



MARCIO BALESTRE

**MAPEAMENTO BAYESIANO DE MÚLTIPLOS
CARACTERES: EFEITOS PLEIOTRÓPICOS NO
ESTUDO DA HERANÇA DOS COMPONENTES
DA PRODUÇÃO EM MILHO**

**LAVRAS-MG
2012**

MARCIO BALESTRE

**MAPEAMENTO BAYESIANO DE MÚLTIPLOS CARACTERES:
A IMPORTÂNCIA DOS EFEITOS PLEIOTRÓPICOS NO ESTUDO
DOS COMPONENTES DA PRODUÇÃO EM MILHO**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Genética e Melhoramento de Plantas,
área de concentração em Genética e Melhoramento de
Plantas, para obtenção do título de Doutor.

Orientador

Dr. Renzo Garcia Von Pinho

Coorientador

Dr. Cláudio Lopes de Souza Júnior

LAVRAS-MG

2012

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Balestre, Marcio.

Mapeamento Bayesiano de múltiplos caracteres : efeitos pleiotrópicos no estudo da herança dos componentes da produção em milho / Marcio Balestre. – Lavras : UFLA, 2012.

89 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2012.

Orientador: Renzo Garcia Von Pinho.

Bibliografia.

1. Pleiotropia. 2. Caracteres complexos. 3. Análise Bayesiana. 4. Redes genéticas. 5. Arquitetura genética. I. Universidade Federal de Lavras. II. Título.

CDD – 631.523

MARCIO BALESTRE

**MAPEAMENTO BAYESIANO DE MÚLTIPLOS CARACTERES:
EFEITOS PLEIOTRÓPICOS NO ESTUDO DA HERANÇA DOS
COMPONENTES DA PRODUÇÃO EM MILHO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de Doutor.

APROVADA em 09 de Janeiro de 2012.

Dr. Júlio Sílvio de Sousa Bueno Filho	UFLA
Dr. João Cândido de Souza	UFLA
Dr. José Airton Rodrigues Nunes	UFLA
Dr. Lauro José Moreira Guimarães	EMBRAPA

Dr. Renzo Garcia Von Pinho
Orientador

LAVRAS-MG

2012

*A Deus, pela sua infinita engenhosidade.
A Laercio Balestre, meu pai, (in memoriam); a Maria Elena, minha mãe;
a Fernanda, minha esposa; a Rúbia e Vanessa, minhas irmãs;*

DEDICO

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Departamento de Genética e Melhoramento de Plantas pela oportunidade de realização deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado.

Ao professor Cláudio Lopes de Souza Júnior (ESALQ/USP) pela disponibilidade do conjunto de dados, pelo incentivo, paciência, confiança e amizade.

Ao professor Renzo Garcia Von Pinho pela orientação, incentivo, conselhos, liberdade de pesquisa e, acima de tudo, pela amizade.

Aos membros da banca avaliadora pelas valiosas sugestões, especialmente ao Dr. Lauro José Moreira Guimarães (EMBRAPA).

Aos Professores João Cândido de Souza, Júlio Silvío de Sousa Bueno Filho, José Airton Nunes, João Bosco dos Santos e César Brasil pelos ensinamentos transmitidos e pelo ótimo convívio durante esse período.

RESUMO

A pleiotropia tem desempenhado papel fundamental na dissecação de caracteres quantitativos. Porém a extensão desse efeito no genoma e suas consequências para o melhoramento de plantas ainda não foram completamente elucidadas. Esse trabalho foi realizado com o objetivo de identificar QTLs pleiotrópicos na cultura do milho utilizando o mapeamento Bayesiano de múltiplos intervalos. Além disso, procurou-se obter uma maior compreensão acerca da herança, extensão e distribuição dos efeitos pleiotrópicos em alguns componentes da produção do milho. Utilizou-se o delineamento III a partir de uma população proveniente do cruzamento entre as linhagens L-14-04B e L-08-05F. Foram genotipadas 250 plantas com 177 marcadores de microssatélites que posteriormente foram retrocruzadas com ambos parentais gerando 500 progenies de retrocruzamento que foram avaliadas em seis ambientes para produção de grãos e seus componentes. Os resultados deste trabalho sugerem que a análise de caracteres isolados restringe a compreensão acerca da arquitetura genética de caracteres quantitativos. Contudo, essa arquitetura pode ser melhor compreendida através de redes pleiotrópicas que permitam visualizar a complexidade da herança desses caracteres e, a partir disto, traçar novas estratégias de seleção. Também foi possível confrontar a ideia de que é possível identificar QTLs “para” caracteres complexos tais como a produção de grãos, visto que a pleiotropia atua de forma destacada em seus subcaracteres e que esse “caráter” pode ser decomposto e predito quase que na sua totalidade pelos QTLs dos seus componentes. Além disso, a pleiotropia de QTLs não necessariamente significa pleiotropia das interações alélicas e, sendo assim, a ampla distribuição dos efeitos pleiotrópicos não restringe a adaptabilidade de plantas.

Palavras-chave: Pleiotropia. Caracteres complexos. Análise Bayesiana. Redes genéticas. Arquitetura Genética.

ABSTRACT

Pleiotropy has played an important role in understanding quantitative traits. However, the extensiveness of this effect in the genome and its consequences for plant improvement have not been fully elucidated. The aim of this study was to identify pleiotropic quantitative trait loci (QTL) in maize using Bayesian multiple interval mapping. Additionally, we sought to obtain a better understanding of the inheritance, extent and distribution of pleiotropic effects of several components in maize production. The design III procedure was used from a population derived from the cross of the inbred lines L-14-04B and L-08-05F. Two-hundred and fifty plants were genotyped with 177 microsatellite markers and backcrossed to both parents giving rise to 500 backcrossed progenies, which were evaluated in six environments for grain yield and its components. The results of this study suggest that mapping isolated traits limits our understanding of the genetic architecture of quantitative traits. This architecture can be better understood by using pleiotropic networks that facilitate the visualization of the complexity of quantitative inheritance, and this characterization will help to develop new selection strategies. It was also possible to confront the idea that it is feasible to identify QTLs for complex traits such as grain yield, as pleiotropy acts prominently on its subtraits and as this “trait” can be broken down and predicted almost completely by the QTLs of its components. Additionally, pleiotropic QTLs do not necessarily signify pleiotropy of allelic interactions, and this indicates that the pervasive pleiotropy does not limit the genetic adaptability of plants.

Keywords: Pleiotropy, Complex traits, Bayesian analysis, Genetic Networks, Genetic Architecture.

SUMÁRIO

PRIMEIRA PARTE		
1	INTRODUÇÃO	10
2	REFERENCIAL TEÓRICO	12
2.1	Utilização do Delineamento III na identificação de QTL	12
2.2	Abordagem bayesiana de múltiplos intervalos	17
2.3	Pleiotropia e suas implicações na arquitetura genética	25
2.4	Componentes da produção e heterose multiplicativa	34
2.5	Análise de QTLs sob múltiplos caracteres	37
3	CONCLUSÕES	42
	REFERÊNCIAS	43
	SEGUNDA PARTE – ARTIGO	49
	ARTIGO 1 Mapeamento bayesiano de múltiplos caracteres em milho: A importância dos efeitos pleiotrópicos no estudo da herança de caracteres quantitativos	50

PRIMEIRA PARTE

1 INTRODUÇÃO

O estudo detalhado da arquitetura genética relacionada a caracteres quantitativos e/ou caracteres complexos tornou-se um grande desafio para geneticistas de diversas áreas. A identificação e localização de genes associados a esses caracteres - comumente denominados pela sigla QTL (*Quantitative trait loci*) - tornou-se possível graças ao advento dos marcadores moleculares e procedimentos estatísticos avançados que proporcionaram a detecção de QTL em diversos arranjos de cruzamentos e também a sua localização em mapas genéticos saturados.

No estudo do controle genético de caracteres quantitativos alguns pontos são de extrema importância e devem ser levados em consideração: i) são caracteres controlados por um grande número de locos; ii) esses locos podem atuar de forma aditiva, dominante, epistática e podem interagir com fatores ambientais; iii) a magnitude dos efeitos de cada loco pode variar consideravelmente; iv) o mesmo gene pode afetar diferentes caracteres apresentando efeito pleiotrópico; v) a distribuição desses genes pode ser aleatória no genoma ou concentrada em blocos. O item iv (pleiotropia) tem grande importância tanto na genética evolutiva quanto no melhoramento de plantas.

A importância da pleiotropia no melhoramento de plantas advém das correlações genéticas entre caracteres de importância agrônômica e econômica, que também são observadas no desequilíbrio de ligação de locos localizados em regiões próximas. No entanto, as correlações obtidas através do desequilíbrio de ligação são transitórias e a velocidade com que o desequilíbrio é dissipado depende principalmente da distância entre os genes. Por outro lado, sendo a pleiotropia um fenômeno onde o mesmo loco controla diferentes caracteres, seu

efeito é mais estável, embora não haja evidências de que seja permanente. Mesmo sobre essa última circunstância e sob o ponto de vista de seleção de múltiplos caracteres e estratégias de melhoramento, torna-se relevante distinguir os efeitos de ligação e de pleiotropia atuando no controle genético dos caracteres, embora outros fenômenos, como a distribuição dos efeitos de pleiotropia ao longo do genoma, devam ser estudados.

A abordagem bayesiana de QTL envolvendo múltiplos intervalos e múltiplos caracteres tornou-se uma ferramenta poderosa em estudos de análise genômica. Por essa razão essa abordagem será aplicada nesse trabalho com o objetivo de obter uma maior compreensão acerca da importância da pleiotropia no estudo da herança do caráter produção de grãos e alguns dos seus componentes na cultura do milho.

2 REFERENCIAL TEÓRICO

2.1 Utilização do Delineamento III na identificação de QTL

O delineamento III da (DIII) proposto por Comstock e Robinson (1952) foi idealizado - a princípio - para estimar componentes da variância e também o grau médio de dominância em estudos de caracteres quantitativos. Esse delineamento consiste da escolha aleatória de plantas da geração F_2 e posterior retrocruzamento com cada linhagem parental, obtendo-se então duas populações retrocruzadas. Os dois retrocruzamentos conjuntamente constituem uma população com frequências alélicas e genotípicas semelhante a uma população F_2 , porém estruturada em progênies de meios irmãos. Sob essa estrutura é possível estimar a variância genética aditiva - pela esperança da variância genética entre progênies - e estimar a variância de dominância que pode ser obtida pela esperança da variância da interação entre progênies e seus parentais.

Para a interpretação dos componentes da variância genética em termos de variância aditiva e de dominância, algumas restrições tais como a ausência de desequilíbrio de ligação e epistasia são adotadas (COMSTOCK; ROBINSON, 1952). Essas pressuposições são difíceis de serem atendidas, já que geralmente este delineamento recorre ao retrocruzamento entre plantas da geração F_2 , onde o desequilíbrio de ligação é máximo e o real valor de epistasia é desconhecido. O desequilíbrio de ligação pode gerar viés na estimativa das variâncias aditivas e de dominância devido à covariância presente nesses componentes que é função do desequilíbrio de ligação. Para a variância aditiva, as estimativas ficam superestimadas quando os genes estão em fase de associação e subestimadas quando em fase de repulsão. Em relação à estimativa da variância de dominância, essa sempre fica superestimada independente da fase de ligação.

Dada essas restrições, Cockerham e Zeng (1996) sugeriram uma nova abordagem do delineamento III por meio marcadores moleculares aplicando o

método de marcas simples, ou valor fenotípico associado à marca. Esses autores sugeriram quatro contrastes ortogonais com objetivo de estimar o grau médio de dominância e parte dos efeitos epistáticos presentes no controle de um caráter de interesse. Porém, esse método possui algumas desvantagens que podem limitar sua aplicação. A primeira desvantagem se refere ao viés embutido na estimativa da média do QTL devido a sua frequência de recombinação com a marca associada. Em outras palavras, a menos que o genoma esteja extremamente saturado por marcadores, onde assintoticamente espera-se que as marcas estejam tão próximas ou até mesmo dentro do QTL de modo a segregarem conjuntamente, o método de Cockerham e Zeng (1996) é claramente viciado. Além disso, como apontado por Garcia et al. (2008), se dois ou mais QTLs estão ligados à mesma marca, não é possível separar os efeitos aditivos e de dominância adequadamente, além do método se limitar à identificação de apenas parte dos efeitos epistáticos.

Em alguns estudos de mapeamento de QTLs utilizando populações provenientes do DIII tem-se sugerido que os retrocruzamentos (RC_1 e RC_2) sejam analisados separadamente, onde os efeitos aditivos e de dominância são obtidos por contrastes simples entre as médias dos QTLs (LU; ROMERO-SEVERSON; BERNARDO, 2003; STUBER et al., 1992). Essa metodologia permite a utilização do mapeamento por intervalo e elimina o viés contido no método de Cockerham e Zeng (1996). Contudo, essa abordagem possui baixa capacidade de identificação de QTL devido à alta interação QTL por retrocruzamento (QTLxRC). A interação QTLxRC pode acontecer quando um QTL exibe dominância completa ou com valores próximos a 1, impossibilitando sua identificação no RC_1 (AA-Aa), sendo então possivelmente identificados no RC_2 (Aa-aa). Assim, a análise conjunta dos dois retrocruzamentos (RC_1 e RC_2) pode aumentar o poder de identificação do QTL, como já destacado por Garcia et al. (2008).

Estudos recentes na área de genética quantitativa têm contribuído enormemente para a identificação de QTLs utilizando tanto o DIII como também diversos outros delineamentos experimentais (MELCHINGER et al., 2007). Por exemplo, Yang (2004) e Zeng, Wang e Zou (2005) levantaram diversas questões acerca da correta modelagem e interpretação dos efeitos de QTLs em diversas configurações de cruzamentos e populações. Esses autores destacam que os efeitos genéticos de um QTL precisam ser expressos de acordo com uma população de referência porque não dependem somente dos valores genotípicos, mas também das frequências alélicas.

Diversos estudos visando à identificação de QTLs têm utilizado a modelagem sugerida por Venn (1959) denominada métrica F_{∞} . Nessa abordagem, os valores genotípicos normalmente são expressos em função do cruzamento entre dois genitores cuja média representa a esperança de uma população endogâmica em F_{∞} , ou seja, não faz qualquer referência à real população sendo mapeada. Para exemplificar, considere um loco com dois alelos A_i ($i=1,2$), sendo o alelo A_1 favorável para um caráter de interesse. Assumindo a seguinte notação: $G_{11}=A_1A_1$; $G_{21}=A_1A_2$ e $G_{22}=A_2A_2$, a esperança dos valores genotípicos sobre a métrica F_{∞} pode ser dada como: $G_{22}=+a$; $G_{21}=d$ e $G_{11}=-a$. Sobre esse modelo, o efeito aditivo (a) é dado pela metade da diferença dos dois indivíduos homozigotos $a = (G_{22} - G_{11})/2$ e o efeito de dominância é dado pela diferença do valor do heterozigoto em relação à média dos genótipos homozigotos $d = G_{21} - (G_{22} + G_{11})/2$. Assim, aplicando a métrica F_{∞} , os efeitos genéticos são definidos apenas em função das médias, contrariando o modelo de Fisher onde os efeitos genéticos são estimados dentro de uma população de referência, em equilíbrio de Hardy-Weinberg e com as frequências alélicas conhecidas.

O efeito de uma substituição alélica ou efeito médio de um alelo, no modelo de Fisher, depende tanto dos efeitos de a e d como também das frequências alélicas da população. Aplicando esse conceito, Cockerham (1954) propôs uma abordagem geral de análise, constituída por um conjunto de contrastes ortogonais que possibilitam a partição do valor genotípico em efeitos aditivos, epistáticos e de dominância. Considerando uma população com frequências alélicas de $\frac{1}{2}$ e em equilíbrio de Hardy-Weinberg, o modelo adaptado de Cockerham é denominado métrica F_2 , e possui características ortogonais que não são observadas na métrica F_∞ quando se ajustam modelos epistáticos (KAO; ZENG, 2002).

Como demonstrado por Yang (2004), sob a métrica F_2 , os valores genotípicos podem ser dados por: $G_{11} = a - \frac{1}{2}d$; $G_{21} = \frac{1}{2}d$ e $G_{22} = -a - \frac{1}{2}d$.

