e 90, observou-se que o perfil da curva e os testes para as estimativas dos efeitos e segregações foram muito semelhantes as obtidas no primeiro corte, indicando provavelmente QTLs estáveis ao longo dos cortes. Nota-se ainda que no grupo de ligação 20 há um QTL no segundo corte na posição 19 cM, que é distinta em relação ao primeiro corte (0,0 cM). Este resultado é relevante, pois pode indicar uma presença de expressão diferencial de genes localizados neste grupo de ligação, ao longo dos cortes.

O QTL detectado para TPH no grupo de ligação 2 apresentou resultados interessantes, pois foi detectado um QTL em 118,58 cM com *LOD* Score 9,3 e segregação 1:1:1:1, sendo os dois efeitos marginais (α_p^* e α_q^*) significativos, porém sem efeito de dominância. Este resultado somente foi possível de ser obtido porque no grupo de ligação havia marcadores com os três tipos de segregações (C , D_1 e D_2), mostrando que a abordagem multiponto para obtenção das probabilidades condicionais conseguiu recuperar a informação contida na região.

Pode-se destacar o QTL posicionado no grupo de ligação 90, que apresentou segregação 1:2:1, com alelo favorável ligado em repulsão no genitor SP80-4966 e com efeito de dominância significativo, explicando 3% da variação fenotípica, e apresentando a mesma segregação e fase em relação ao primeiro corte. Os demais QTLs apresentam segregação 1:1 com um QTL originário de SP80-180 (grupo de ligação 48) e outros três provenientes de SP80-4966, sendo que o QTL mapeado no grupo de ligação 8 apresentava fase de ligação em associação, e os posicionados em 20 e 32 em repulsão.

Para TCH houve a detecção de QTLs nos grupos de ligação 32 (26,0 cM), 88 (0,0 cM), 90 (0,0 cM) e 96 (0,0 cM) (Tabela 12 e Figura 9), explicando 46% da variação fenotípica. Destaca-se que para este carácter não houve o claro posicionamento dos QTLs ao longo do grupos de ligação 88, 90 e 96, resultado do pequeno número de marcadores existentes e também pela possível localização do QTL em regiões adjacentes ao marcadores. Em todos os quatro QTLs mapeados verificou-se segregação 1:1, sendo os dois primeiros segregantes para o genitor SP80-4966 (ambos em repulsão), e os dois últimos segregantes para SP80-180, em que o QTL mapeado no grupo de ligação 90 está em associação; o QTL posicionado no grupo de ligação 96 está ligado em repulsão.

Para teor de Fibra houve a detecção de seis QTLs, localizados nos grupos de ligação 7 (21,0 cM), 10 (11,2 cM), 68 (11,4 cM), 71 (0,0 cM), 78 (8,2 cM), 92 (4,0 cM) e explicando

11% da variação fenotípica do caráter. Novamente, é possível perceber que o mapeamento para grupos de ligação apresentaram valores de *LOD* Score próximos do limiar, além do comportamento ascendente da curva de mapeamento com a localização de QTLs em posições terminais, o que pode sugerir a presença de QTL fora da região coberta por marcadores.

Com relação a segregação encontrada destacam-se os QTLs mapeados nos grupos de ligação 10 e 78, que apresentaram segregação 1:2:1 devido a significância de α_q^* e δ_{pq}^* . Esses foram os dois QTLs que apresentaram maiores valores de R^2 (3,44% e 3,06%, respectivamente), indicando que a maior contribuição do QTL para a variação do caráter ocorre quando há mais efeitos presentes. Os outros quatro QTLs apresentam segregação 1:1, em que um QTL (grupo de ligação 71) está presente no genitor SP80-180 em repulsão, e três QTLs estão no genitor SP80-4966, sendo um deles em repulsão (grupo de ligação 68) e três em associação (grupos de ligação 7, 78 e 92)

Para PCC foram detectados sete QTLs nos grupos de ligação 1 (101,8 cM), 3 (33,0 cM), 6 (14,0 cM), 30 (21,0 cM), 35 (23,0 cM), 67 (0,0 cM), 81 (2,0 cM) (Tabela 12 e Figura 10). Todas estas regiões explicaram conjuntamente 26,19% da variação fenotípica. No entanto, os QTLs mapeados nos grupos de ligação 1 e 6 estão localizados em posições distintas em relação ao primeiro corte (15,7 e 67 cM, respectivamente), o que sugere a presença de QTLs distintos localizados nos mesmos grupos de ligação. Os QTLs posicionados nos grupos de ligação 3 e 6 apresentaram maiores valores de *LOD* Score (7,2 e 5,1, respectivamente), assim como maiores valores de R^2 (2,67% e 18,39%), o que poderia ser explicado pelo efeitos dos QTLs, já que os mesmos possuíam geralmente mais de um efeito significativo.

Os QTLs dos grupos de ligação 1, 3, 30, 35, 67 e 81 possuem segregação 1:1, sendo apenas o QTL mapeado no grupo de ligação 35 associado ao genitor SP80-180 com ligação em repulsão; os demais segregavam para SP80-4966, sendo dois em associação (grupos de ligação 1 e 30) e três em repulsão (grupos de ligação 3, 67 e 81).

Tabela 11 – QTLs mapeados com modelo CIM, para características TPH, TCH, PCC e Fibra (primeiro corte)

Caráter	G.L. (1)	cM (2)	Posição										Efeitos			Conclusão					
			Marcadores flaqueadores										LOD	$\hat{\mu}$	$\hat{\alpha}_p^*$	LOD	$\hat{\alpha}_q^*$	LOD	$\hat{\delta}_{pq}^*$	LOD	$\hat{\delta}^{2*}$
TPH	8	44,0	ESTA47m2-D2	ESTC109m2-D2	7,3	9,38	0,00	0,00	0,70	7,35	0,00	0,96	1:1	Q^1Q^2	0,55%						
TPH	16	14,5	ESTB131m2-C	ESTA53m8-C	5,3	9,53	-0,31	1,01	-0,31	1,01	0,36	0,82	1:1	n.d.(4)	12,84%						
TPH	20	0,0	ESTA53m4-D2	ESTA14m3-D2	4,0	9,34	0,00	0,00	-0,52	3,95	0,00	1,14	1:1	Q^2Q^1	0,53%						
TPH	21	0,0	SG26D-C	SG23B-C	6,1	9,40	-0,59	6,11	0,00	0,00	0,00	0,95	1:1	P^2P^1	0,58%						
TPH	26	0,0	SG26B-C	SG23C-C	5,5	9,38	0,00	0,00	0,57	5,53	0,00	0,94	1:1	Q^1Q^2	0,53%						
TPH	32	26,3	ESTA63m3-D2	ESTA48m2-D2	10,4	9,46	-0,03	0,00	-0,73	7,98	0,03	1,04	1:1	Q^2Q^1	8,89%						
TPH	69	11,3	SG41H-C	SG94H-C	3,3	9,38	-0,25	1,36	-0,25	1,36	0,27	1,07	3:1(5)	$P^2P^1 \times Q^2Q^1$	29,00%						
TPH	90	4,1	ESTB41m7-D1	ESTB41m1-C	3,4	9,43	-0,06	0,34	-0,33	1,60	0,33	0,86	1:2:1(6)	Q^2Q^1	1,96%						
															54,92%						
TCH	23	0,0	ESTC80m2-D2	ESTB80m2-D2	5,1	50,93	0,00	0,00	3,44	5,13	0,00	0,74	1:1	Q^1Q^2	0,52%						
TCH	32	26,3	ESTA63m3-D2	ESTA48m2-D2	7,8	52,80	-0,83	0,09	-3,58	5,09	0,65	0,47	1:1	Q^2Q^1	8,31%						
TCH	78	0,0	ESTA34m8-D1	ESTA34m1-C	3,7	51,90	3,20	3,72	-0,16	0,09	-0,29	0,12	1:1	P^1P^2	16,87%						
TCH	90	0,0	ESTB41m7-D1	ESTB41m1-C	9,3	51,72	1,39	0,55	-4,72	3,83	4,42	3,59	1:2:1(7)	Q^2Q^1	29,19%						
															54,89%						
PCC	1	15,7	ESTA08m1-D2	ESTC81m4-D2	3,2	14,00	0,00	0,00	-0,30	3,15	0,00	0,53	1:1	Q^2Q^1	0,83%						
PCC	3	41,2	ESTB56m4-D1	ESTB40m2-C	4,1	14,48	-0,02	0,00	-0,19	0,79	0,36	2,59	1:1	n.d.(4)	3,91%						
PCC	6	67,0	ESTA63m1-C	ESTB111m2-C	8,4	14,28	0,26	0,94	0,26	0,94	-0,35	3,03	1:1	n.d.(4)	19,70%						
PCC	11	35,6	ESTB63m2-D2	ESTA17m1-D2	3,1	13,44	0,00	0,00	-0,26	3,24	0,00	0,48	1:1	Q^2Q^1	0,91%						
PCC	81	2,0	ESTC113m-D2	ESTC24m1-D2	6,7	14,42	0,00	0,00	-0,42	7,02	0,00	0,37	1:1	Q^2Q^1	0,86%						
															26,1%						
Fibra	73	5,0	ESTA54m3-D1	ESTB94m6-D1	4,3	0,09	-2,67(8)	4,31	0,00	0,00	0,00	3,96(9)	1:1	P^1P^2	1,10%						

(1) GL: Grupo de Ligação.

(2) Distância em relação ao primeiro marcador à esquerda do respectivo grupo de ligação.

(3) São apresentados apenas os alelos do QTL; por exemplo, P^1P^2 significa configuração $P^1P^1P^1/P^2P^2P^2$.

(4) n.d.: Fase de ligação não determinável, pois não há efeitos marginais significativos.

(5) As Hipóteses H_{04} , H_{05} e H_{06} foram não rejeitadas (P -valores 0,13; 0,15 e 0,15, respectivamente).(6) A Hipótese H_{06} foi não rejeitada (P -valor 0,47).(7) A Hipótese H_{06} foi não rejeitada (P -valor 1,00).(8) Valor deve ser multiplicado por 10^{-3} .(9) Valor deve ser multiplicado por 10^{-5} .

Tabela 12 – QTLs mapeados com modelo CIM, para características TPH, TCH, PCC e Fibra (segundo corte).