Quando não se ajustam os efeitos epistáticos ao modelo, as estimativas dos efeitos aditivos e de dominância podem ser obtidas através dos contrastes $(G_{22} - G_{11})/2$ e $[(G_{21} - G_{11}) + (G_{21} - G_{22})]/2$ respectivamente.

Resumindo, constata-se que na ausência de um modelo epistático, as métricas F_2 e F_∞ se equivalem (ZENG; WANG; ZOU, 2005). No delineamento III, a distinção entre as métricas torna-se importante somente no ajuste de modelos epistáticos ou quando se deseja estudar a contribuição desses efeitos no controle genético da heterose (MELCHINGER et al., 2007; ZENG; WANG; ZOU, 2005). Esses autores demonstraram que, quando se ajusta um modelo epistático, existe uma grande diferença entre a adoção das métricas F_2 e F_∞ na interpretação dos resultados em termos de efeitos de QTL. Por exemplo, sob a métrica F_∞ , a heterose é função dos efeitos de dominância e das epistasias do tipo dominante x dominante e aditiva x aditiva ($h^* = [d_1 + d_2 + d_1d_2 - aa_{12}]$). Por outro lado, na

métrica F_2 , a heterose seria função apenas dos efeitos de dominância e da epistasia do tipo aditiva x aditiva ($h^* = [d_1 + d_2 - aa_{12}]$).

Segundo Melchinger et al. (2007), esse último modelo torna a interpretação da heterose mais adequada e simples e, além disso, põe em dúvida o argumento de que, ao longo dos anos, devido ao aumento na média das linhagens em detrimento do ganho em heterose, o principal efeito sobre esse fenômeno seria a complementação alélica expressa em termo de dominância completa/parcial (TROYER, 2006). Tomando a expressão dos efeitos ampliados de dominância derivados por Melchinger et al. (2007) para o DIII; $d^* = (d_1 + d_2 - aa_{12})$, e que também são equivalentes a esperança da heterose (ZENG; WANG; ZOU, 2005), observa-se que, à medida que ocorre fixação de locos favoráveis nos parentais, a heterose diminui e mascara o real efeito de dominância no controle do caráter. Assim, mesmo sob a hipótese de sobredominância é possível obter baixa heterose e linhagens com média alta, devido à fixação de locos favoráveis durante o processo de obtenção de linhagens endogâmicas. Da mesma forma, é possível obter heterose mesmo em locos que não apresentam dominância. Para isso, basta que o somatório da epistasia do tipo aditiva-aditiva presente em cada um dos parentais seja negativa; o que corresponde à complementação de locos puramente aditivos. Devido a esses fatores, não é prudente inferir sobre o controle genético da heterose com base na relação da média dos parentais e o vigor híbrido dos seus descendentes.

Estudos utilizando o delineamento III apresentam resultados pouco concordantes em relação ao controle genético da heterose para peso de grãos em milho. Em trabalhos anteriores, Lu, Romero-Severson e Bernardo (2003) observaram que mais de 80% dos QTL identificados apresentaram efeito de sobredominância e sugeriram duas explicações para esses resultados: 1) os locos apresentam pseudo-sobredominância e três gerações de acasalamento em uma

população F_2 não foram suficientes para dissipá-la ou 2) os QTLs para produção de grãos realmente são controlados por locos com sobredominância.

Em outro trabalho utilizando diversos delineamentos experimentais, incluindo o DIII, Frascaroli et al. (2007) observaram que a maior parte dos QTLs identificados apresentaram grau de dominância >1 . De forma semelhante, Garcia et al. (2008) verificaram que a dominância pode ser o principal agente da heterose, com a epistasia desempenhando importância marginal no controle desse caráter. Além disso, esses autores destacam que efeito de sobredominância foi observado em vários locos e, sendo assim, a sua importância no controle genético da heterose não pode ser ignorada. Resultados semelhantes foram obtidos por Shön et al. (2010), porém esses autores não observaram nenhum efeito significativo de epistasia. Em nenhum desses trabalhos o efeito de epistasia apresentou grande contribuição e a partir desses resultados duas hipóteses ficam em aberto: 1) o ajuste dos efeitos epistáticos apenas em QTLs de grande efeito pode não desenhar a real contribuição desse fenômeno no controle genético da heterose, 2) a epistasia realmente exerce efeito marginal e sendo assim sua exclusão na modelagem de QTLs pode ser justificada. Se a hipótese 2 for verdadeira, a utilização da métrica F_∞ no lugar da métrica F_2 pode ser aplicada sem prejuízos à análise.

Com isso, é possível observar que ainda existe uma grande incerteza acerca da arquitetura genética da heterose e que são necessários novos estudos e novas abordagens que contribuam na elucidação desse fenômeno tão importante em diversas culturas, principalmente na cultura do milho.

2.2 Abordagem bayesiana de múltiplos intervalos

Quando se deseja realizar um estudo utilizando a abordagem clássica da genética quantitativa, normalmente assume-se um modelo infinitesimal, em que

inúmeros locos de pequeno efeito controlam o caráter sob investigação. Nesse modelo, supõe-se que a variância individual dos QTLs é tão pequena que não pode ser investigada separadamente, mas antes deve ser avaliada em conjunto por meio de uma estrutura populacional conhecida ou parentesco entre indivíduos (LYNCH; WASH, 1998). Embora a arquitetura poligênica tenha sido observada em estudos recentes, o modelo infinitesimal que fundamenta essa hipótese tem sido questionado uma vez que a distribuição dos efeitos de QTLs tende a aproximar uma função Gamma em forma de L (*L-shaped*), ou seja, milhares de regiões sem efeito algum e dezenas ou centenas de regiões com pequeno efeito (FLINT; MACKAY, 2009).

Recentemente, algumas técnicas de mapeamento têm permitido a identificação conjunta de QTLs no genoma, melhorando a percepção da arquitetura genética do caráter. Quando o conjunto de marcadores que cobrem os grupos de ligação de uma espécie é analisado simultaneamente em apenas um modelo, define-se então um método de análise genômica ampla (MEUWISSEN; HAYES; GODDARD, 2001). Xu (2003) estendeu a análise genômica ampla sob a ótica bayesiana com objetivo de identificar marcadores ligados a QTLs. Para isso, assumiu variância individual para cada marca e utilizou as propriedades de “encolhimento” inerente ao método para obter uma estimativa do número de QTLs controlando o caráter. Essa abordagem é o fundamento desse trabalho e será descrita com detalhes em seguida.

A detecção e localização de QTLs são possíveis somente quando se utiliza marcadores moleculares densamente distribuídos ao longo do genoma. O genótipo do QTL não é observável, porém pode ser predito com base nas marcas que o flanqueiam, ou seja, essas marcas definem um intervalo que pode ou não conter um QTL. Baseado nisso, o método de mapeamento por intervalo tem sido amplamente utilizado (LANDER; BOTSTEIN, 1989). Nesse método, cada posição é testada separadamente, ou seja, para cada QTL, de forma individual,

são obtidos valores de LOD que posteriormente são plotados em um gráfico descrevendo a provável distribuição do QTL ao longo do genoma. Como cada QTL é ajustado individualmente no modelo, os outros possíveis QTLs presentes no genoma são ignorados, aumentando dessa forma o resíduo e superestimando seu efeito e/ou sua porcentagem de explicação da variância genotípica e fenotípica. Além disso, QTLs localizados próximos a um intervalo específico podem levar à identificação de QTLs “fantasmas” devido à ligação do real QTL, presente em intervalos adjacentes, com o intervalo sendo testado (WU; MA; CASELLA, 2007). Uma maneira de aliviar esse problema seria a inclusão de covariáveis ou cofatores na análise. Essa abordagem é denominada mapeamento por intervalo composto (CIM) que utiliza testes estatísticos através da combinação do mapeamento por intervalo sobre duas marcas e regressão múltipla sobre os demais marcadores. Essa estratégia, que foi proposta independentemente por Jansen (1992, 1993) e Zeng (1993, 1994), pode remover apenas parte da interferência do QTL fora do intervalo sendo testado se o genoma for pobremente saturado e, além disso, se inúmeros QTLs são identificados torna-se necessário reescrever o modelo e reestimar os seus efeitos com as posições fixas nas suas estimativas (WANG et al., 2005). Assim, podemos inferir que o ideal seria eliminar os efeitos dos QTLs à jusante ou à montante do intervalo utilizando a análise de múltiplos intervalos.

Mapeamento por múltiplos intervalos tornou-se o “estado da arte” dentro dos procedimentos de busca por QTLs. Uma vez que um caráter complexo pode ser controlado por inúmeros QTLs, é crucial realizar um mapeamento onde se possam modelar vários QTLs simultaneamente e, assim, identificar e localizar todos os QTLs responsáveis pelo caráter em questão. Algumas metodologias que incluem a análise de múltiplos intervalos (MIM) têm sido propostas (KAO; ZENG; TEASDALE, 1999; ZENG et al., 2000). No entanto, sua implementação é onerosa, principalmente quando se tem que lidar

com a seleção de modelos ocorrendo simultaneamente à identificação de QTLs (XU, 2003). Segundo Xu (2003), esse fato caracteriza a principal desvantagem do método, sendo que em alguns casos o critério de adição e deleção de QTLs durante o processo de seleção de modelos pode ser arbitrário. Além disso, segundo esse autor, o espaço amostral de possíveis modelos pode ser tão grande que dificilmente é explorado por completo. Também, a seleção de modelos pode causar um viés na estimativa dos efeitos genéticos se o modelo selecionado é de apenas um único QTL.

O mapeamento por múltiplos intervalos original, proposto por Kao, Zeng e Teasdale (1999), foi denominado por Xu (2003) como *mapeamento seletivo de múltiplos intervalos*, ou seja, na realidade esse método não tem como lidar com múltiplos intervalos e por esse fato realiza a seleção de modelos. Segundo Xu (2003), o verdadeiro mapeamento por intervalos múltiplos seria aquele onde todos os intervalos - que são definidos pelos marcadores distribuídos ao longo do genoma - são explorados simultaneamente.

Recentemente, a análise bayesiana de múltiplos intervalos via Monte Carlo Cadeias de Markov (MCCM) tem sido aplicada (SATAGOPAN et al., 1996; WANG et al., 2005; XU et al., 2009). A abordagem bayesiana permite grande flexibilidade de análise em modelos complexos e também em situações onde o número de parâmetros a ser estimado (ex. número de marcadores e intervalos) é maior que o número de observações (XU, 2003).

Na abordagem bayesiana, o número de QTLs no modelo pode ser definido pela aplicação direta do fator de Bayes (SATAGOPAN et al., 1996) ou utilizando a técnica de MCCM com saltos reversíveis (GREEN, 1995; SILLANPÄÄ; ARJAS, 1999). Saltos reversíveis ou “*reversible jump*” (RJ) realiza simultaneamente a estimação de parâmetros (efeito e posição do QTL) e seleção de modelos (relacionado ao número de QTLs). O número de QTLs controlando o caráter é tido como uma variável aleatória, sendo assim, a cada

iteração um número de QTL é sugerido e a cadeia de Markov salta dentro desses modelos e seleciona aquele com maior probabilidade. Assim, a quantidade de vezes que o modelo é selecionado caracteriza o número de QTLs no controle do caráter. Normalmente tem sido observada uma baixa taxa de convergência na aplicação deste método, exigindo assim cadeias muito longas além de intensa demanda computacional e ainda sim resulta em amostragem deficiente.

Recentemente, Yi, George e Allison (2003) utilizaram o método denominado modelo de espaço composto também conhecido como “*Stochastic Search Variable Selection*” (SSVS). Esse algoritmo introduz variáveis binárias indicadoras do estado de inclusão ou exclusão de um loco durante a análise. Nesse método, o efeito do QTL é condicionado a uma mistura de duas distribuições normais. Uma distribuição tem média igual a zero e uma pequena variância, enquanto a outra distribuição possui também média zero porém uma alta variância. Se existe uma alta probabilidade a posteriori de que um efeito do modelo possua grande variância, esse efeito é selecionado.

Uma alternativa às técnicas de seleção de modelos seria não realizar seleção de modelo, ou seja, realizar o mapeamento de múltiplos intervalos assumindo que os intervalos são fixos e condicionados ao número de marcadores utilizados. Além disso, pode-se assumir, a priori, que cada QTL é uma variável aleatória proveniente de uma distribuição normal com média zero e variância individual. Sob essa suposição, cada possível QTL é penalizado pela razão da sua variância com a variância residual, ou, por assim dizer, pela “herdabilidade do QTL”. Sob esse enfoque, aqueles QTLs de pequeno efeito e baixa variância ou baixa herdabilidade, têm esses efeitos “encolhidos” a valores próximos a zero, pois são penalizados pela variância residual do modelo. Por outro lado, aqueles QTLs de grande efeito tendem a apresentar destacada variância e são menos penalizados pela variância residual.

Essa abordagem tem sido denominada como “*Bayesian Shrinkage Analysis of Multiple Interval*” (BSAMI) ou análise bayesiana de múltiplos intervalos com efeitos penalizados pelo resíduo. O protótipo dessa ideia foi o trabalho desenvolvido por Xu (2003) que aplicou a abordagem bayesiana sobre inúmeros marcadores com o objetivo de visualizar o padrão de distribuição de QTLs ao longo do genoma adotando um modelo de “encolhimento” em marcas individuais. Nesse trabalho, cada marcador foi assumido como um provável QTL, admitindo dessa forma que a matriz de incidência fosse integralmente observável. Além disso, os efeitos desses supostos QTLs foram inseridos conjuntamente no modelo, não havendo inclusão ou exclusão de marcadores ao longo do processo de MCCM. O método basicamente consiste do ajuste de um modelo aleatório de regressão dado por:

$$y_i = b_0 + \sum_j^m x_{ij} a_j + \sum_j^m w_{ij} d_j + e_{ij}$$

onde y_i é o fenótipo do indivíduo i , b_0 é a média geral da população sob estudo, m é o número total de marcas no genoma, x_{ik} e w_{ij} são variáveis representativas que descrevem o genótipo aditivo e dominante do marcador j respectivamente, a_j e d_j são os efeitos aditivo e dominante do QTL associados com o marcador j e e_{ij} é o resíduo assumido $N(0, \sigma_e^2)$.

Nesse modelo é assumido que a_j e d_j são tomados a partir de uma distribuição normal com média zero e variância σ_{aj}^2 e σ_{dj}^2 . As variáveis observáveis são os dados fenotípicos e os genótipos dos marcadores, enquanto que as variáveis não observáveis são os efeitos aditivos, dominantes e suas variâncias. Como priori são assumidos que:

$$p(b_0) \propto 1, p(a_j) \propto N(0, \sigma_{aj}^2), p(d_j) \propto N(0, \sigma_{dj}^2), \\ p(\sigma_{aj}^2) \propto 1/\sigma_{aj}^2, p(\sigma_{dj}^2) \propto 1/\sigma_{dj}^2, p(\sigma_e^2) \propto 1/\sigma_e^2$$

As distribuições a priori dos efeitos dos QTLs (a_j e d_j), da média geral (b_0) e das variâncias ($\sigma_{aj}^2, \sigma_{dj}^2$ e σ_e^2) serão assumidos como distribuições \mathbf{b} e \mathbf{v} respectivamente, apenas por facilidade de notação. Essas distribuições podem ser convenientemente descritas dentro de uma função de probabilidade conjunta $p(\mathbf{b}, \mathbf{v})$. A verossimilhança das variáveis observáveis e não-observáveis é dada por:

$$p(y | b, v) = \prod_{i=1}^n p(y_i | b, \sigma_e^2) \propto (\sigma_e^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^m x_{ij} a_j - \sum_{j=1}^m w_{ij} d_j \right)^2 \right\}$$

As marcas perdidas podem ser estimadas a cada iteração a partir da sua recombinação com as marcas adjacentes. Para isso, uma distribuição Bernoulli com probabilidades modeladas em termos de frequência de recombinação entre a marca perdida as marcas flaqueadoras podem ser usadas para atualizar os dados perdidos individualmente (SATAGOPAN et al., 1996).

Baseado nessas suposições e assumindo que as distribuições dos parâmetros são conhecidas, Xu (2003) implementou o algoritmo MCCM utilizando a amostragem de Gibbs nas distribuições condicionais posteriores de todos os parâmetros do modelo.