Carácter	Posição			Efeitos										Conclusão	
	G.L. (1)	cM (2)	Marcas flaqueadores	LOD	$\hat{\mu}$	$\hat{\alpha}_p^*$	LOD	$\hat{\alpha}_q$	LOD	$\hat{\delta}_{pq}^*$	LOD	$\hat{\delta}_2^*$	Seg.	Fase (3)	R^2
TPH	2	118,6	ESTA66m3-C ESTC134m2-D2	9,3	10,02	0,28	1,78	0,92	6,19	-0,10	2,08	1:1:1:1(4)	$P^1 P^2 \times Q^1 Q^2$	6,12%	
TPH	8	45,0	ESTA47m2-D2 ESTC109m2-D2	7,2	9,88	0,00	0,00	0,93	6,73	0,00	2,04	1:1	$Q^1 Q^2$	0,92%	
TPH	20	19,0	ESTC132m1-D2 ESTC131m1-D2	6,2	11,18	0,00	0,00	-0,93	6,14	0,00	2,12	1:1	$Q^2 Q^1$	0,95%	
TPH	32	26,3	ESTA63m3-D2 ESTA48m2-D2	9,9	10,22	-0,24	0,17	-0,99	7,65	0,17	1,98	1:1	$Q^2 Q^1$	3,81%	
TPH	48	18,1	SG56B-C SG155A-C	4,3	10,03	-0,68	4,20	0,00	0,00	0,00	1,98	1:1	$P^2 P^1$	0,87%	
TPH	90	5,0	ESTB41m7-D1 ESTB41m1-C	5,0	9,92	0,14	0,68	-0,55	2,49	0,54	1,61	1:2:1(5)	$Q^2 Q^1$	2,99%	
														26,5%	
TCH	32	26,0	ESTA63m3-D2 ESTA48m2-D2	5,6	103,81	-1,51	0,28	-5,11	4,73	0,95	119,00	1:1	$Q^2 Q^1$	9,83%	
TCH	88	0,0	ESTB60m5-D2 ESTB92m1-D2	3,4	118,83	0,00	0,00	-3,77	3,36	0,00	131,80	1:1	$Q^2 Q^1$	0,61%	
TCH	90	0,0	ESTB41m7-D1 ESTB41m1-C	4,1	165,07	7,53	2,61	-3,00	0,99	3,35	1288,43	1:1	$P^1 P^2$	35,06%	
TCH	96	0,0	ESTA34m9-D1 ESTA34m7-D1	3,6	76,02	-4,44	3,58	0,00	0,00	0,00	123,59	1:1	$P^2 P^1$	0,56%	
														46,00%	
PCC	1	101,8	ESTA15m4-D2 ESTB39m11-D2	3,4	14,11	0,00	0,00	0,23	3,40	0,00	0,29	1:1	$Q^1 Q^2$	0,83%	
PCC	3	33,0	ESTC12m1-C ESTC120m2-D2	7,2	14,67	0,12	0,58	-0,34	6,00	0,13	0,53	1:1	$Q^2 Q^1$	2,67%	
PCC	6	14,0	ESTB07m1-C ESTA68m2-C	5,1	13,94	0,22	1,92	0,22	1,92	-0,20	0,30	3:1(6)	$P^1 P^2 \times Q^1 Q^2$	18,39%	
PCC	30	21,0	SG96G-C ESTA07m1-D2	4,1	14,29	0,00	0,00	0,26	4,28	0,00	0,26	1:1	$Q^1 Q^2$	0,88%	
PCC	35	23,0	ESTB69m2-D1 ESTB65m3-D1	4,1	14,11	-0,25	4,06	0,00	0,00	0,00	0,26	1:1	$P^2 P^1$	0,87%	
PCC	67	0,0	ESTA55m4-D2 ESTC109m5-D2	4,2	14,84	0,00	0,00	-0,25	4,24	0,00	0,31	1:1	$Q^2 Q^1$	0,80%	
PCC	81	2,0	ESTC113m-D2 ESTC24m1-D2	3,7	13,72	0,00	0,00	-0,31	3,71	0,00	0,55	1:1	$Q^2 Q^1$	0,88%	
														26,19%	
Fibra	7	21,0	ESTA03m1-D2 ESTC22m5-D2	4,2	0,10	0,00	0,00	2,11(7)	3,48	0,00	4,20(8)	1:1	$Q^1 Q^2$	1,22%	
Fibra	10	11,2	SG12D-C ESTC30m3-D1	5,4	0,11	0,77(7)	0,20	-2,00	2,00	1,96(7)	3,90(8)	1:2:1(9)	$Q^2 Q^1$	3,44%	
Fibra	68	11,4	ESTC42m2-D2 ESTB16m2-D2	3,2	0,11	0,00	0,00	-2,00(7)	3,01	0,00	3,40(8)	1:1	$Q^2 Q^1$	0,95%	
Fibra	71	0,0	ESTC42m3-D1 ESTC21m1-D1	7,4	0,11	-3,82(7)	7,35	0,00	0,27	0,00	3,40(8)	1:1	$P^2 P^1$	0,95%	
Fibra	78	8,2	ESTA34m8-D1 ESTA34m1-C	5,1	0,11	-0,06(7)	0,04	2,30	1,81	2,00(7)	1,72	1:2:1(10)	$Q^1 Q^2$	3,06%	
Fibra	92	4,0	ESTB65m1-D2 ESTC44m1-D2	3,9	0,10	0,00	0,00	2,41(7)	3,70	0,00	3,60(8)	1:1	$Q^1 Q^2$	0,95%	
														11,63%	

(1) GL: Grupo de Ligação.

(2) Distância em relação ao primeiro marcador à esquerda do respectivo grupo de ligação.

(3) São apresentado apenas os alelos do QTL; por exemplo, $P^1 P^2$ significa configuração $P_m^1 P_{m+1}^1 / P_m^2 P_{m+1}^2$.

(4) As Hipóteses H_{04} , H_{05} e H_{06} foram rejeitadas (P -valores $3,52 \times 10^{-4}$, $1,57 \times 10^{-2}$ e $9,42 \times 10^{-5}$, respectivamente).

(5) A Hipótese H_{06} foi não rejeitada (P -valor de 0,29).

(6) As Hipóteses H_{04} , H_{05} e H_{06} foram não rejeitadas (P -valores 0,18; 0,15 e 0,15, respectivamente).

(7) Valores devem ser multiplicados 10^{-3}

(8) Valores devem ser multiplicados 10^{-5}

(9) A Hipótese H_{06} foi não rejeitada (P -valor de 1,00)

(10) A Hipótese H_{06} foi não rejeitada (P -valor de 0,25).

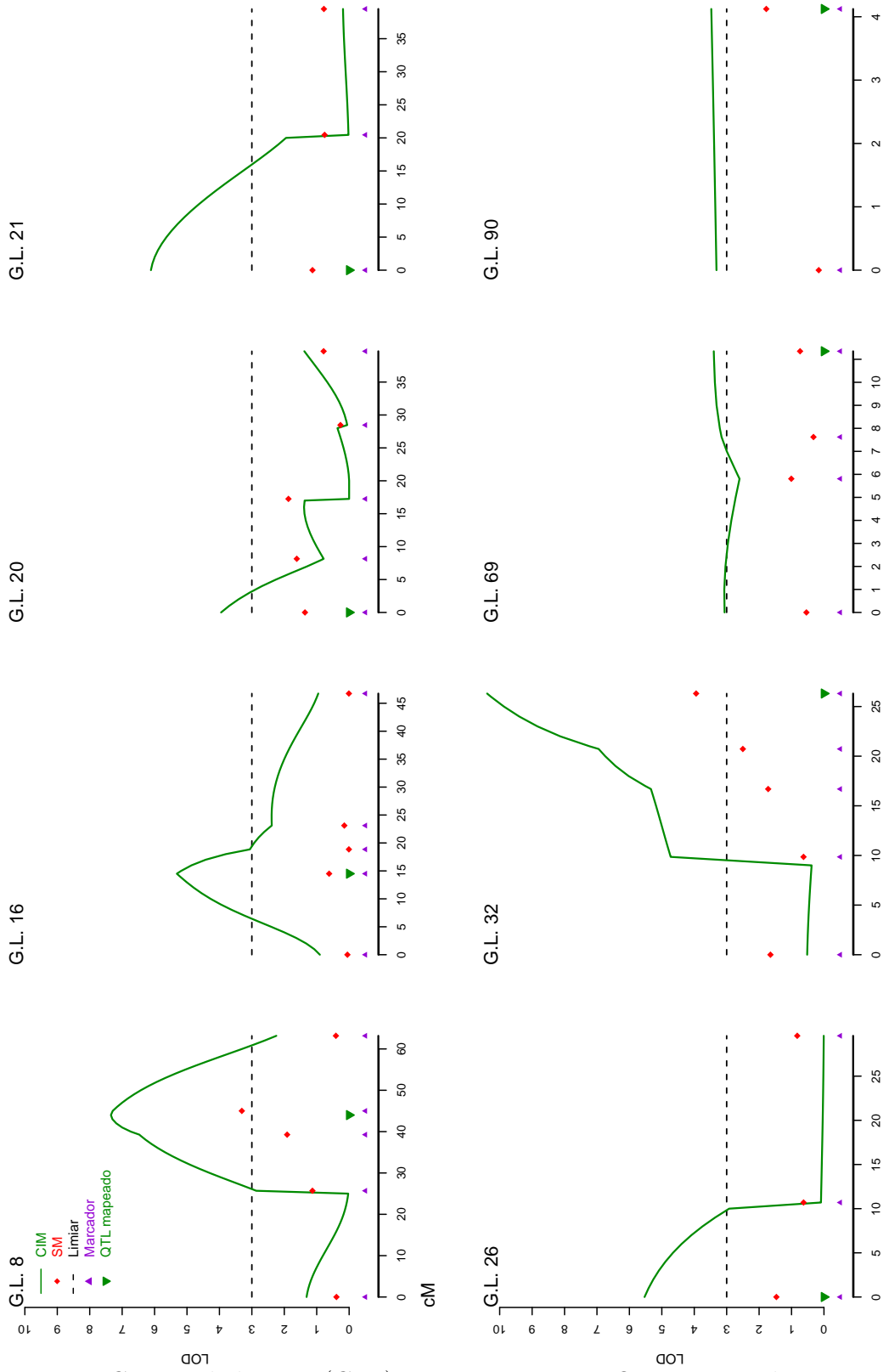


Figura 6 – Grupos de ligação (G.L.) que apresentaram QTLs mapeados para o caráter TPH, no primeiro corte

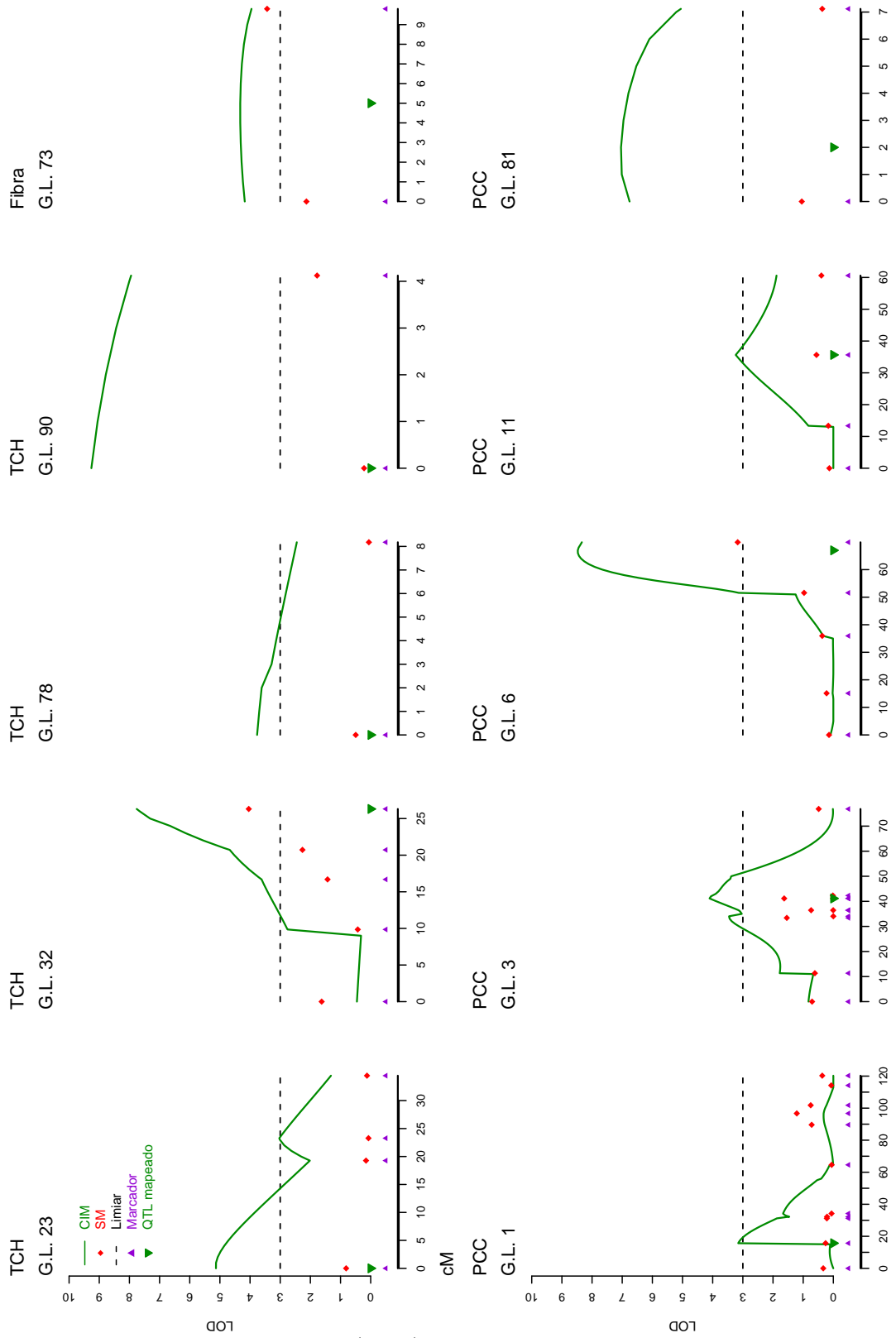


Figura 7 – Grupos de ligação (G.L.) que apresentaram QTLs mapeados para os caracteres TCH, PCC e Fibr (primeiro corte)