A distribuição posteriori condicional da média geral foi assumida como normal, com média $b_0 = (1/n) \sum_{i=1}^n \left(y_i - \sum_{j=1}^m x_{ij} a_j - \sum_{j=1}^m w_{ij} d_j \right)$ e variância $(1/n)\sigma_e^2$

A distribuição posteriori condicional dos efeitos aditivos e dominantes foram assumidas como normal com médias dadas respectivamente por:

$$a_j = \left(\sum_{i=1}^n x_{ij}^2 + \sigma_e^2 / \sigma_{aj}^2 \right)^{-1} \sum_{i=1}^n x_{ij}^2 \left(y_i - b_0 - \sum_{k \neq j}^m x_{ik} a_k - \sum_{j=1}^m w_{ij} d_j \right)$$

$$d_j = \left(\sum_{i=1}^n w_{ij}^2 + \sigma_e^2 / \sigma_{dj}^2 \right)^{-1} \sum_{i=1}^n w_{ij}^2 \left(y_i - b_0 - \sum_{i=j}^m x_{ij} a_j - \sum_{k \neq j}^m w_{ij} d_j \right)$$

e variâncias $\left(\sum_{i=1}^n x_{ij}^2 + \sigma_e^2 / \sigma_{aj}^2 \right)^{-1} \sigma_e^2$ e $\left(\sum_{i=1}^n w_{ij}^2 + \sigma_e^2 / \sigma_{dj}^2 \right)^{-1} \sigma_e^2$,

respectivamente.

As variâncias aditivas, dominantes e residuais foram amostradas de distribuições qui-quadrado escalonada invertida, ou seja:

$\sigma_{aj}^2 = a_j^2 / \chi_1^2$, $\sigma_{dj}^2 = d_j^2 / \chi_1^2$ onde χ_1^2 é um número aleatório amostrado de uma distribuição qui-quadrado com 1 grau de liberdade e

$$\sigma_e^2 = \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^m x_{ij} a_j - \sum_{j=1}^m w_{ij} d_j \right)^2 / \chi_n^2 \text{ onde } \chi_n^2 \text{ é um número aleatório}$$

amostrado de uma distribuição qui-quadrado com n graus de liberdade.

A metodologia proposta por Xu (2003) é bem simples e provou ser mais eficiente no estudo da arquitetura genética do que o método de análise de marcas individuais e também superior ao método de “*ridge regression*” que é muito utilizado no mapeamento tradicional. No entanto, a abordagem apresentada por Xu (2003) não pode ser tomada como uma metodologia direta de mapeamento de QTL, pois o genótipo do QTL não é observado e, dessa forma, a matriz de incidência dos efeitos do QTLs torna-se incerta e por consequência uma variável aleatória. Além disso, a posição do QTL no genoma é um parâmetro adicional a ser estimado e requer técnicas estatísticas que estão além da aplicação de modelos de “encolhimento”.

Utilizando a ideia de Xu (2003), Wang et al. (2005) apresentaram a técnica de mapeamento de múltiplos intervalos utilizando a análise bayesiana penalizada. O modelo aplicado é semelhante ao apresentado por Xu, no entanto, novos parâmetros são estimados, a saber: o genótipo do QTL (x_j e w_j) - antes

assumido como genótipo dos marcadores - e também sua posição (λ_j). Wang et al. (2005) assumiram que cada intervalo no genoma contém um QTL em potencial, e sua posição pode variar aleatoriamente dentro desse intervalo. Na realidade muitos intervalos não contém QTL algum, ou seja, são intervalos nulos.

A aplicação da análise bayesiana “penalizada” força os intervalos com pequeno efeito a valores próximos de zero, fazendo com que sua inclusão tenha pouco efeito nos demais intervalos. Nessa metodologia, os dados observados são as marcas (m) e o fenótipo (y). As variáveis não-observáveis são os genótipos dos QTLs (x_j e w_j), seus efeitos (a_j e d_j), suas variâncias junto com a variância residual ($\sigma_{aj}^2, \sigma_{dj}^2$ e σ_e^2) e também suas posições (λ_j). As distribuições a priori para os efeitos dos QTLs e suas variâncias são as mesmas utilizadas anteriormente no modelo de Xu. A priori relativa à posição do QTL pode ser uniforme, ou seja, assumindo que M_{j-1} e M_{j+1} são duas marcas que flanqueiam o genótipo do QTL Q_j , e que L_j e U_j sejam as distâncias entre $M_{j-1} \leftrightarrow Q_j$ e $Q_j \leftrightarrow M_{j+1}$ respectivamente, a priori uniforme para cada intervalo é dada por um conjunto de números ordenados de mesma probabilidade e que variam de $[L_j$ até $U_j]$.

A nova distribuição a priori conjunta pode ser assumida como:

$$p(\lambda, b, v) = p(b_0) p(\sigma_e^2) \prod_{i=1}^{m-1} p(\lambda_i) p(a_i) p(d_i) p(\sigma_{ai}^2) p(\sigma_{di}^2)$$

A verossimilhança de b , v e λ , pode então ser dada como:

$$p(\lambda, b, v | y, m) = \prod_{i=1}^n p(y | b, x, w, \sigma_e^2) p(x, w | \lambda, m)$$

Assim, a posteriori conjunta pode ser tomada por:

$$p(\lambda, b, v | y, m) \propto p(y | b, x, w, \sigma_e^2) p(x, w | \lambda, m) p(\lambda, b, v)$$

onde $p(x, w | \lambda, m)$ é a probabilidade do genótipo do QTL dada sua posição no genoma e as marcas que o flanqueiam.

Assumindo que uma população F_2 pode ser definida em termos de G_{22} , G_{21} e G_{11} , codificados como (1, 0 e -1), cada genótipo é amostrado diretamente de uma distribuição Bernoulli com probabilidade dada por;

$$p(x_{ij}, w_{ij} | \lambda_j, m_{ij-1}, m_{ij+1}, r_{1j}, r_{2j}) = \frac{p(x_{ij}, w_{ij} | r_{1j}, r_{2j}) p(y_i | b, x_{ij}, w_{ij}, \sigma_e^2)}{\sum_{k=1}^3 p(x_{ij}, w_{ij} | r_{1j}, r_{2j}) p(y_i | b, x_{ij}, w_{ij}, \sigma_e^2)}$$

Onde r_{1j}, r_{2j} é frequência de recombinação entre o QTL j com as marcas m_{ij-1} , e m_{ij+1} respectivamente.

O parâmetro λ não possui uma função conhecida, sendo assim, a utilização do amostrador de Gibbs é proibitiva. Isso porque o amostrador de Gibbs faz uso de uma função conhecida retirando amostras de uma cadeia de Markov utilizando um processo iterativo. Quando essa função não é conhecida (não necessariamente a distribuição posteriori), outro método Monte Carlo via Cadeias de Markov (MCCM), denominado Metropolis-Hastings (HASTINGS, 1970; METROPOLIS et al., 1953) pode ser utilizado. Esse algoritmo não exige que o parâmetro tenha uma função de probabilidade conhecida mas, ao invés disso, faz uso de uma função auxiliar possível de ser amostrada, retirando valores candidatos que podem ser aceitos com um α de probabilidade. No caso de mapeamento por intervalo, pode-se utilizar uma distribuição uniforme como função auxiliar que é amostrada sobre cada intervalo delimitado por $\max(\lambda_{j-1}, \lambda_j - d)$ e $\min(\lambda_{j+1}, \lambda_j + d)$ onde d é uma constante que define o caminhamento dentro do intervalo j , normalmente fixada entre 1 e 2 cM. Essa

função é denotada por $u(\lambda^*, \lambda)$, e a nova posição será aceita na k-ésima iteração com $\min(1, \alpha)$ de probabilidade sendo α dado por:

$$\alpha = \frac{p(\lambda_j^* | y, b, x, w, v)u(\lambda_j^*, \lambda_j)}{p(\lambda_j^0 | y, b, x, w, v)u(\lambda_j, \lambda_j^*)}$$

Assim, se α for aceito, uma nova posição é estabelecida e um novo genótipo é sugerido para as matrizes x e w fechando um ciclo de MCCM (BANERJEE; YANDELL; YI, 2008; SATAGOPAN et al., 1996; WANG et al., 2005). Uma das grandes desvantagens da análise bayesiana de múltiplos intervalos é que todos os possíveis QTL são mantidos no modelo violando, por assim dizer, a ideia de modelos parcimoniosos, convergindo para uma alta demanda computacional. Essa maior demanda computacional pode ser compensada pela menor exigência do tamanho da cadeia, pois como os intervalos são curtos, cadeias reportadas entre 20.000 a 50.000 têm se mostrado suficientes para convergência (GAO et al., 2009; WANG et al., 2005; XU et al., 2009).

Apesar de aparentemente violar o princípio da parcimônia de modelos, esse método se apresenta como o mais representativo da possível arquitetura de um caráter oligogênico/poligênico, onde genes principais apresentam de moderada/alta influência sobre o caráter, e quando combinados com vários outros de genes de pequeno efeito determinam a variação genética de um caráter quantitativo (FLINT; MACKAY, 2009; MACKAY, 2001). Assim, é importante manter esses QTLs de pequeno efeito no modelo, pois coletivamente podem realizar um importante papel na variância genética do caráter (WANG et al., 2005).

Além disso, em comparação com os outros métodos de análise por múltiplos intervalos com seleção de modelos, incluindo o método SSVS, a

abordagem bayesiana de múltiplos intervalos com efeitos penalizados tem se demonstrado superior tanto na estimativa dos efeitos, quanto na estimativa da real posição do QTL (WANG et al., 2005). No entanto, ainda poucos estudos de mapeamento de QTL são realizados utilizando o método, principalmente devido à alta demanda computacional.

2.3 Pleiotropia e suas implicações na arquitetura genética

O termo pleiotropia - que recentemente completou 100 anos - foi cunhado pelo geneticista Ludwig Plate em 1910. Plate definiu pleiotropia da seguinte maneira: Eu denomino uma unidade de herança como pleiotrópica se várias características são dependentes desta unidade; essas características deverão aparecer sempre juntas e podem também ser correlacionadas. Essa definição, de modo geral, ainda é muito utilizada (STEARNS, 2010).

Um dos primeiros estudos relacionados à pleiotropia foi conduzido por Gruneberg (1938). Segundo Stearns (2010), uma das maiores contribuições do trabalho de Gruneberg foi a divisão do termo pleiotropia em “verdadeira pleiotropia” e “falsa pleiotropia”. A “verdadeira pleiotropia” - segundo Gruneberg - era caracterizada por dois ou mais “produtos primários distintos” provenientes do mesmo gene e responsáveis por diferentes fenótipos, enquanto a “falsa pleiotropia” se caracteriza quando apenas um “produto primário” é utilizado de diferentes maneiras na constituição fenotípica. Além dessas duas definições, esse autor caracterizou também outro aspecto - também denominado de “falsa pleiotropia”- onde um produto primário seria o responsável por uma cascata de eventos com diferentes consequências para o fenótipo. É importante destacar que o conceito de “verdadeira pleiotropia” de Gruneberg foi ignorado durante vários anos devido ao trabalho realizado por Beadle e Tatum (1941), que

inseriram a famosa relação onde um gene produziria no máximo uma proteína (“um gene/uma proteína”). Essa ideia foi prontamente adotada e, por consequência, a expressão “falsa pleiotropia” também caiu em desuso justamente por corroborar com as teorias de Beadle e Tatum, ou seja, a terminologia “falsa” tornou-se inapropriada, dado o postulado teórico da época.

Através da ideia “um gene/uma proteína”, novos conceitos relacionados ao funcionamento da pleiotropia foram estabelecidos. Hadorn (1961) denominou dois novos conceitos de pleiotropia, separando-a em pleiotropia “mosaica” e “relacional”. A “pleiotropia mosaica” é descrita quando um único loco altera diretamente dois caracteres fenotípicos, ao passo que a “pleiotropia relacional” se refere a uma cascata de eventos iniciada por um único loco que interfere em múltiplos caracteres.

Nos anos seguintes, diversos autores na área de genética de populações entenderam que poderia ocorrer uma “pleiotropia universal” (Figura 1), ou seja, uma mutação em qualquer loco tinha o potencial de afetar direta ou indiretamente todo o fenótipo do indivíduo (WRIGHT, 1968).

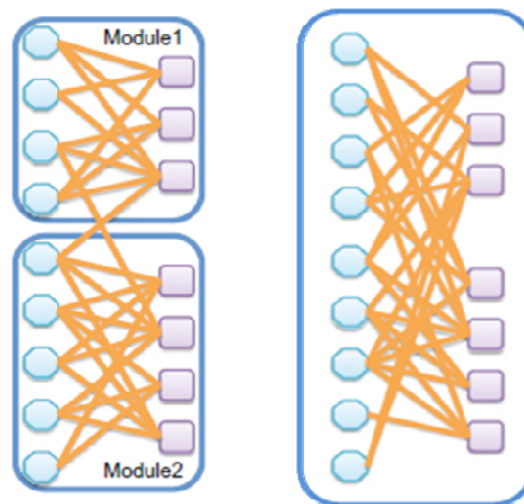


Figura 1 Modelo de pleiotropia modular (esquerda) e pleiotropia universal (Direita).
Fonte: Wang, Liao e Zhang (2010)

A “pleiotropia universal” acarreta diversos entraves ao processo evolutivo devido seu amplo espectro de atuação, convergindo no denominado “custo da complexidade” que pode ser predito pelo Modelo Geométrico de Fisher (WAGNER; ZHANG, 2011). Nesse modelo, o fenótipo de um indivíduo pode ser representado como um ponto em um espaço n -dimensional cujas dimensões correspondem ao número de caracteres que determinam o fenótipo do indivíduo (Figura 2). Quanto mais dimensões de variação um fenótipo possui, maior a dificuldade de adaptação do indivíduo por mutações aleatórias. Em outras palavras, se existem k pontos possíveis no espaço n -dimensional, é extremamente improvável que mutações aleatórias ajustem a combinação correta dos caracteres de modo a proporcionar uma melhor adaptação do indivíduo, acarretando, dessa forma, um custo adaptativo em organismos extremamente complexos. Fisher afirma que o efeito da mutação deveria ser infinitesimal e, somente sobre essa ótica, seria possível obter 50% de chance de melhorar um fenótipo via mutação aleatória. Segundo Wagner e Zhang (2011) o “custo da complexidade” deveria ser na verdade denominado como “custo da pleiotropia” uma vez que o modelo geométrico de Fisher assume que cada mutação afeta potencialmente todo o espaço dimensional. A “pleiotropia universal” é considerada hoje o principal paradigma dentro da teoria de genética de populações.

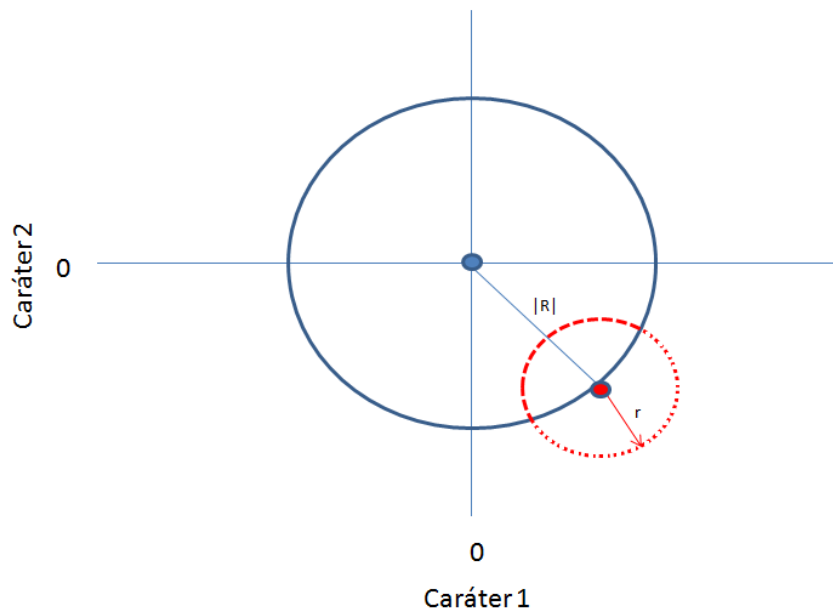


Figura 2 Modelo Geométrico de Fisher considerando 2 caracteres. O ponto azul, situado nas coordenadas $z(0,0)$, corresponde ao genótipo com adaptação ideal. O vetor $|R|$ corresponde à distância de um indivíduo qualquer em relação ao genótipo adaptado e o círculo azul corresponde aos indivíduos com a mesma adaptação relativa. O ponto vermelho representa um indivíduo que sofreu mutação, cuja direção é representada pelo vetor r . O círculo vermelho se relaciona a todas direções possíveis da mutação e que são igualmente prováveis.

Contudo, em estudos recentes, a “pleiotropia universal” não tem sido comumente observada e, ao contrário, se tem evidenciado a denominada “pleiotropia modular” (SU; ZENG; GU, 2010; WAGNER et al., 2008). A pleiotropia modular (Figura 1) é restrita a um determinado grupo de genes controladores de alguns caracteres, e não universal a todos os possíveis caracteres do indivíduo. Na pleiotropia modular, a taxa adaptativa de um indivíduo depende do grau de pleiotropia relacionado às mutações individuais, e não da complexidade do indivíduo. Assim, mutações de grande magnitude não necessariamente influenciariam todos os caracteres dado que seu espectro de atuação é restrito.

Evidências obtidas em estudos na área de genética molecular têm também demonstrado que a “verdadeira pleiotropia” descrita por Gruneberg pode ser a forma mais comum de pleiotropia (STEARNS, 2010; WAGNER; ZHANG, 2011). Tem sido constatado que um loco pode ter múltiplas regiões de início ou mesmo regiões sobrepostas de leitura, ou seja, uma única fita de um mesmo loco poderia ser lida de diferentes pontos produzindo diferentes proteínas. Como observado por Stearns (2010), esses locos pleiotrópicos muitas vezes são considerados como genes distintos, embora não possam ser separados por recombinação, pois estão situados na mesma região do genoma. Outros mecanismos como o “*splice alternativo*” e “*mRNA editing*” contribuem para que um único loco possa produzir proteínas /enzimas que atuem em diversas rotas metabólicas (GOMMANS; MULLEN; MAAS, 2009; MAYDANOVYCH; BEAL, 2006).

Proteínas multifuncionais ou proteínas de múltiplos propósitos também têm sido observadas (HE; ZANG, 2006). Essas proteínas se encaixam dentro da ideia de “falsa pleiotropia” descrita por Gruneberg, onde um único produto gênico pode atuar em diferentes rotas metabólicas (“*moonlighting proteins*”). No entanto, como destacado por Stearns (2010), no caso das “*moonlighting proteins*”, tem sido constatado que uma mutação em um loco pode alterar um processo metabólico, mas não necessariamente altera todos os processos dependentes da mesma proteína, ou seja, a mesma proteína possui funções independentes. Isso faz com que esse tipo de fenômeno não seja considerado uma “verdadeira pleiotropia”, pois os fenótipos controlados por essas proteínas podem ser independentes (HUBERTS; KLEI, 2010).

Recentemente, vários autores têm observado que a epistasia pode gerar correlação entre caracteres (Epistasia pleiotrópica) (WOLF et al., 2005, 2006), ou que a epistasia entre locos pode influenciar os efeitos pleiotrópicos (Epistasia Diferencial) (CHEVERUD et al., 2004; PAVLICEV et al., 2008). O termo

“epistasia diferencial” foi cunhado por Cheverud et al. (2004). A “epistasia diferencial” pode ser observada quando a interação entre pares de locos é diferente para os diferentes caracteres em que esses locos atuam de forma pleiotrópica, ou seja, o efeito de pleiotropia pode ser fortemente influenciado pela interação de locos pleiotrópicos (sinergia e antagonismo). Esses autores observaram que, sob o ponto de vista de eficiência evolutiva e adaptativa, a pleiotropia deveria ser um fenômeno aleatório e modular e não fixo e universal. De forma similar, Wolf et al. (2005) apresentou o termo “epistasia pleiotrópica”. Segundo esses autores, quando se estuda a arquitetura genética de múltiplos caracteres deve-se considerar a epistasia como uma forma de gerar covariância entre esses caracteres. Segundo Wolf et al. (2005), existem duas formas de interpretar a epistasia pleiotrópica: i) pode ser observada sob a ótica dos efeitos genéticos, onde a interação entre locos afeta a co-expressão de múltiplos caracteres e sendo assim, a pleiotropia seria uma propriedade de um sistema multi-locos, ii) Os locos individuais são pleiotrópicos, porém a magnitude dos efeitos de pleiotropia depende dos alelos presentes em cada loco individual.

Sob o postulado epistático modular, podemos supor que o efeito de pleiotropia pode ser dependente do *background* genético e, nesse caso, supõe-se que seja um fenômeno quase impossível de se prever. Essa suposição tem grande impacto dentro da eficiência adaptativa e evolutiva, pois fornece ao indivíduo maior capacidade adaptativa e menor dano quando mutações são realizadas em conjuntos de genes que atuam em caracteres independentes. Do ponto de vista do melhoramento genético, a pleiotropia epistática modular pode fornecer vantagens e desvantagens. A principal vantagem seria a possibilidade de “quebra pleiotrópica”, onde seria possível realizar seleção dentro de caracteres controlados pelo mesmo gene, porém que apresentam correlações genéticas indesejáveis. A principal desvantagem seria a possibilidade de diminuição da correlação entre caracteres quando se utiliza seleção assistida por

meio de QTLs pleiotrópicos e que são ao mesmo tempo hipostáticos. Como observado por Wolf et al. (2005), a pleiotropia epistática costuma aumentar a variância genética e influenciar de maneira sutil a covariância. Portanto, pode ocorrer diminuição da correlação genética, sendo esta altamente influenciada pela interação pleiotropia x background x frequências alélicas.

Sob o que foi enunciado posteriormente, podemos concluir que a pleiotropia é um fenômeno genético altamente complexo e tem papel fundamental na compreensão da arquitetura genética de qualquer espécie. No melhoramento genético, poucos estudos têm dado a devida atenção a esse fenômeno, e quando o fazem, não atentam para as diversas implicações desse elemento, ou seja, infere-se sobre a pleiotropia apenas como um fenômeno fixo de correlação genética e busca-se simplesmente distingui-la de ligação entre locos que é um evento transitório. Embora a estimativa de correlação genética entre caracteres tenha aplicação prática no melhoramento, seria ingênuo considerar apenas essa faceta da pleiotropia como aplicável, ou seja, a pleiotropia está relacionada à própria adaptação do indivíduo e a compreensão da sua arquitetura pode fornecer novas estratégias de melhoramento. Além disso, não há garantias que genes pleiotrópicos apresentem covariâncias estáveis sob qualquer circunstância, ou seja, esse efeito pode ser variável quando diferentes interações epistáticas ocorrem entre os locos que afetam múltiplos caracteres (MACKAY; STONE; AYROLES, 2009).

2.4 Componentes da produção e heterose multiplicativa

No melhoramento genético de plantas a decomposição de caracteres complexos em caracteres mais simples ajuda o melhorista a direcionar a seleção para algumas características desejáveis, realizar seleção indireta e obter uma maior compreensão do controle genético do caráter de interesse. O estudo de

componentes da produção, ou de subcaracteres se encaixam nesse contexto e ajudam o melhorista a: i) identificar os componentes específicos de acordo com os objetivos do programa de melhoramento; ii) gerar hipóteses que permitam aumentar os ganhos de produção; iii) detectar quais genes de um componente específico podem ser usados para aumentar a produção quando incorporado em um conjunto gênico elite.

Segundo Grafuis (1959), o caráter produção de grãos *per se* pode não possuir um grupo de genes específicos, pelo contrário, esse caráter tende a ser decomposto em vários componentes independentes ou subcaracteres. Em outras palavras, assumindo W como um caráter principal (produção de grãos), podemos descrevê-lo em função do número de espigas (R), número de grãos por fileira(S), número de fileiras de grãos (T), e peso de grãos (U) ($W = R \times S \times T \times U$).

Sob a hipótese de que a produção é um caráter complexo e resultado de um modelo multiplicativo, Schnell e Cockerham (1992) avaliaram a possibilidade de se entender a heterose como produto de uma ação multiplicativa entre subcaracteres. Sob o universo das possibilidades de interações multiplicativas entre genes que podem convergir em efeitos epistáticos, é possível obter heterose mesmo quando o efeito de dominância se encontra ausente nos subcaracteres. Esses autores descreveram a expressão da heterose sob a hipótese multiplicativa como sendo: $h = \bar{X}\bar{Y} - (X_1Y_1 + X_2Y_2)/2$, onde X_1 e Y_1 são os efeitos dos componentes da produção X e Y no parental 1, X_2 e Y_2 são os efeitos dos componentes da produção no parental 2, e $\bar{X} = (X_1 + X_2)/2$ e $\bar{Y} = (Y_1 + Y_2)/2$. De forma alternativa, esse modelo pode ser descrito como: $h = -(X_1 - X_2)(Y_1 - Y_2)/4$. Portanto, poderá ocorrer heterose positiva se $X_1 > X_2$ e $Y_2 > Y_1$ ou vice-versa.

Utilizando uma expressão linear, esses autores descreveram o valor genotípico de um indivíduo para um loco como sendo: $A_i = (p_1A_1 + p_2A_2) + r_i(A_1 - A_2)$, onde p_1 e p_2 são as frequências alélicas nos componentes de produção 1 e 2 e $r_1=p_2$ e $r_2=-p_1$. Aplicando um modelo multiplicativo entre dois locos, esses autores derivaram a média geral, os efeitos aditivos, dominantes e epistáticos para uma população F_2 da seguinte forma: $\mu = \mu_1\mu_2, a_1 = \alpha_1\mu_2, a_2 = \alpha_2\mu_1, d_1 = \delta_1\mu_2, d_2 = \delta_2\mu_1, i_{aa} = \alpha_1\alpha_2, i_{ad} = \alpha_1\delta_2$ e $i_{dd} = \delta_1\delta_2$, onde: μ_1 e μ_2 são as médias genotípicas para os componentes de produção 1 e 2, α_i é o efeito aditivo do i -ésimo componente que pode ser obtido pela métrica F_2 e δ_i é o efeito de dominância do i -ésimo componente.

Utilizando esses componentes, Schnell e Cockerham (1992) derivaram a expressão da heterose utilizando o efeito multiplicativo entre subcaracteres. Essa expressão da heterose corresponde àquela já descrita anteriormente no item 2.1 e, segundo esses autores, possivelmente vários caracteres são produtos da multiplicação de dois ou mais subcaracteres e a heterose multiplicativa pode ser mais frequente do que se tem considerado.

Dentro dos conceitos já discutidos acerca dos efeitos de pleiotropia, Schnell e Cockerham (1992) argumentam que, na pleiotropia modular, é possível que o efeito multiplicativo deva acontecer entre genes não-pleiotrópicos e esporadicamente entre genes pleiotrópicos, gerando dessa forma uma epistasia entre e dentro dos módulos. Por outro lado, assumindo que a hipótese de que a “pleiotropia universal” seja verdadeira, pode ocorrer um efeito parcial de multiplicação entre os componentes se os genes de efeitos principais responsáveis por cada caráter são diferentes entre si.

Assim, de forma generalizada, o modelo de Schnell e Cockerham (1992) permite prever os efeitos genéticos aditivos, dominantes e epistáticos de um

caráter complexo como a produção utilizando as estimativas dos QTL de seus componentes.

Com isso, observa-se que a análise de múltiplos caracteres oferece não somente a possibilidade de maior compreensão acerca dos efeitos de pleiotropia, como também permite prever heterose em caracteres complexos como a produção de grãos.

2.5 Análise de QTLs sob múltiplos caracteres

No melhoramento genético, a inferência acerca dos efeitos pleiotrópicos em estudos de mapeamento de QTLs normalmente é realizada empregando-se uma análise subjetiva de coincidência de posição do QTL ou intervalo de LOD, além de testes estatísticos via razão de verossimilhança como sugerido por Jiang e Zeng (1995) que pode não ser adequado para esse fim. No teste de razão de verossimilhança, duas hipóteses são comparadas com base na maximização de duas funções aninhadas. Sob hipótese nula, um ou mais parâmetros contidos no modelo completo são fixados no modelo reduzido (ex. média igual a zero, efeito de tratamento igual a zero etc.). Sob a suposição de que o modelo reduzido é adequado, a transformação $-2\ln$ da razão de verossimilhança deverá seguir assintoticamente uma distribuição qui-quadrado com graus de liberdade correspondentes à diferença entre o número de parâmetros maximizados no modelo completo e número de parâmetros especificado no modelo reduzido. No entanto, esse teste somente terá distribuição qui-quadrado se o modelo reduzido for aninhado com o modelo completo, ou seja, as duas distribuições devem possuir os mesmos parâmetros, sendo que em um modelo alguns desses parâmetros são fixos (modelo reduzido) e no outro esses mesmos parâmetros podem variar (modelo completo), sendo os demais parâmetros comuns a ambos os modelos. O teste de Jiang e Zeng (1995) claramente viola essa pressuposição

por assumir no modelo completo dois QTLs distintos, cada um afetando um caráter independente, e no modelo reduzido apenas um único QTL afetando ambos os caracteres. Nesses modelos, diferentes parâmetros são maximizados em cada hipótese, e assumir que dois QTLs possuem a mesma posição pode não resolver o problema por completo (WELLER, 2009). Além disso, à medida que se aumenta o número de parâmetros a serem maximizados, o poder da análise de verossimilhança multivariada é penalizado pelo aumento no número de graus de liberdade dos parâmetros (LIU et al., 2007).

Considerando a violação das pressuposições do teste de verossimilhança nesse tipo de análise, Sorensen et al. (2003) sugeriram uma distribuição empírica para teste da razão de verossimilhança. Utilizando uma análise similar, via abordagem bayesiana, Liu et al. (2007) e Varona et al. (2004) sugeriram o fator de Bayes como um teste alternativo livre de suposições assintóticas, onde o modelo que melhor se ajusta aos dados é tomado como o mais provável. Outra vantagem desse método é que o fator de Bayes é expresso em termos de probabilidade, sendo assim, seu resultado tem interpretação direta tornando os resultados comparáveis entre os experimentos.

Somente alguns estudos têm considerado o aspecto teórico da análise de QTLs com vários caracteres simultaneamente (BANERJEE; YANDELL; YI, 2008; EEUWIJK et al., 2010; GILBERT; LEROY, 2004; JIANG; ZENG, 1995; LIU et al., 2007; MALOSETTI et al., 2008; WELLER et al., 1996; XU et al., 2009). Nesses estudos, algumas metodologias têm sido aplicadas visando à identificação de QTLs com múltiplos caracteres. A primeira delas foi proposta por Jiang e Zeng (1995) e utiliza a máxima verossimilhança de uma função multivariada. Esses autores aplicaram a metodologia em dados simulados considerando uma situação bivariada e distribuição normal para os resíduos em ambos os caracteres. O grande problema da abordagem de máxima verossimilhança é que à medida que se aumenta o número de parâmetros,

aumenta-se o número de graus de liberdade, o que pode restringir o ganho da análise conjunta, perdendo sua vantagem em relação à análise de um único caráter (BANERJEE; YANDELL; YI, 2008; LIU et al., 2007). Outro aspecto importante a ser destacado é que no modelo multi-caracteres de Jiang e Zeng (1995) apenas a covariância residual é levada em consideração no mesmo. Assim, o efeito do QTL é ajustado considerando apenas a direção e magnitude da covariância residual, ou seja, não leva em consideração - de forma direta - a covariância do QTL, que pode ser contrária em direção e magnitude em relação à covariância residual, e também variável de QTL para QTL.

Outra metodologia proposta baseia-se no uso da técnica de redução dimensional, ou seja, análise de componentes principais (PCA), análise discriminante (DA) e transformação canônica dos caracteres originais (GILBERT; LEROY, 2004; MANGIN; THOQUET; GRIMSLEY, 1998; WELLER et al., 1996). A partir desses métodos utiliza-se a combinação linear dos caracteres com o objetivo de mapear QTLs. A grande dificuldade desse tipo de abordagem é que a combinação linear dos caracteres pode não ser biologicamente interpretável e ainda pode causar falsos efeitos de ligações (GILBERT; LEROY, 2004).