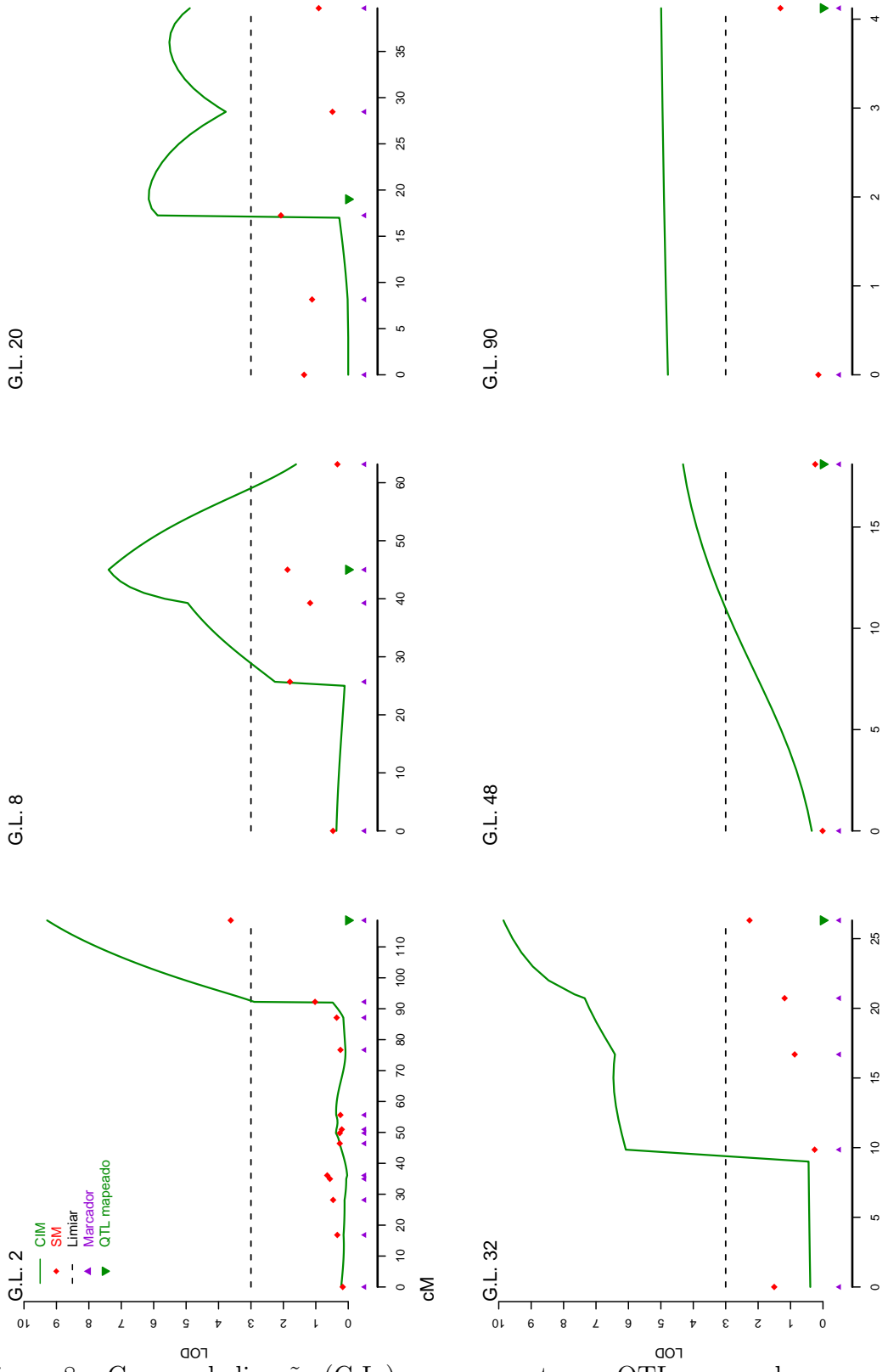


Figura 8 – Grupos de ligação (G.L.) que apresentaram QTLs mapeados para o caráter TPH (segundo corte)