Outra abordagem recente sugere a modelagem mista de múltiplos caracteres e múltiplos ambientes em um único modelo (EEUWIJK et al., 2010; MALOSETTI et al., 2008). Esses autores consideram o efeito de QTL como fixo e os efeitos marginais das marcas como aleatórios (mapeamento por intervalo composto). Embora a modelagem conjunta de múltiplos ambientes e múltiplos caracteres seja desejável, a utilização de análise por intervalo simples pode contrabalancear esse possível ganho e, neste caso, novas modelagens são requeridas para inserir tanto os efeitos ambientais quanto o mapeamento por múltiplos intervalos em um único modelo linear misto. Nesta abordagem, a covariância do QTL também é confundida com a covariância residual, tanto em

direção quanto em magnitude, não explorando também as (co)variâncias específicas de cada QTL ao longo dos caracteres. Além disso, conceitualmente tomar um efeito de QTL como fixo aparentemente contrasta com toda a incerteza presente na arquitetura genética de um caráter (BUCHANAN et al., 2009), ou seja, se os efeitos são fixos, assume-se que eles variam não por suas propriedades biológicas, mas sim por suposições assintóticas da amostra observada ou amostras que poderiam ser observadas, mas não foram. A suposição de que um QTL possui um efeito fixo pode ser estatisticamente e/ou computacionalmente útil, mas pode não corroborar com a realidade dos efeitos interferentes na expressão de um loco quantitativo. Assim, pode ser preferível assumir que os efeitos dos QTLs são incertos e aleatórios e, baseados nos dados fenotípicos, obter uma distribuição de probabilidade de todos os possíveis efeitos de um suposto QTL e não apenas uma estimativa pontual cuja incerteza depende da aproximação assintótica de uma amostra em relação a uma população de tamanho infinito.

A abordagem bayesiana, conceitualmente, pode amenizar essas deficiências e, segundo Liu et al. (2007), essa metodologia representa um novo horizonte na análise genômica porque apresenta algumas vantagens significativas sobre as demais abordagens, e dentre elas podemos citar: i) maior flexibilidade em lidar com modelos mais complexos; ii) apresentação de resultados de fácil interpretação; iii) capacidade de incorporar informação de diferentes origens (priors). A vantagem i tem tornado a abordagem bayesiana extremamente atrativa, fornecendo a possibilidade de se trabalhar com modelos tão complexos que podem descrever a arquitetura genética de forma mais realística como, por exemplo, considerar que cada QTL tem variância individual e que seus efeitos, por menores que sejam, contribuem para a arquitetura genética do indivíduo (FLINT; MACKAY, 2009). Além disso, na análise bayesiana de múltiplos intervalos, o efeito do QTL é encolhido, o que faz com

que cada QTL, individualmente, tenha pouca influência na variância genética, apresentando um quadro mais próximo da arquitetura genética poligênica discutida no trabalho de Flint e Mackay (2009) e Mackay, Stone e Ayroles (2009). A análise Bayesiana de múltiplos intervalos também separa as (co)variâncias do QTL das (co)variâncias residuais, permitindo que a recuperação da informação admita integralmente o efeito e a direção do QTL nas estimativas das (co)variâncias, e que esta seja específica para cada loco. Outra vantagem é a possibilidade de aplicar a seleção genômica ampla diretamente sobre os efeitos dos QTLs, o que pode ser vantajoso em relação aos efeitos das marcas.

A abordagem bayesiana tem sido aplicada com sucesso no mapeamento de QTL para um único caráter e recentemente para múltiplos caracteres (BANERJEE; YANDELL; YI, 2008; LIU et al., 2007; XU et al., 2009). Devido a essas inúmeras vantagens essa abordagem será considerada nesse trabalho.

3 CONCLUSÕES

Pelo o que foi discutido na revisão de literatura, podemos inferir que a pleiotropia é um fenômeno pouco compreendido e explorado de forma superficial por geneticistas e melhoristas. Sendo assim, novos estudos são necessários para dissecação desse fenômeno buscando melhorar a compreensão da arquitetura genética de caracteres quantitativos. Podemos também concluir que o estudo de caracteres complexos exige uma modelagem estatística mais sofisticada. Nesse aspecto a inferência bayesiana pode ser de grande utilidade no estudo da pleiotropia ou mesmo na modelagem conjunta dos efeitos pleiotrópicos/epistáticos.

REFERÊNCIAS

- BANERJEE, S.; YANDELL, B. S.; YI, N. Bayesian quantitative trait loci mapping for multiple traits. **Genetics**, Baltimore, v. 179, n. 4, p. 2275-2289, Aug. 2008.
- BEADLE, G. W.; TATUM, E. L. Genetic control of biochemical reactions in neurospora. **Proceedings of the National Academy of Sciences of United States of America**, Washington, v. 27, n. 11, p. 499-506, Nov. 1941.
- BUCHANAN, A. V. et al. What are genes “for” or where are traits “from”? what is the question? **Bioessays**, Hoboken, v. 31, n. 2, p. 198-208, Feb. 2009.
- CHEVERUD, J. et al. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. **Journal of Experimental Zoology**, Hoboken, v. 302, n. 5, p. 424-435, Sept. 2004.
- COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. **Genetics**, Baltimore, v. 39, n. 6, p. 859-882, Nov. 1954.
- COCKERHAM, C. C.; ZENG, Z. B. Design III with marker loci. **Genetics**, Baltimore, v. 143, n. 3, p. 1437-1456, July 1996.
- COMSTOCK, R. E.; ROBINSON, H. F. Estimation of average dominance genes. In: GOWEN, J. W. (Ed.). **Heterosis**. Ames: Iowa State College, 1952. p. 494-516.
- EEUWIJK, F. A. van et al. Detection and use of QTL for complex traits in multiple environments. **Current Opinion Plant Biology**, London, v. 13, n. 2, p. 193-205, Apr. 2010.
- FLINT, J.; MACKAY, T. F. C. Genetic architecture of quantitative traits in mice, flies, and humans. **Genome Research**, Woodbury, v. 19, n. 5, p. 723-730, May 2009.
- FRASCAROLI, E. et al. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Baltimore, v. 176, n. 1, p. 625-644, May 2007.

- GAO, H. et al. Bayesian shrinkage mapping for multiple QTL in half-sib families. **Heredity**, London, v. 103, n. 5, p. 368-376, Nov. 2009.
- GARCIA, A. A. et al. Quantitative trait mapping and the genetic basis of heterosis in maize and rice. **Genetics**, Baltimore, v. 180, n. 3, p. 1707-1724, Nov. 2008.
- GILBERT, H.; LEROY, P. Comparison of three multitrait methods for QTL detection. **Genetics, Selection, Evolution**, London, v. 35, n. 3, p. 281-304, May 2003.
- GOMMANS, W. M.; MULLEN, S. P.; MAAS, S. RNA editing: a driving force for adaptive evolution? **Bioessays**, Hoboken, v. 31, n. 10, p. 1137-1145, Oct. 2009.
- GRAFUIS, J. E. Heterosis in barley. **Agronomy Journal**, Madison, v. 51, n. 9, p. 551-554, 1959.
- GREEN, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. **Biometrika**, London, v. 82, n. 4, p. 711-732, Dec. 1995.
- GRUNEBERG, H. An analysis of the "pleiotropic" effects of a new lethal mutation in the rat: *mus norvegicus*. **Proceeding of the Royal Society**, London, v. 125, n. 3, p. 123-144, Mar. 1938.
- HADORN, H. **Developmental genetics and lethal factors**. London: Methuen, 1961. 355 p.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, Feb. 1970.
- HE, X.; ZHANG, J. Toward a molecular understanding of pleiotropy. **Genetics**, Baltimore, v. 173, n. 4, p. 1885-1896, Aug. 2006.
- HUBERTS, D. H.; KLEI, I. J. van der. Moonlighting proteins: an intriguing mode of multitasking. **Biochimica et Biophysica Acta**, Amsterdam, v. 1803, n. 4, p. 520-525, Apr. 2010.
- JANSEN, R. C. A general mixture model for mapping quantitative trait loci by using molecular markers. **Theoretical and Applied Genetics**, Berlin, v. 85, n. 3, p. 252-260, Apr. 1992.

_____. Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. **Biometrics**, Malden, v. 49, n. 1, p. 227-231, Mar. 1993.

JIANG, C.; ZENG, Z. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Baltimore, v. 140, n. 3, p. 1111-1127, July 1995.

KAO, C. H.; ZENG, Z. B. Modeling epistasis of quantitative trait loci using Cockerham's model. **Genetics**, Baltimore, v. 160, n. 3, p. 1243-1261, Mar. 2002.

KAO, C. H.; ZENG, Z. B.; TEASDALE, R. D. Multiple interval mapping for quantitative trait loci. **Genetics**, Baltimore, v. 152, n. 3, p. 1203-1216, July 1999.

LANDER, E. S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Baltimore, v. 121, n. 1, p. 185-199, Jan. 1989.

LIU, J. et al. Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. **American Journal of Human Genetics**, Cambridge, v. 81, n. 2, p. 304-320, Aug. 2007.

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Genetics basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. **Theoretical and Applied Genetics**, Berlin, v. 107, n. 3, p. 494-502, Aug. 2003.

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

MACKAY, T. F. C. The genetic architecture of quantitative traits. **Annual Review of Genetics**, Palo Alto, v. 35, p. 336-339, Sept. 2001.

MACKAY, T. F.; STONE, E. A.; AYROLES, J. F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews Genetics**, London, v. 10, n. 8, p. 565-577, Aug. 2009.

MALOSETTI, M. et al. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Dordrecht, v. 161, n. 1, p. 241-257, Jan. 2008.

- MANGIN, B.; THOQUET, P.; GRIMSLEY, N. Pleiotropic QTL analysis. **Biometrics**, Malden, v. 54, n. 1, p. 88-99, Apr. 1998.
- MAYDANOVYCH, O.; BEAL, P. A. Breaking the central dogma by RNA editing. **Chemical Reviews**, Washington, v. 106, n. 8, p. 3397-3411, Aug. 2006.
- MELCHINGER, A. E. et al. The role of epistasis in the manifestation of heterosis: a systems-oriented approach. **Genetics**, Baltimore, v. 177, n. 3, p. 1815-1825, Nov. 2007.
- METROPOLIS, N. et al. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, New York, v. 21, p. 1087-1092, 1953.
- MEUWISSEN, N. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Baltimore, v. 157, n. 4, p. 1819-1829, Apr. 2001.
- PAVLICEV, M. et al. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. **Evolution**, Malden, v. 62, n. 1, p. 199-213, Jan. 2008.
- SATAGOPAN, J. M. et al. A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. **Genetics**, Baltimore, v. 144, n. 2, p. 805-816, Oct. 1996.
- SCHNELL, F. W.; COCKERHAM, C. C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics**, Baltimore, v. 131, n. 2, p. 461-469, June 1992.
- SCHÖN, C. C. et al. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010.
- SILLANPÄÄ, M. J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. **Genetics**, Baltimore, v. 151, n. 4, p. 1605-1619, Apr. 1999.
- SØRENSEN, P. et al. A comparison of bivariate and univariate QTL mapping in livestock populations. **Genetics, Selection, Evolution**, London, v. 35, n. 6, p. 605-622, Nov./Dec. 2003.
- STEARNS, F. W. One hundred years of pleiotropy: a retrospective. **Genetics**, Baltimore, v. 186, n. 3, p. 767-773, Nov. 2010.

STUBER, C. W. et al. Identification of genetic factors contributing to heterosis in a hybrid from two maize inbred lines using molecular markers. **Genetics**, Baltimore, v. 132, p. 823-839, Nov. 1992.

SU, Z.; ZENG, Y.; GU, X. A preliminary analysis of gene pleiotropy estimated from protein sequences. **Journal of Experimental Zoology**, Hoboken, v. 314, n. 2, p. 115-122, Mar. 2010.

TROYER, A. F. Adaptedness and heterosis in corn and mule hybrids. **Crop Science**, Madison, v. 46, n. 2, p. 528-543, Mar. 2006.

VARONA, L. et al. Derivation of a Bayes factor to distinguish between linked or pleiotropic quantitative trait loci. **Genetics**, Baltimore, v. 166, n. 2, p. 1025-1035, Feb. 2004.

VEEN, J. H. van der. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. **Genetica**, Dordrecht, v. 30, n. 1, p. 201-232, 1959.

WAGNER, G. P. et al. Pleiotropic scaling of gene effects and the 'cost of complexity'. **Nature**, London, v. 27, n. 7186, p. 470-472, Mar. 2008.

WAGNER, G. P.; ZHANG, J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. **Nature Reviews, Genetics**, London, v. 12, p. 204-213, Mar. 2011.

WANG, H. et al. Bayesian shrinkage estimation of quantitative trait loci parameters. **Genetics**, Baltimore, v. 170, n. 1, p. 465-480, May 2005.

WANG, Z.; LIAO, B. Y.; ZHANG, J. Genomic patterns of pleiotropy and the evolution of complexity. **Proceedings of the National Academic of Science of the United State of America**, Washington, v. 107, n. 42, p. 18034-18039, 2010.

WELLER, J. I. **Quantitative trait loci analysis in animals**. Cambridge: CABI, 2009. 261 p.

WELLER, J. I. et al. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment. **Theoretical Applied Genetics**, Berlin, v. 92, n. 4, p. 998-1002, Oct. 1996.

WOLF, J. B. et al. Contribution of epistatic pleiotropy to the genetic architecture of covariation among polygenic traits in mice. **Evolution Development**, Malden, v. 8, n. 5, p. 468-476, Oct. 2006.

_____. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. **Genetics**, Baltimore, v. 171, n. 2, p. 1-12, Oct. 2005.

WRIGHT, S. **Evolution and the genetics of populations**. Chicago: University of Chicago, 1968. 480 p.

WU, R.; MA, C. X.; CASELLA, G. **Statistical genetics of quantitative traits: linkage, maps and QTL**. New York: Springer-Verlag, 2007. 361 p.

XU, C. et al. Mapping QTL for multiple traits using Bayesian statistics. **Genetics Research**, New York, v. 91, n. 1, p. 23-37, Feb. 2009.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Baltimore, v. 163, n. 2, p. 789-801, Feb. 2003.

YANG, R. C. Epistasis of quantitative trait loci under different gene action models. **Genetics**, Baltimore, v. 167, n. 3, p. 1493-1505, July 2004.

YI, N.; GEORGE, V.; ALLISON, D. B. Stochastic search variable selection for identifying multiple quantitative trait loci. **Genetics**, Baltimore, v. 164, n. 3, p. 1129-1138, July 2003.

ZENG, Z. B. Precision mapping of quantitative trait loci. **Genetics**, Baltimore, v. 136, n. 4, p. 1457-1468, Apr. 1994.

_____. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings of the National Academic of Science of the United State of America**, Washington, v. 90, n. 23, p. 10972-10976, Dec. 1993.

ZENG, Z. B. et al. Genetic architecture of a morphological shape difference between two *Drosophila* species. **Genetics**, Baltimore, v. 154, n. 1, p. 299-310, Jan. 2000.

ZENG, Z. B.; WANG, T.; ZOU, W. Modeling quantitative trait loci and interpretation of models. **Genetics**, Baltimore, v. 169, n. 3, p. 1711-1725, Mar. 2005.

SEGUNDA PARTE - ARTIGO

ARTIGO-1

MAPEAMENTO BAYESIANO DE MÚLTIPLOS CARACTERES EM MILHO: A IMPORTÂNCIA DOS EFEITOS PLEIOTRÓPICOS NO ESTUDO DA HERANÇA DE CARACTERES QUANTITATIVOS

Artigo redigido nas normas da Revista Científica Theoretical and Applied Genetics (Versão preliminar)

Mapeamento bayesiano de múltiplos caracteres em milho: a importância dos efeitos pleiotrópicos no estudo da herança de caracteres quantitativos

Marcio Balestre¹, Renzo Garcia Von Pinho², Cláudio Lopes de Souza Júnior³, Júlio Sílvio de Souza Bueno Filho⁴

¹ Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brazil. e-mail marciobalestre@hotmail.com

² Departamento de Agricultura, Universidade Federal de Lavras, Lavras MG, Brazil.

³ Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo (ESALQ/USP), Departamento de Genética, Piracicaba, SP, Brazil

⁴ Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras MG, Brazil.

Resumo

A pleiotropia tem desempenhado papel fundamental na dissecação de caracteres quantitativos, porém a extensão desse efeito e suas consequências para o melhoramento de plantas ainda não foram completamente elucidadas. Esse trabalho foi realizado com o objetivo de identificar QTLs pleiotrópicos na cultura do milho utilizando o mapeamento Bayesiano de múltiplos intervalos. Além disso, procurou-se obter uma maior compreensão acerca da herança, extensão e distribuição dos efeitos pleiotrópicos em alguns componentes da produção do milho. Utilizou-se 177 marcadores de microssatélites no mapeamento de 250 plantas F_2 - representadas por suas progênies $F_{2:3}$ – que posteriormente foram retrocruzadas com as duas linhagens parentais (L-14-04 B e L-08-05 F) obtendo-se dessa forma 500 progênies que, em conjunto, caracterizam o delineamento III. Os resultados deste trabalho sugerem que a análise de caracteres isolados restringe a compreensão acerca da arquitetura genética de caracteres quantitativos. Contudo, essa arquitetura pode ser melhor compreendida através de redes pleiotrópicas que permitam visualizar a complexidade da herança desses caracteres e, a partir disto, traçar novas

estratégias de seleção. Também foi possível contestar a ideia de que é possível identificar QTLs “para” caracteres complexos tais como a produção de grãos, visto que a pleiotropia atua de forma destacada em seus subcaracteres e que esse “caráter” pode ser decomposto e predito quase que na sua totalidade pelos QTLs dos seus componentes. Além disso, a pleiotropia de QTLs não necessariamente significa pleiotropia das interações alélicas e, sendo assim, a ampla distribuição dos efeitos pleiotrópicos não restringe os ganhos com o melhoramento genético.

Palavras-chave: Pleiotropia. Caracteres complexos. Análise Bayesiana. Múltiplos caracteres.

Introdução

A compreensão da arquitetura genética de caracteres complexos ainda é um grande desafio aos geneticistas. Fenômenos como epistasia, pleiotropia e heterose tornam o entendimento do controle genético de caracteres quantitativo extremamente complexo e, por assim dizer, uma verdadeira caixa preta. Estudos recentes buscam lançar luz acerca desses fenômenos (Melchinger et al., 2007; Flint e Mackay, 2009; Stears, 2010). No entanto, a dissecação desses elementos só é possível utilizando técnicas moleculares complexas e modelos estatísticos sofisticados.

Um desses elementos de alta complexidade é a pleiotropia. A importância da pleiotropia no melhoramento de plantas advém das correlações genéticas entre caracteres de importância agrônômica e econômica, que também são observadas no desequilíbrio de ligação. No entanto, as correlações obtidas através do desequilíbrio de ligação são transitórias e a velocidade com que o desequilíbrio é dissipado depende principalmente da distância entre os genes. Por outro lado, sendo a pleiotropia um fenômeno onde o mesmo loco controla diferentes caracteres, seu efeito é mais estável e por isso a identificação de QTLs

pleiotrópicos torna-se importante. Sob o ponto de vista de seleção de múltiplos caracteres e estratégias de melhoramento, é relevante distinguir as correlações genéticas provenientes dos efeitos de ligação das correlações oriundas de genes pleiotrópicos, embora outros fenômenos envolvidos na arquitetura dos genes pleiotrópicos devam ser melhor compreendidos (Flint e Mackay, 2009; Mackay et al., 2009).

O entendimento do arcabouço pleiotrópico ainda é muito restrito e, sendo assim, diversas hipóteses foram levantadas para melhor compreensão desse fenômeno. Um dos primeiros estudos relacionados à pleiotropia foi conduzido por Gruneberg (1934), que definiu os termos “verdadeira pleiotropia”, o qual infere sobre vários “produtos genéticos primários” provenientes do mesmo gene e que atuam em diferentes rotas metabólicas, e a “falsa pleiotropia”, onde apenas um produto primário atua em diferentes rotas metabólicas ou ainda podem dar início a uma cascata de eventos com diferentes consequências para o fenótipo. A “falsa pleiotropia” é observada em proteínas de múltiplas funções (“*moonlighting*” *proteins*), porém alguns autores não consideram esses genes como pleiotrópicos uma vez que as mutações nesses locos não necessariamente alteram outros processos dependentes da mesma proteína (Huberts e Van der Klei, 2010; Stears, 2010). Por outro lado, evidências obtidas em estudos na área de genética molecular têm demonstrado que a “verdadeira pleiotropia” descrita por Gruneberg pode ser a forma mais comum de pleiotropia (Maydanovych e Beal, 2006; Gommans et al., 2009; Stearns, 2010, Wagner e Zhang, 2011).

Outra dúvida se refere à distribuição dos efeitos pleiotrópicos ao longo do genoma. Na hipótese de “pleiotropia universal”, todos os genes afetam, em maior ou menor intensidade, todos os caracteres que constituem o fenótipo de um indivíduo e dentro dessa ideia uma mutação em qualquer loco tem o potencial de afetar direta ou indiretamente todo o fenótipo do indivíduo (Wright,

1968). Sob a hipótese pleiotrópica universal, não existe um gene específico “para” um caráter, mas sim uma rede extremamente complexa onde todos os genes contribuem com diferentes magnitudes de efeitos para o caráter em questão; e se esse for o caso, a identificação de QTLs “para” um determinado caráter pode ser enganosa (Buchanan et al., 2009; Mackay et al., 2009). Embora essa ideia seja ainda defendida por alguns autores (Carbone et al. 2006; Buchanan et al., 2009; Mackay et al., 2009), esse fenômeno não tem sido comumente observado em estudos recentes, onde se tem evidenciado a denominada “pleiotropia modular” (Wagner et al., 2008; Su et al., 2010, Wagner e Zhang, 2011). A pleiotropia modular é restrita a um determinado grupo de genes controladores de alguns caracteres, e não universal a todos os possíveis caracteres do indivíduo.

Além do que já foi comentado acerca do funcionamento e distribuição dos genes pleiotrópicos, evidências apontam que a epistasia pode gerar correlação entre caracteres (Epistasia Pleiotrópica: Wolf et al., 2005; Wolf et al., 2006), ou que a epistasia entre locos pode influenciar nos efeitos pleiotrópicos de locos individuais (Epistasia Diferencial: Cheverud, et al., 2004; Pavlicev et al., 2007). Esses autores levantam a questão de que a pleiotropia pode ser fortemente influenciada pela epistasia e sendo assim não há garantias que genes pleiotrópicos apresentem covariâncias estáveis sob qualquer circunstância, ou seja, a magnitude desse efeito pode ser variável quando diferentes interações epistáticas ocorrem entre os locos que afetam múltiplos caracteres (Mackay et al., 2009).

Sob o que foi enunciado anteriormente, podemos concluir que a pleiotropia é um fenômeno genético altamente complexo e tem papel fundamental na compreensão da arquitetura genética de qualquer espécie. No melhoramento genético, poucos estudos têm dado a devida atenção a esse fenômeno, e quando o fazem, não atentam para as diversas implicações desse

fenômeno no melhoramento, ou seja, infere-se sobre a pleiotropia como um fenômeno fixo de correlação genética e busca-se simplesmente distingui-la do efeito de ligação entre locos que é um evento transitório. Sendo assim, além de distinguir pleiotropia de ligação, torna-se necessário também verificar a extensão desse efeito ao longo do genoma sob diferentes caracteres com o objetivo de se buscar novas estratégias de seleção além de maior compreensão acerca da herança de caracteres quantitativos.

Para a identificação de QTLs pleiotrópicos e estudo da sua distribuição ao longo do genoma, a análise bayesiana de múltiplos intervalos pode ser considerada um ferramenta poderosa (Wang et al. 2005, Xu et al., 2009). A análise sugerida por esses autores é irrestrita ao número de parâmetros e livre de seleção de modelos, ou seja, cada intervalo delimitado por dois marcadores distribuídos ao longo do genoma é aceito como candidato a conter um QTL e a posição desse QTL pode variar aleatoriamente dentro de cada intervalo. A suposição de que cada QTL possui uma variância individual força os intervalos com pequeno efeito a valores próximos de zero, fazendo com que sua inclusão tenha pouco efeito nos demais intervalos. Por outro lado, QTLs com efeito expressivo tendem a ter uma variância menos penalizada pelo modelo e se destacam sobre os demais. Por essas propriedades essa análise é denominada análise bayesiana “shrinkage” (Wang et al. 2005).

Apesar de aparentemente violar o princípio da parcimônia de modelos, esse método se apresenta como o mais representativo da possível arquitetura de um caráter oligogênico/poligênico, onde genes principais apresentam de moderada/alta influência sobre o caráter, e quando combinados com vários outros genes de pequeno efeito determinam a variação genética de um caráter quantitativo (Mackay, 2001; Flint e Mackay, 2009). Assim, é importante manter esses QTLs de pequeno efeito no modelo, pois coletivamente podem realizar um importante papel na variância genética do caráter (Wang et al., 2005).

Devido à importância da pleiotropia na dissecação de caracteres quantitativos, neste trabalho, buscou-se realizar o mapeamento de múltiplos caracteres utilizando a abordagem bayesiana para a identificação de QTLs pleiotrópicos em milho. Além disso, procurou-se obter uma maior compreensão acerca da herança, extensão e distribuição dos efeitos pleiotrópicos em alguns componentes da produção do milho.

Material e Métodos

Material genético

Para obtenção da população F_2 realizou-se o cruzamento de duas linhagens contrastantes para diversos caracteres de importância agrônômica (L-14-04 B e L-08-05F). Essas linhagens pertencem a grupos heteróticos distintos, sendo que a L-14-04 B foi derivada da população BR-106 (grãos amarelos e dentados) e a L-08-05 F da população IG-1 (grãos alaranjados e duros). A população BR-106 foi obtida pela EMBRAPA Milho e Sorgo através do cruzamento de populações brasileiras e mexicanas. A população IG-1 foi obtida pelo Departamento de Genética da ESALQ/USP e foi desenvolvida pelo cruzamento de populações brasileiras e tailandesas.

Duzentos e cinquenta plantas da geração F_2 foram autofecundadas dando origem a progênies $F_{2,3}$. Essas progênies foram retrocruzadas para o parental L-14-04 B, formando as progênies RC_1 e também retrocruzadas para o parental L-08-05 F, formando as progênies RC_2 . Esses cruzamentos foram realizados em lotes isolados de despendoamento, formando assim 500 progênies de retrocruzamento e caracterizando assim o delineamento III. As sementes das progênies $F_{2,3}$ retrocruzadas para ambos os parentais foram utilizadas nas avaliações experimentais.

Procedimento experimental

As 500 progênies obtidas nos retrocruzamentos foram avaliadas em seis ambientes na estação experimental ESAL/USP, Piracicaba - SP. O delineamento utilizado foi o látice 10x10 com duas repetições por ambiente. Em cada látice foram alocadas 100 progênies de retrocruzamento, sendo 50 progênies $F_{2:3}$ retrocruzadas para cada um dos parentais. Assim, formaram-se cinco látices (experimentos) por ambiente. Em todos os ambientes os experimentos e suas repetições foram aleatorizados dentro das áreas em que foram instalados. As parcelas foram formadas por linhas de 4,0 metros espaçadas 0.80 metros umas das outras contendo 20 plantas por parcela ($62.500 \text{ plantas ha}^{-1}$). Os caracteres avaliados foram produção de grãos (PG) em kg parcela^{-1} , posteriormente ajustado para 15% de umidade dos grãos e para o estande médio dos experimentos; diâmetro de espiga (DE) comprimento da espiga (CE) em milímetro, número de grãos por fileira (NGF) e número de fileiras (NFI) . Os caracteres DE, CE, NGF e NFI foram analisados utilizando a média de cinco observações de cada parcela. Detalhes a cerca dos procedimentos experimentais podem ser encontrados em Silva (2002).

Mapa genético

O mapa genético utilizado, bem como os procedimentos para obtenção do mesmo, foi previamente descrito por Sibov et al. (2003). As análises moleculares foram realizadas no CEBEMEG/Unicamp. Brevemente, as plantas F_2 que deram origem às progênies $F_{2:3}$ foram genotipadas com marcadores microsatélite. O mapa genético foi desenvolvido utilizando o programa MAPMAKER/EXP versão 3.0b (Lincoln et al., 1992) com um LOD de 3.0 e uma distância máxima entre marcas adjacentes de 50 cM para formar os grupos de ligação. Sessenta novos marcadores microsatélite foram adicionados ao mapa de Sibov et al. (2003), totalizando 177 marcadores distribuídos ao longo

dos 10 cromossomos do milho. O mapa genético cobriu 2.052 cM do genoma do milho com uma distância média de 11,6 cM entre marcadores.

Análise de QTL

Modelo linear

A análise bayesiana de múltiplos intervalos e múltiplos caracteres utilizada nesse trabalho foi semelhante à apresentada por Xu et al. (2009). Nessa abordagem assume-se que cada intervalo possui um QTL em potencial. Os efeitos dos QTLs são aleatórios e com variância individual. Nessa metodologia, os dados observados são as marcas (m) e o fenótipo (y). As variáveis não observáveis são os genótipos dos QTLs (x e w), seus efeitos (a e d), as matrizes de (co)variâncias aditiva, dominante e residual ($A_k, D_k, \epsilon \Sigma$) e também as posições dos QTL (2)

Assim, utilizando-se as médias fenotípicas obtidas da análise conjunta nos seis ambientes e mais os 177 marcadores de microssatélites, adotou-se o seguinte modelo linear:

$$y_{ij} = b_j + \sum_{k=1}^p x_{ijk} a_k + \sum_{k=1}^p w_{ijk} d_k + e_i \quad (1)$$

onde y_{ij} é o vetor das observações fenotípicas $[y_{1ij} \dots y_{qij}]^T$ sendo q o número de caracteres avaliados na i -ésima progênie no j -ésimo retrocruzamento; $b_j = [b_{1j} \dots b_{qj}]^T$ é o vetor das médias populacionais para o caráter q no j -ésimo retrocruzamento; $a_k = [a_{1k} \dots a_{qk}]^T$ e $d_k = [d_{1k} \dots d_{qk}]^T$ são os efeitos aditivos e dominantes respectivamente para o k -ésimo loco. O resíduo é um vetor $q \times 1$ assumido como multivariado normal $MVN(0, \Sigma)$ onde Σ é uma matriz $q \times q$. As variáveis x_{ijk} e w_{ijk} foram assumidas sobre a métrica F_∞ , sendo então ortogonais e semelhantes à métrica F_2 quando os efeitos epistáticos não são

ajustados no modelo (Yang, 2004; Zeng et al., 2005). Dessa forma podemos assumir que:

$$x_{ijk} = \begin{cases} 1 & \text{se } QQ \\ 0 & \text{se } Qq \\ -1 & \text{se } qq \end{cases} \text{ e } w_{ijk} = \begin{cases} 0 & \text{se } QQ \\ 1 & \text{se } Qq \\ 0 & \text{se } qq \end{cases} \quad (2)$$

As variáveis x_{ijk} e w_{ijk} não são observáveis, porém podem ser inferidas a partir da informação dos genótipos das marcas flaqueadoras e a da posição do QTL.

Função de verossimilhança

Para simplificar será considerada a seguinte equivalência vetorial $y=y_{ij}$, $b=b_j$, $x=x_{ij}$ e $w=w_{ij}$. Sob a suposição de normalidade dos resíduos pode-se assumir que a probabilidade condicional de y é dada por:

$$p(y | b, a, d, x, w, \Sigma) = N \left(b + \sum_{k=1}^p x_k a_k + \sum_{k=1}^p w_k d_k, \Sigma \right) \quad (3)$$

Dessa forma, função de verossimilhança pode ser dada por:

$$p(y | \lambda, b, a, d, x, w, \Sigma) \propto |\Sigma \otimes I|^{-1/2} \exp \left\{ -\frac{1}{2} \left(y - b - \sum_{k=1}^p x_k a_k - \sum_{k=1}^p w_k d_k \right) (\Sigma \otimes I)^{-1} \left(y - b - \sum_{k=1}^p x_k a_k - \sum_{k=1}^p w_k d_k \right) \right\} \quad (4)$$

Os parâmetros serão assumidos como $\theta = \{\lambda, b, a, d, \Sigma, A_k, D_k\}$. O número de QTLs normalmente é assumido como um parâmetro de interesse no mapeamento clássico, mas na análise bayesiana “shrinkage” o número de QTLs é uma constante condicionada ao número de marcadores ou intervalos. Se um intervalo não contém um QTL seu valor encolhe a zero, o que é equivalente a um QTL sendo excluído do modelo.