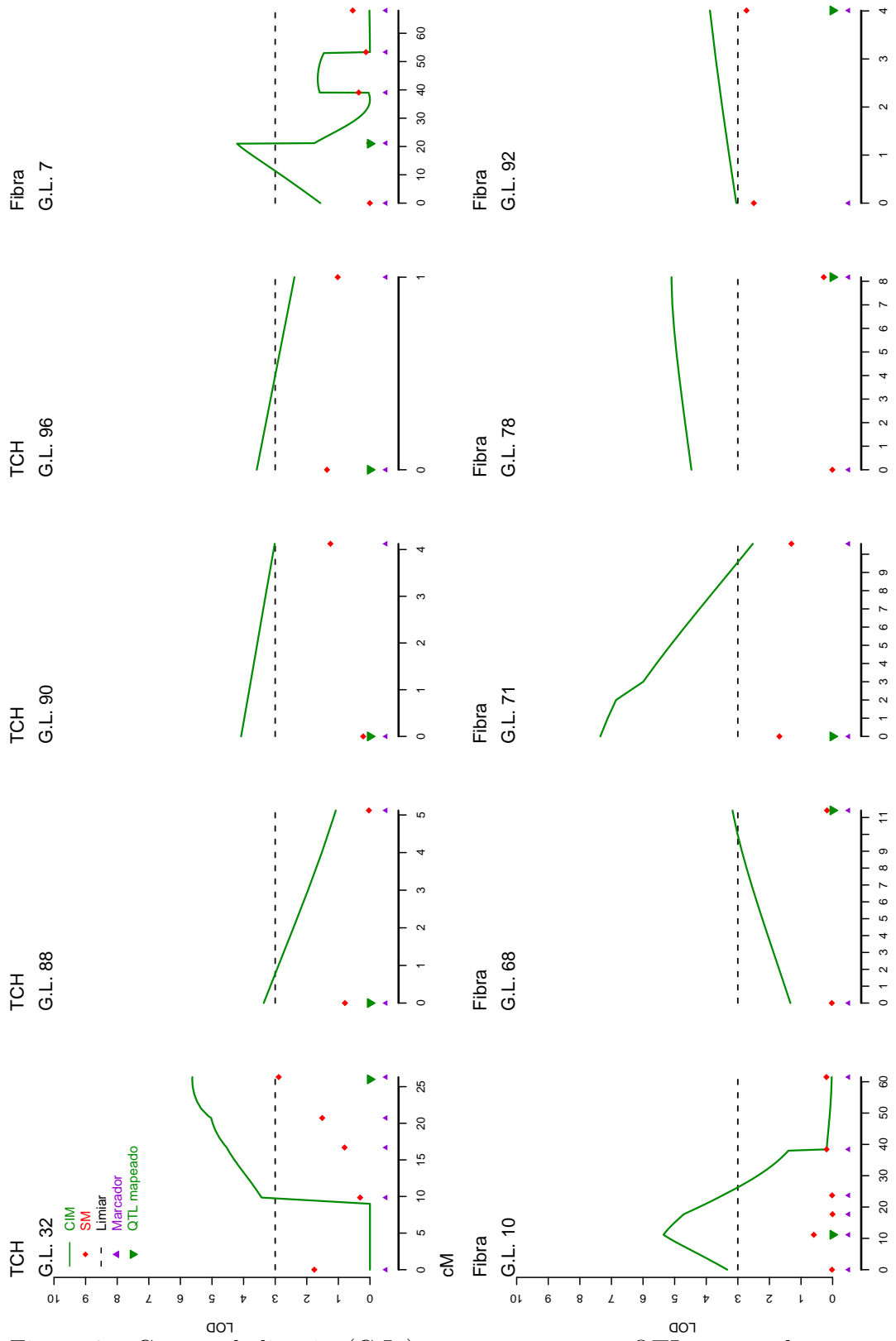


Figura 9 – Grupos de ligação (G.L.) que apresentaram QTLs mapeados para o caráter TCH e Fibra (segundo corte)

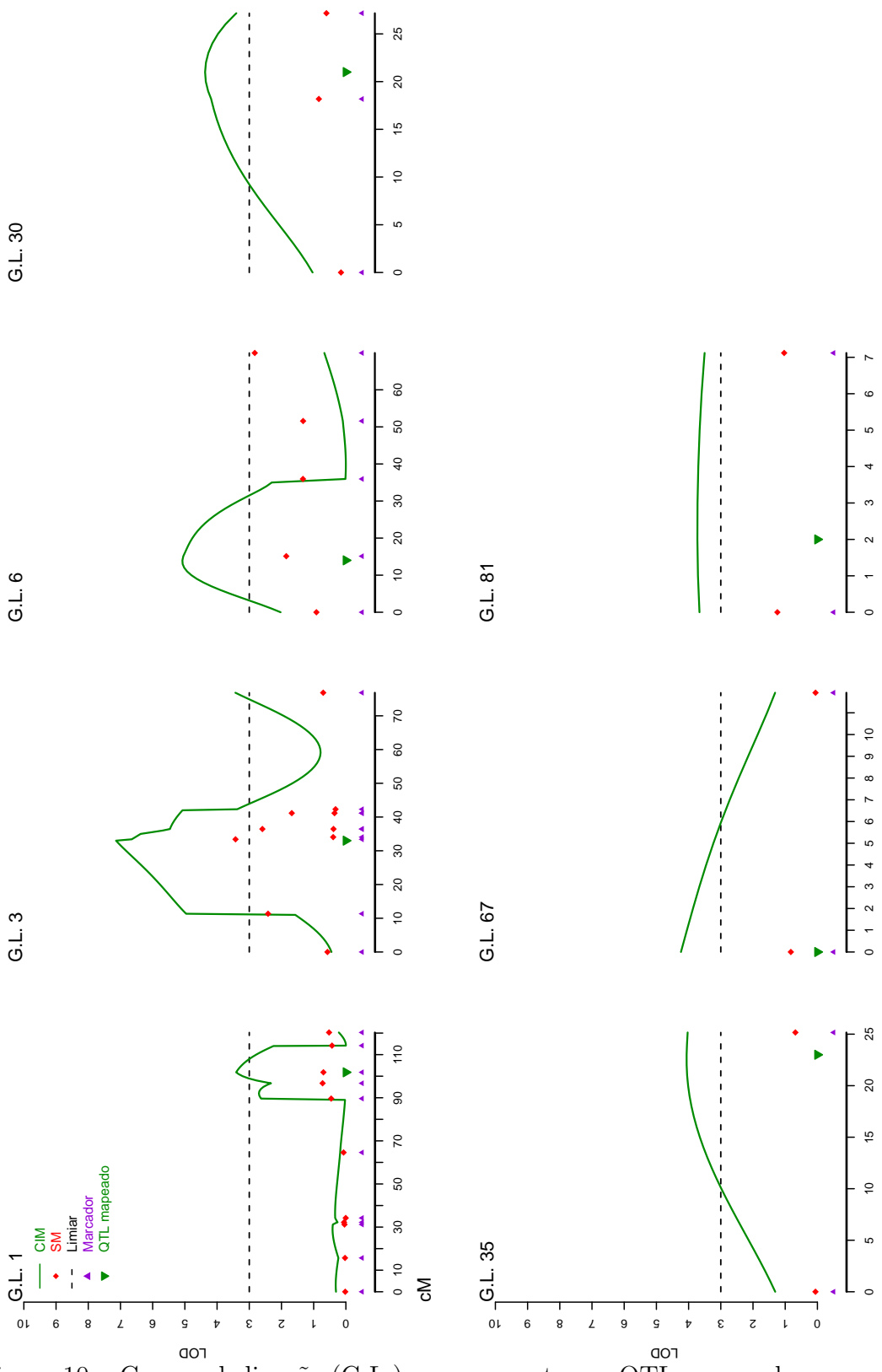


Figura 10 – Grupos de ligação (G.L.) que apresentaram QTLs mapeados para o caráter PCC (segundo corte)

4.4 Discussão

No presente estudo realizou-se o mapeamento de QTLs utilizando modelo CIM (GA-ZAFFI²), a partir de um mapa construído com marcadores de diferentes padrões de segregação (1:1 e 3:1). De forma geral, esta abordagem permitiu o mapeamento de um número elevado de QTLs (43), maior que o obtido pela análise de marcas individuais (11). Esse aumento provavelmente é devido ao maior poder estatístico que o modelo CIM possui para mapear QTLs. Isto pode ser verificado pelo exame dos gráficos, que indicam que os *LOD* Scores obtidos com o CIM são superiores aos obtidos com as marcas analisadas individualmente. O modelo CIM geralmente resultou em valores de *LOD* Score com cerca de duas a três vezes maiores.