Distribuição a priori

Cada parâmetro do modelo possui uma distribuição a priori. Dessa forma podemos assumir que $p(b) \propto 1$, $p(a_k) \propto N(0, A_k)$, $p(d_k) \propto N(0, D_k)$. As matrizes de variâncias e covariâncias aditiva (A_k) e dominantes (D_k) possuem dimensão $q \times q$. Essas matrizes possuem valores que variam para os diferentes locos, ou seja, são consideradas de efeito específico. A distribuição a priori dessas matrizes seguem uma distribuição Wishart invertida e podem ser dadas por: $p(A_k) = \text{inv-Wishart}(\tau, \Gamma)$ e $p(D_k) = \text{inv-Wishart}(\tau, \Gamma)$ sendo $\tau > q$ e $\Gamma > 0$. Nesse trabalho assumiu-se valores mínimos para esses hiperparâmetros de modo a não alterar a inferência a posteriori, ou seja, foi considerado que $\tau > q$ e $\Gamma > 0$ como sugerido por Xu et al. (2009). Para a matriz dos resíduos Σ é assumida a mesma priori $p(\Sigma) = \text{inv-Wishart}(\tau, \Gamma)$ e, neste caso, devido ao número de graus de liberdade contido nos dados, os hiperparâmetros terão pouca influência nos valores na estimativa de Σ . Em dados simulados, Xu et al. (2009) demonstraram que essas suposições praticamente não alteraram as inferências acerca dos efeitos e posições dos QTLs. A priori relativa à posição do QTL pode ser uniforme, ou seja, assumindo que M_k^L e M_k^R são duas marcas que flanqueiam o genótipo do QTL Q_k , e que L_k e U_k sejam as distâncias entre $M_k^L \leftrightarrow Q_k$ e $Q_k \leftrightarrow M_k^R$ respectivamente, a priori uniforme para cada intervalo é dada por um conjunto de números ordenados de mesma probabilidade e que variam de $[L_k \text{ até } U_k]$.

Assim, a distribuição priori conjunta é dada por:

$$p(\theta) = p(b) p(\Sigma) \prod_{k=1}^p p(\lambda_k) p(a_k) p(d_k) p(A_k) p(D_k) \quad (5)$$

Distribuição posteriori

A distribuição posteriori dos parâmetros pode ser dada por:

$$p(\theta, x, w | y) \propto p(y | b, a, d, x, w, \Sigma) p(x, w | \lambda) p(\theta) \quad (6)$$

onde $p(x, w | \lambda)$ é a probabilidade do genótipo do QTL dada sua posição.

Embora essa distribuição seja de difícil tratamento, pode-se utilizar o método Monte Carlo Cadeias de Markov (MCCM) para obter amostras da distribuição posteriori conjunta. Para facilitar o processo de amostragem, cada parâmetro é amostrado por vez, e essa amostragem é condicional a todos os demais parâmetros. Se a distribuição de probabilidade condicional dos parâmetros tem uma forma explícita, pode-se aplicar o amostrador de Gibbs, retirando amostras diretamente de distribuições conhecidas. As distribuições condicionais dos parâmetros podem ser dadas como disposto.

A distribuição posteriori condicional da média populacional referente aos dois retrocruzamentos é multivariada normal com média e variância dadas por:

$$b_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(y_{ij} - \sum_{k=1}^p x_{ijk} a_k - \sum_{k=1}^p w_{ijk} d_k \right) \quad \text{e} \quad \text{var}(b_j) = \left(\frac{1}{n_j} \right) \Sigma \quad (7) \quad (8)$$

Dessa forma, para cada retrocruzamento é assumida uma média distinta.

A distribuição condicional posteriori dos efeitos aditivos (a_k) é multivariada normal com média e variância dadas por:

$$a_k = \left(\sum_{i=1}^n x_{ijk}^2 \Sigma^{-1} + A_k^{-1} \right)^{-1} \sum_{i=1}^n x_{ijk}^2 \Sigma^{-1} \left(y_{ij} - b_j - \sum_{k' \neq k}^p x_{ijk'} a_{k'} - \sum_{k=1}^p w_{ijk} d_k \right) \quad (9)$$

$$\text{e} \quad \text{var}(a_k) = \left(\sum_{i=1}^n x_{ijk}^2 \Sigma^{-1} + A_k^{-1} \right)^{-1} \quad (10)$$

De modo semelhante, a posteriori condicional para os efeitos de dominância é normal multivariada com média e variâncias dadas por:

$$d_k = \left(\sum_{i=1}^n w_{ijk}^2 \Sigma^{-1} + D_k^{-1} \right)^{-1} \sum_{i=1}^n w_{ijk}^2 \Sigma^{-1} \left(y_{ij} - b_j - \sum_{k=1}^p x_{ijk} a_k - \sum_{k' \neq k}^p w_{ijk'} d_{k'} \right) \quad (11)$$

$$\text{e } \text{var}(d_k) = \left(\sum_{i=1}^n x_{ijk}^2 \Sigma^{-1} + D_k^{-1} \right)^{-1} \quad (12)$$

A priori das matrizes de (co)variâncias é conjugada, e assim as condicionais posterioris de A_k e D_k são também Wishart invertida e podem ser dadas por:

$$p(A_k | \dots) = \text{Inv-Wishart}(\tau + 1, \Gamma + a_k a_k^T) \quad (13)$$

$$p(D_k | \dots) = \text{Inv-Wishart}(\tau + 1, \Gamma + d_k d_k^T) \quad (14)$$

A posteriori condicional da matriz de (co)variância residual é também Wishart invertida e pode ser dada por:

$$p(\Sigma | \dots) = \text{Inv-Wishart}(\tau + n, \Gamma + SS) \quad (15)$$

onde

$$SS = \sum_{i=1}^n \left(y_{ij} - b_j - \sum_{k=1}^p x_{ijk} a_k - \sum_{k=1}^p w_{ijk} d_k \right) \left(y_{ij} - b_j - \sum_{k=1}^p x_{ijk} a_k - \sum_{k=1}^p w_{ijk} d_k \right)^T \quad (16)$$

No delineamento III, cada retrocruzamento configura uma população distinta. Embora os retrocruzamentos sejam analisados conjuntamente, a probabilidade condicional do genótipo do QTL dado os marcadores é obtida para cada população separadamente (Tabela 1). Assumindo a codificação dada em (2) e tomando $g=(g_1=1, g_2=0 \text{ e } g_3=-1)$ para x_{ijk} e $h=(h_1=0, h_2=1 \text{ e } h_3=0)$ para w_{ijk} , a probabilidade condicional posteriori para o genótipo do QTL é calculada utilizando o teorema de Bayes da seguinte forma:

$$p(x_{ijk} = g | \dots) = \frac{p(x_{ijk} = g) H_{kL}(g, m_l) H_{kR}(g, m_r) p(y_{ij} | b_j, a_k, d_k, x_{ijk}, w_{ijk}, \Sigma)}{\sum_{z=1}^3 p(x_{ijk} = z) H_{kL}(z, m_l) H_{kR}(z, m_r) p(y_{ij} | b_j, a_k, d_k, x_{ijk}, w_{ijk}, \Sigma)}$$

(17)

em que : $p(x_{ijk} = g)$ é a probabilidade a priori do genótipo do QTL dada a segregação esperada nos retrocruzamentos, por exemplo:
 $p(x_{ijk} = g_1) = p(x_{ijk} = g_2) = \frac{1}{2}$ para RC1 e $p(x_{ijk} = g_2) = p(x_{ijk} = g_3) = \frac{1}{2}$ para RC2. As variáveis $H_{kL}(g, m_l)$ e $H_{kR}(g, m_r)$ são as matrizes de transição entre as marcas M_k^L e M_k^R e o QTL Q_k . Essas matrizes são construídas em função do produto de kronecker das probabilidades condicionais expressas na Tabela 1.

O parâmetro λ não possui uma função conhecida, sendo assim, a utilização do amostrador de Gibbs é proibitiva. No entanto, outro método Monte Carlo Cadeias de Markov (MCCM), denominado Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) pode ser utilizado. Esse algoritmo não exige que o parâmetro tenha uma função de probabilidade conhecida e, ao invés disso, faz-se uso de uma função auxiliar possível de ser amostrada, retirando valores candidatos que podem ser aceitos com um α de probabilidade. No caso de mapeamento por intervalo, pode-se utilizar uma distribuição uniforme como função auxiliar que é amostrada sobre cada intervalo delimitados por $\max(\lambda_{j-1}, \lambda_j + \delta)$ e $\min(\lambda_{j+1}, \lambda_j + \delta)$ onde δ é uma constante que define o caminhamento dentro do intervalo j , normalmente fixada entre 1 e 2 cM. Essa função é denotada por $u(\lambda^*, \lambda)$, e a nova posição λ^* será aceita na k -ésima iteração com $\min(1, \alpha)$ de probabilidade sendo α dado por:

$$\alpha = \frac{p(\lambda_j^* | y, b, x, w, a, d, \Sigma) u(\lambda_j^*, \lambda_j)}{p(\lambda_j^0 | y, b, x, w, a, d, \Sigma) u(\lambda_j, \lambda_j^*)} \quad (18)$$

E, finalmente, os genótipos perdidos das marcas são amostrados de sua distribuição condicional posteriori que pode ser calculada utilizando o teorema de Bayes

$$p(m_{ijk} = g | \dots) = \frac{p(m_{ijk} = g)M_{k-1}(g, m_{k-1})M_{k+1}(g, m_{k+1})}{\sum_{z=1}^3 p(m_{ijk} = g)M_{k-1}(g, x_{k-1})M_{k+1}(g, x_{k+1})}$$

em que: $p(m_{ijk} = g_1) = p(m_{ijk} = g_3) = \frac{1}{4}$ e $p(m_{ijk} = g_2) = \frac{1}{2}$ e M_{k-1} e M_{k+1} são as matrizes referentes ao genótipo dos marcadores que flanqueiam a marca perdida e suas distâncias.

Por essas expressões fica claro que as marcas seguem uma segregação de F_2 ao passo que os QTLs segregam de acordo com o esperado em populações de retrocruzamento. Por esse fato, a predição da marca perdida com base no genótipo dos QTLs torna-se onerosa e requer matrizes que comportem a probabilidade do RC1 e RC2. Assim, devido à densidade de marcadores utilizados, a predição das marcas perdidas com base nas marcas flanqueadoras não causa danos à análise. Todas as análises foram realizadas utilizando um programa desenvolvido no SAS/IML.

Análise pós-MCCM

Na inferência bayesiana, o teste de significância não tem grande importância quanto na análise de verossimilhança. Antes, o objetivo da análise bayesiana via MCCM é obter uma distribuição posteriori empírica de onde toda informação a respeito do QTL pode ser obtida. Na análise bayesiana simples, a posição do QTL é inferida com base na quantidade de vezes que o efeito do QTL passa por pequena região (bin) em determinada posição do genoma. Essa curva descreve o perfil de intensidade do QTL. Na abordagem de Wang et al. (2005), é assumido que cada intervalo está associado a um QTL, de modo que

em todos os intervalos, o suposto QTL irá passar por todas as regiões do genoma, e em cada intervalo o mesmo número de “batida” do QTL irá ocorrer independente do seu efeito. No entanto, é esperado que se ocorrer um verdadeiro QTL em um dado intervalo, sua posição demonstrará um pico ao passo que se o efeito é nulo a distribuição dentro do intervalo é uniforme (Yang e Xu, 2007). O perfil da intensidade do QTL é representada por Yang e Xu (2007) como uma função da posição $f(\lambda)$. No entanto $f(\lambda)$ pode não ser suficientemente informativa para inferência acerca do QTL na análise bayesiana “shrinkage”. Baseado nisto, Yang e Xu (2007), propuseram a descrição dos efeitos dos QTLs de acordo com suas formas quadráticas e ponderadas pela intensidade da posição: $g(\lambda) = W(\lambda)f(\lambda)$ onde $W(\lambda) = a^T V_a^{-1} a + d^T V_d^{-1} d$ onde V_a^{-1} e V_d^{-1} são as inversas das variâncias dos efeitos dos QTLs dadas em (10) e (12) que correspondem à matriz de informação do efeito. Esse teste segue uma distribuição qui-quadrado com dois graus de liberdade (Yang e Xu, 2007).

Pleiotropia vs. ligação

O modelo (1) é claramente um modelo pleiotrópico e descreve apenas um QTL afetando simultaneamente dois caracteres, apresentando apenas um QTL por intervalo e matrizes de variâncias e covariâncias definidas. Para testar se em um intervalo existe um QTL afetando simultaneamente dois ou mais caracteres, ou dois QTLs ligados onde cada QTL afeta um caráter por vez, torna-se necessário definir um modelo de ligação. Esse modelo pode ser facilmente adaptado do modelo de pleiotropia assumindo duas posições de QTLs por intervalos e covariância nula ao longo dos caracteres (Liu et al. 2007). O teste de pleiotropia vs. ligação pode ser realizado comparando o modelo original de pleiotropia ($model_{plei.}$) com o modelo de ligação ($model_{liga.}$) utilizando o fator de Bayes (Liu et al. 2007).

$$BF = \left[\frac{p(y | \text{mod}_{pleiotropic})}{p(y | \text{mod}_{linkage})} \right]$$

onde $p(y | \text{mod} \dots)$ é a probabilidade dos dados condicionais a todos os demais parâmetros. Nesse caso, cada amostra é guardada durante o processo MCMC e $p(y | \text{mod} \dots)$ é calculado e posteriormente armazenado, obtendo-se assim uma cadeia de valores de BF. A média harmônica dessa cadeia corresponde à probabilidade das observações dados os parâmetros (Liu et al. 2007). O critério adotado para inferir acerca do modelo mais favorável foi dado pelo log do fator de Bayes considerando: $\text{Log}(\text{BF}) > 10$, decisivo a favor do modelo de pleiotropia; $10 > \text{Log}(\text{BF}) > 5$, forte evidência de pleiotropia, $5 > \text{Log}(\text{BF}) > 0$ moderada evidência de pleiotropia. Valores negativos do $\text{Log}(\text{BF})$ indicam evidência a favor do modelo de ligação considerando a mesma escala descrita para pleiotropia. O teste de pleiotropia foi realizado apenas nas análises com dois caracteres, que equivale à decomposição da análise conjunta com os cinco caracteres.

Distribuição dos efeitos pleiotrópicos

A direção e magnitude dos efeitos pleiotropicos pode ser representada através do uso de um biplot. Esse biplot pode ser construído utilizando a decomposição por valor singular de uma tabela de dupla entrada, com os efeitos dos QTL (significativos no mínimo em um caráter), nos diferentes caracteres avaliados. A tabela centrada nos caracteres e com os efeitos corrigidos para a mesma escala gera um padrão semelhante à análise GGE biplot em estudos de interação genótipos por ambientes (Yan et al., 2001). Nesse gráfico é possível visualizar quais caracteres são fortemente influenciados por grupos de QTLs pleiotrópicos.

Além do biplot, a extensão dos efeitos de pleiotropia e identificação de possíveis módulos foram analisadas por meio de redes hierárquicas bipartidas (Buchanan et al., 2009; Wang et al. 2010). Segundo Buchanan et al. (2009), a análise da distribuição de genes em sistemas ou redes hierárquicas fornece uma maior compreensão da arquitetura genética, removendo a ideia de especificidade gênica e tornando a interpretação biologicamente aceitável. Essas análises foram realizadas utilizando o programa Cytoscape v. 2.8.1 (Shanon et al., 2003).