Os genitores SP80-180 e SP80-4966 apresentavam fenótipos contrastantes e complementares, porém a análise de marcas individuais detectou mais alelos de QTLs associados a marcadores que segregavam no genitor SP80-4966. Esta tendência também foi confirmada no mapeamento por intervalo composto, pois dentre os 41 efeitos marginais significativos, apenas 11 (26,8%) eram referentes ao genitor SP80-180. Isto possivelmente ocorre pois algumas regiões que continham QTLs somente apresentavam um tipo de marcador. Por exemplo, o grupo de ligação 1, no qual foi mapeado dois QTLs para PCC apresentava marcadores apenas do tipo D_2 (informativos para SP80-4966). Neste caso, mesmo que houvesse a presença de um alelo do QTL para o genitor SP80-180, o mesmo não poderia ser detectado, já que não há informação para tanto. Situação análoga ocorre no grupo de ligação 32 que apresentou QTLs para TPH e TCH nos dois cortes, mas havia apenas a presença de marcadores D_2 e C , o que dificultou a detecção de alelos originários de SP80-180. Isso torna clara, a importância de se obter bons mapas genéticos, com alelos segregantes nos dois genitores, o que ainda é um desafio em cana-de-açúcar.

Para os quatro caracteres mapeados verificou-se que a ação gênica predominante do caráter é aditividade, pois no geral os maiores valores de *LOD* estiveram associados com efeitos α_p^* ou α_q^* . Em apenas três situações o modelo apontou maiores *LOD* Score para dominância, todas no primeiro corte (PCC, 2 QTLs; TPH, 1QTL). Nos outros sete QTLs que apresentaram efeito de dominância significativo, os efeitos eram de dominância completa, pois o efeito aditivo em SP80-4966 foi estatisticamente igual a dominância, tais QTLs estavam distribuídos

de forma aleatória entre os caracteres (Tabelas 11 e 12).

Para as fases de ligação há um equilíbrio entre QTLs em associação (5 QTLs) e repulsão (6 QTLs) para o genitor SP80-180. Para o genitor SP80-4966 há uma maior tendência em detectar QTLs em repulsão (17 QTLs), em relação à associação (11 QTL). Esta diferença ocorre principalmente para o caráter PCC no primeiro corte. Esse resultado é muito importante caso deseje-se estudar as regiões a DNA que contém os alelos favoráveis; que puderam ser identificados. O mesmo ocorre caso trabalhos de MAS sejam conduzidos no futuro.

Com relação as estimativas de R^2 é possível verificar que houve variação sobre o quanto que os QTLs explicavam da variação fenotípica para um dado caráter. Os maiores valores foram encontrados para TPH e TCH, no primeiro corte (55%), que apresentaram redução nos valores no segundo corte (TPH: 26% e TCH: 46%). Para PCC houve um estimativa de R^2 semelhantes para ambos os cortes (próxima a 26%) e teor de fibra apresentou um aumento de R^2 de 1% a 11,6% do primeiro para o segundo corte. A comparação dos valores R^2 obtidos com os de outros trabalhos é dificultada, pois as populações, marcadores e ambientes são muito diferentes. Ming et al. (2001; 2002b) mapeando QTLs em diferentes cruzamentos, obtiveram estimativas de R^2 para TPH que variavam entre 30% a 68%, o que está de acordo com os valores obtidas no presente estudo. Sills et al. (1995) e Ming et al. (2002b) encontraram estimativas de R^2 para teor de Fibra próximos a 60%. De qualquer forma, os valores aqui apresentados são muito promissores, uma vez que o mapa continua sendo saturado com outros marcadores e as populações continuam sob avaliação fenotípica. A eventual inclusão de mais dados, com a expansão do modelo para múltiplos caracteres e ambientes podem implicar em grandes aumentos de R^2 .

Os resultados para o primeiro corte indicam a possível existência de dois QTL relacionados na expressão dos caracteres TPH e TCH nos grupos de ligação 32 e 90. No grupo de ligação 32 pode haver indicação da existência de QTL com efeito pleiotrópico, pois as curvas de mapeamento apresentam o mesmo padrão, assim como as segregações e fase de ligação entre marcadores e QTLs. Contudo, o mesmo não ocorre para o grupo de ligação 90, muito embora os resultados não sejam evidentes, pois o reduzido tamanho do grupo de ligação (4,2 cM) dificulta o correto posicionamento dos QTLs. No segundo corte, observa-se que os mesmos grupos mostraram evidências de QTLs para TPH e TCH, mas as interpretações no entanto, ao considerar TCH não são simples dada a baixa quantidade de marcadores.

Uma comparação entre os cortes também é possível. Para TPH houve uma redução no número de QTLs detectados, o que explicaria a redução dos valores de R^2 do primeiro (54,9%) para o segundo corte (26,5%), porém 3 QTLs que foram mapeados em ambas as situações (grupos de ligação 8, 32 e 90), sendo dois deles com LOD Score elevados (acima de 7,0). O grupo de ligação 20 indicou a presença de QTLs tanto para primeiro corte (0 cM) como para segundo corte (19 cM), sendo possivelmente o resultado de interação entre QTL e corte. Para TCH, há também evidências de interação entre QTL e corte, já que há casos em que a posição (ou presença) do QTL mudou em cana-soca, por exemplo, o grupo de ligação 3 apresentou um QTL para cada corte localizados em regiões próximas (41 cM e 33 cM, no primeiro e segundo corte, respectivamente).

O fenótipo teor de fibra também evidências de interação, pois assim como a análise de marcas individuais, o modelo CIM apresentou apenas um QTL no primeiro corte que não foi mapeado no segundo corte, em que por outro lado houve a detecção de outros seis QTLs.

A comparação das regiões mapeadas entre diferentes estudos já realizados com cana-de-açúcar é difícil, pois nem sempre os grupos de ligação obtidos são os mesmos, por usarem diferentes marcadores e cruzamentos, e também porque a espécie não tem o número de cromossomos e ploidia conhecida (HEINZ; TEW, 1987). Neste contexto, o maior desafio para estudos de mapeamento de QTLs em cana-de-açúcar é a construção de mapas de ligação que permitam cobrir o genoma de forma integral, o que aumentaria a possibilidade de detectar QTLs. Contudo, o mapeamento aqui realizado mostrou resultados promissores, pois mesmo considerando que apenas parte do genoma foi amostrada foi possível detectar QTLs que explicavam conjuntamente 54% da variação fenotípica. Um fato que deve ser mencionado e que possivelmente explique tais resultados foi a utilização de marcadores funcionais do tipo EST (EST-RFLP e EST-SSR). Por considerarem apenas regiões expressas, eles podem ter aumentado a chance de sucesso no mapeamento.

Também vale destacar a utilização da abordagem multiponto para construção do mapa de ligação e cálculo das probabilidades *a priori* dos genótipos dos QTLs permitiu mapear QTLs com diferentes padrões de segregação, o que não é possível com a estratégia do *duplo pseudo testcross*. Neste caso, foram detectados oito QTLs com diferentes segregações (1:1:1:1, 1:2:1 e 3:1), o que deve contribuir para a obtenção de R^2 elevados.

Mapear QTLs em cana-de-açúcar é um grande desafio. Resultados ainda melhores po-

deriam possivelmente ser obtidos com a inclusão de mais marcadores em outras doses, pois além de contribuir para saturação do mapa também permitiriam estudar os caracteres quantitativos específicos de poliploides, sem a necessidade de fazer suposições irrealistas. Para tanto, faz-se necessário o desenvolvimento de estatísticos tanto para a construção do mapa de ligação como para o mapeamento de QTLs incluindo interação entre QTL e cortes e mapeamento simultâneo de vários caracteres. Isso poderia permitir um maior entendimento da base genética dos caracteres quantitativos em cana-de-açúcar. De qualquer forma, supõe-se que o modelo aqui apresentado contribua para estudos dessa espécie tão importante.

4.5 Conclusão

O modelo CIM aplicado em uma progênie de irmãos completos em cana-de-açúcar permitiu a detecção de mais QTLs do que as abordagens comumente utilizadas, inclusive com o mapeamento de QTLs com diferentes padrões de segregações.

Referências

- 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.; 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.
- 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.
- BUTCHER, A.; WILLIAMS, R.; WHITAKER, D.; LING, S.; SPEED, P.; MORAN F. Improving linkage analysis in outcrossed forest trees - an example from *Acacia mangium*, **Theoretical and Applied Genetics**, New York, v. 104, p. 1185-1191, 2002.
- CHARCOSSET, A.; MOREAU, L. Use of molecular markers for the development of new cultivars and the evaluation of genetic diversity, **Euphytica**, Wageningen, 137, p. 81-94, 2004.
- DOERGE, R.W.; ZENG, Z.B.; WEIR, B.S. Statistical issues in the search for genes affecting quantitative traits in experimental populations, **Statistical Science**, Hayward, v. 12, p. 195-219, 1997.

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

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.

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.

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

HOARAU, J.Y.; OFFMAN, 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.; OFFMAN, B.; RABOIN, L.M.; DIORFLAR, J.P.; PAYET, J.; HELLMAN, M.; D'HONT, A.; GLASZMANN, J.C. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTL's 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. 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.

KAO, C.H.; ZENG, Z.B. General formulae 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.

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

LANDER, E.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALEY, M.; LINCOLN, S.; 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.

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

LIU, B.H. **Statistical genomics**, p. linkage, mapping, and QTL analysis. 2 ed. Boca Raton: CRC Press, 1998. 611 p

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.

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, 2005.

MEER, J.M.; MANLY, K.F.; CUDMORE, R.H. **Software for genetic mapping of mendelian markers and quantitative traits loci**. Roswell Cancer Park Institute, 2002.

MING, R.; LIU, S.C.; MOORE, P.H.; IRVINE, J.E.; PATERSON, A.H. QTL Analysis in a complex autoploid: genetic control of sugar content in sugarcane. **Genome Research**, London, v. 11, p. 2075-2084, 2001.

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

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

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

MORGANTE, M.; SALAMINI, F. From plant genomics to breeding practice, **Current Opinion in Biotechnology**, Philadelphia, v. 14, p. 214-219, 2003.

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

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

PINTO, L.R.; OLIVEIRA, K.M.; MARCONI, T.G.; GARCIA, A.A.F.; ULIAN, E.C.; SOUZA, A.P. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs, **Plant Breeding**, Berlin, v. 125, p. 378-384, 2006.

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.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2009. Disponível em: <<http://www.R-project.org>> Acessado: 25 jan. 2009.

SCHWARZ, G. Estimating The Dimension of a model, **Annals of statistics**, Hayward, v. 6, p. 461-464, 1978.

SILLS, G.R.; BRIDGES, W.C.; 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.

TAKEDA, S.; MATSUOKA, M. Genetic approaches to crop improvement: responding to environmental and population changes, **Nature Reviews Genetics**, v. 9, p. 444-457, 2008.

WANG, S., BASTEN, C.J.; ZENG, Z.B. **Windows QTL Cartographer**: Version 2.5. Department of Statistics, North Carolina State University, North Carolina: Raleigh, 2007.

WU, R.L.; 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, 2002.

ZENG, Z.B. Theoretical basis of precision mapping of quantitative trait loci, **Proceedings of the National Academy of Sciences 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., KAO, C.H.; BASTEN, C.J. Estimating the genetic architecture of quantitative traits, **Genetical Research**, London, v. 74, p. 279-289, 1999.