Resultados

Controle genético da produção e seus componentes

Neste trabalho, foram identificados 27 QTLs atuando sobre os caracteres produção de grãos (PROD), diâmetro da espiga (DE), comprimento da espiga (CE), número de grãos por fileira (NGF) e número de fileiras por espiga (NFI). Na Figura 1 pode-se observar a distribuição desses 27 QTLs ao longo do genoma. Não foram identificados QTLs de grande efeito nos cromossomos 4 e 6, porém, observou-se uma grande saturação ao longo dos cromossomos 8 e 10. Os picos mais agudos sugerem a presença de apenas um QTL na região, ao passo que os picos mais abertos sugerem QTLs ligados, mas com distribuições $g(\lambda)$ não sobrepostas (Figura 1). Na análise de múltiplos intervalos, os 10 grupos de ligação foram pareados simulando um mega-cromossomo com extensos intervalos entre grupos, porém, na Figura 1 a escala está dada em cM e também em função dos grupos de ligação, apenas para facilidade representativa.

No caráter número de grãos por fileira (NGF) foram identificados 13 QTLs, cujo somatório dos efeitos sugere que o caráter possui dominância completa ($GD=0.88$). No entanto, considerando o efeito individual foram encontrados QTLs com efeitos aditivos (6), com dominância parcial (3), dominância completa (1) e sobredominância (4). Grande parte dos QTLs que

atuaram no caráter NGF (85%) apresentou destacado efeito também sobre o comprimento da espiga (CE); sendo então identificados 14 QTLs que, conjuntamente, apresentaram dominância parcial. As interações alélicas - ou grau de dominância - exercidas por esses QTLs sobre NGF e CE foram bastante distintas e, considerando os efeitos específicos sobre CE, foram encontrados sete QTLs aditivos, seis com sobredominância e apenas um apresentando dominância parcial. O traço da distribuição posteriori dos efeitos dos QTLs para NGF e CE podem ser observados na Figura 2.

Os QTLs atuando sobre o número de fileira (10) foram praticamente os mesmos operando no diâmetro da espiga (17), onde se verificou uma coincidência de 90%. Diferentemente do observado para CE e NGF, conjuntamente, pode-se inferir que o controle genético do diâmetro da espiga (DE) e número fileira (NFI) é praticamente aditivo. Não foi encontrado nenhum QTL dominante ou com sobredominância atuando sobre NFI, sendo que 60% dos seus QTLs apresentaram herança aditiva e os demais exibiram dominância parcial. Apenas um QTL com sobredominância foi encontrado atuando sobre DE, e o restante apresentou efeito aditivo (9) e dominância parcial (7). A herança aditiva desses QTLs pode ser observada na Figura 2, onde se constata poucos picos de dominância na cadeia da distribuição posteriori.

O controle genético da produção de grãos (PROD) apresentou dominância parcial quando considerou-se o somatório dos 17 QTLs que atuaram sobre esse caráter. De forma genérica, a herança desse caráter foi equivalente à mistura dos efeitos dos QTLs presentes nos seus componentes (Figura 2). Nessa figura é possível observar que o traço dos efeitos aditivos referente à produção de grãos foi semelhante aos observados para DE e NFI, e que o traço dos seus efeitos de dominância foi semelhante aos observados nos caracteres CE e NGF. Dentro dos 17 QTLs que influenciam a produção de grãos, houve uma distribuição quase equitativa em números de locos de efeito aditivo (6), de

dominância parcial (6) e sobredominância (5). Esses resultados indicam que o caráter produção de grãos é resultado do produto das interações alélicas dos seus componentes.

Pleiotropia e sua distribuição

A maior parte dos QTLs identificados nesse trabalho (63%) apresentou efeito pleiotrópico. Esses QTLs pleiotrópicos também foram os mais importantes no controle genético dos cinco caracteres analisados (Figura 1); observa-se que as regiões do genoma que apresentaram os maiores valores de $W(\lambda)$ demonstraram também mais de um pico acima do valor crítico. Os QTLs mais importantes foram: o QTL55 - localizado a 110.8 cM do início do cromossomo 3 - que atuou nos cinco caracteres; o QTL 175 - localizado a 140.5 cM do início do cromossomo 10 - que atuou em quatro caracteres e o QTL 141 - localizado a 58.8 cM do início do cromossomo 8 - e que atuou nos cinco caracteres.

As análises comparativas entre os modelos de ligação e modelos de pleiotropia revelaram, em todas as situações, uma forte evidência em favor do modelo de pleiotrópico (Tabela 2). A redução da variância residual no modelo de pleiotropia chegou a 4 vezes quando se analisou conjuntamente os caracteres CE e DE em relação ao modelo de ligação. Os resultados da Tabela 2 demonstram que é pouco provável que dois QTLs muito próximos atuem de forma indireta sobre os caracteres, onde sua covariância seria apenas função da frequência de recombinação.

Na Figura 2 pode-se observar uma forte sobreposição do traço das distribuições posteriori entre CE e NGF o que sugere que não há distinção entre esses caracteres, ou seja, são duas medidas de um mesmo caráter. Da mesma forma, não houve distinção entre DE e NFI, tanto na posição quanto na direção dos efeitos dos QTLs (Figura 2). Também é evidente que a produção de grãos é

função quase que exclusivamente desses quatro componentes, o que pode sugerir, a priori, que não existem QTLs “para” a produção de grãos, mas que esta é função dos QTLs que atuam em seus sub-caracteres.

A forte evidência do modelo de pleiotropia também pode ser observada na Figura 3. Nessa figura fica manifestado que o modelo de pleiotropia remove os efeitos de QTLs muito próximos e concentra esses efeitos em apenas uma região do genoma e que essa característica do modelo converge em menor variância residual (Tabela 2). Por exemplo, na análise de ligação bi-caracteres, que considerou a PROD e NFI, observou-se que o primeiro grupo de ligação apresentou dois picos em intervalos vizinhos, mas que posteriormente foram corrigidos para dois picos dentro do mesmo intervalo no modelo de pleiotropia (Figura 3). A mesma situação foi observada para os caracteres PROD e NGF, onde quase todos os picos vizinhos no modelo de ligação foram alocados em apenas uma região do genoma. Esse padrão se repetiu em todas as análises comparativas entre o modelo de ligação e modelo de pleiotropia.

Os resultados deste estudo sugerem que a produção de grãos não possui módulos de genes pleiotrópicos, ou seja, não foi possível separar os QTLs de acordo com sua grandeza e amplitude de atuação (Figura 4B). Embora os QTLs 168 e 169 - ambos atuando sobre DE e NFI e localizados no grupo 10 - ficassem fora do módulo devido a sua baixa importância relativa, nenhum caráter pôde ser separado de acordo com grupos específicos de QTLs. No módulo encontrado, os QTLs de maior importância ficaram mais próximos aos cinco caracteres e aqueles de menor importância foram alocados à margem do módulo. Esse agrupamento considera tanto a amplitude de atuação do QTL quanto suas magnitudes. Na Figura 4A podemos observar uma rede hierárquica onde os QTLs foram agrupados de acordo com a magnitude de influência exercida sobre os caracteres - mas não em amplitude - e os caracteres agrupados de acordo com a influência recebida por esses QTLs. Nesse caso, os QTLs mais próximos aos

caracteres são aqueles de menor influência média e aqueles mais distantes são os QTLs de maior influência, não importando sua amplitude de atuação. Esse sistema hierárquico também sugere que o controle genético desses cinco caracteres é extremamente complexo e que a pleiotropia é amplamente dispersa. Também sugere que CE e NGF são altamente correlacionados assim como DE e NFI.

A alta correlação entre os cinco caracteres também pode ser observada através do uso de biplots onde os efeitos padronizados dos 27 QTLs sobre os cinco caracteres foram decompostos em componentes principais. Na Figura 5 fica evidente que um grupo de QTLs exerce forte influência aditiva positiva sobre DE e NFI e outro grupo de QTLs apresenta forte influência positiva sobre CE e NGF. Além disso, observa-se que esses QTLs pleiotrópicos apresentam efeitos recíprocos de sinais opostos dentro desses caracteres. Os QTLs de baixa pleiotropia aditiva se encontraram próximos às coordenadas (0,0) no biplot.

Na Figura 5 também é possível estimar a correlação genética aditiva entre os caracteres, que é dada pelo cosseno do menor ângulo entre os vetores de CE, DE, PROD, NFI e NGF. Os valores de correlação estimados pelo biplot foram bem representativos dos valores obtidos via componentes de variância. Por exemplo, a correlação entre DE e NFI no biplot foi de 0.982 e de 0.63 via componentes da variância (CV). A correlação entre CE e NGF foi de 0.72 no biplot e de 0.46 via CV. O caráter com maior correlação aditiva com PROD foi o diâmetro da espiga (0.89 via biplot e 0.52 via CV) e com menor correlação (em escala) foi o comprimento da espiga (-0.50 via biplot e 0.12 via CV). Esse resultado sugere que o biplot aditivo ilustra de forma representativa a importância da pleiotropia na determinação da correlação genética aditiva entre esses caracteres, ou seja, os 17 QTLs pleiotrópicos encontrados foram fortemente descritivos da correlação genética estimada via métodos convencionais por uso de componentes da variância.

Na Figura 5 também podemos observar o efeito de dominância dos 27 QTLs e sua importância no controle genético da produção de grãos e seus componentes. No biplot aditivo a produção de grãos foi fortemente influenciada pelos caracteres DE e NFI, entretanto, considerando o biplot dos efeitos de dominância, a produção de grãos foi fortemente influenciada pelos caracteres CE e NGF. Além disso, observou-se que os QTLs com maior relevância de dominância foram ligeiramente distintos dos QTLs com maior relevância aditiva e esse resultado sugere que: (i) a linhagem com maior produção *per se* possui espigas menores, porém com maior diâmetro e conseqüentemente maior número de fileiras; (ii) no híbrido, a interação alélica e a complementação desses QTLs pleiotrópicos permitem que, ao mesmo tempo em que se obtenha QTLs aditivos aumentando o diâmetro da espiga e número de fileiras, também se permite que outros QTLs com efeito de dominância aumentem o comprimento da espiga e número de grãos por fileira. Esse fato pode compensar aqueles QTLs pleiotrópicos que apresentam diferentes sinais de efeito aditivo ao longo dos caracteres como observado na Figura 2, onde o QTL 55 - localizado no grupo de ligação 3 - apresenta forte sinal positivo sobre NGF e CE e sinal negativo sobre DE e NFI. Dessa forma, a pleiotropia das interações alélicas pode minimizar os possíveis prejuízos de correlações negativas provenientes da pleiotropia do QTL, visto que os sinais dessas interações nem sempre são os mesmos e o grau de dominância que os QTLs pleiotrópicos exercem sobre determinados caracteres também é muito variado.

Discussão

A complexidade da herança dos caracteres quantitativos tem ganhado destacada atenção em recentes revisões sobre o assunto (Mackay, 2001; Flint e Mackay, 2009; Mackay et al., 2009; Buchanan et al., 2009; Wagner e Zhang, 2011). Em todos esses trabalhos tem-se evidenciado que o “custo da

complexidade” de um organismo está diretamente ligado à distribuição da pleiotropia no genoma. Nossos resultados sugerem que o controle genético da produção de grãos e seus componentes é indissociável na amplitude de atuação dos QTLs e que a pleiotropia exerce papel fundamental na compreensão da arquitetura genética desses caracteres. Em todas as análises realizadas foi difícil separar o controle genético destes cinco caracteres com base na atuação dos QTLs (Figura 1, 2 3 e 4). Na Figura 2 fica evidente que a produção de grãos surge da interação dos QTLs presentes em DE, CE, NFI, e NGF o que pode levar à interpretação de um modelo multiplicativo atuando sobre esses caracteres (Schnell e Cockerham, 1992).

Schnell e Cockerham (1992) avaliaram a possibilidade de se entender a heterose como produto de uma ação multiplicativa entre subcaracteres. À medida que o traço da distribuição posteriori referente à produção de grãos é claramente sobreposto pelos traços aditivos de DE e NFI e pelos traços de dominância de CE e NGF (Figura 2), e considerando que os QTLs pleiotrópicos que atuam sobre esses caracteres possuem diferentes interações alélicas (Figura 5), podemos prever o controle genético da produção de grãos por meio da análise dos QTLs presentes nos seus componentes (Schnell e Cockerham, 1992). Por exemplo, se considerarmos o QTL 20 (cromossomo 1- primeiro pico em CE e NGF), podemos observar sua especificidade aditiva sobre CE e NGF, mas efeito nulo em DE e NFI (Figura 2). Assim, sobre o modelo multiplicativo de Schnell e Cockerham (1992), a predição deste QTL apresentaria também efeito nulo sobre o caráter produção de grãos como observado na Figura 2, ou seja, o seu efeito predito seria

$$QTL_{20} = \mu_{el}\mu_{ed} + \alpha_{el}\mu_{ed} + \alpha_{el}\mu_{ed} + d_{el}\mu_{ed} + d_{el}\mu_{ed} + \alpha_{el}\alpha_{ed} + d_{el}d_{ed},$$

onde: $\mu_{ed} = 0$, $\alpha_{ed} = 0$, $d_{ed} = 0$ e $d_{ed} = 0$. A mesma predição pode ser feita para o comportamento do QTL 145 (segundo pico do cromossomo 8) que apresentou efeito de sobredominância sobre CE e NGF e efeito aditivo sobre DE

e NFI, mas exercendo efeito de dominância completa na produção de grãos. Assim é possível prever o comportamento da maioria dos QTLs pleiotrópicos atuando na produção de grãos - mas não o comportamento de QTLs específicos “para” produção de grãos. Embora fossem encontrados 4 QTLs específicos atuando sobre a produção de grãos, não significa que estes não estejam relacionados a outros caracteres não avaliados nesse estudo.

Na universalidade dos efeitos de pleiotropia especula-se um entrave na adaptação do indivíduo devido ao “custo da complexidade” ou “custo pleiotrópico”, onde a rede adaptativa pode ser descrita em um espaço n-dimensional delineada pelo Modelo Geométrico de Fisher (Wagner e Zhang , 2011). Segundo Wagner e Zhang (2011) e vários outros autores (Wagner et al., 2008; Su et al., 2010, Stearns, 2010), a pleiotropia modular seria favorável do ponto de vista adaptativo/evolutivo, ao passo que a pleiotropia universal seria um entrave ao processo. Sob a ótica do melhoramento genético, a divisão da produção de grãos em módulos de genes pleiotrópicos, de certo modo, facilitaria o melhoramento, pois o ajuste entre módulos para chegar ao genótipo ideal seria mais fácil do que o ajuste de vários QTLs pleiotrópicos em um único módulo; seria como montar um complexo quebra-cabeça com partes pré-montadas comparando a montagem peça a peça. No entanto, nossos resultados demonstram que mesmo ocorrendo uma pleiotropia amplamente distribuída entre os QTLs (mas não universal devido ao número de caracteres avaliados), a interação alélica desses QTLs não necessariamente apresentou forte pleiotropia (Figura 5). Nessa figura, podemos observar que os QTLs com forte efeito pleiotrópico aditivo (QTL 55, QTL 37, QTL 36, QTL 104, QTL 141, QTL 145 e QTL 140), não necessariamente apresentam forte pleiotropia de dominância (QTL 175, QTL 172, QTL 145, QTL 141, QTL 8, QTL 176, QTL 174 e QTL 126). Além disso, verifica-se que o QTL 145 apresentou forte influência aditiva sobre DE e NFI e também expressivo controle de dominância sobre CE e NGF,

sugerindo que QTLs pleiotrópicos podem demonstrar diferentes interações alélicas ao longo dos caracteres.

Esses resultados sugerem que, mesmo sob uma pleiotropia amplamente dispersa, é possível obter híbridos altamente produtivos ou genótipos amplamente adaptados devido às diferentes extensões de pleiotropia exercida pelas interações alélicas nos QTLs. Essas observações, de certa forma, são similares às descritas por Carbone et al. (2006), que trabalhando com o gene pleiotrópico *Catsup* em *Drosophila*, verificou que para alguns caracteres os sítios desse gene, descritos pelos seus QTNs, atuavam de forma independente. Sendo assim, mutações no gene *Catsup* não necessariamente agiriam de forma universal em todos os caracteres, pois o efeito alélico do gene apresentou diferentes extensões de pleiotropia.

Devemos também destacar que em estudos de mapeamento por desequilíbrio de ligação dificilmente podemos inferir sobre genes pleiotrópicos visto que um QTL pode conter vários genes (Mackay et al., 2009). Além disso, deve-se enfatizar que nosso modelo não considerou dois QTLs pleiotrópicos por intervalo e sim dois QTLs com efeitos independentes por intervalo. Neste caso, nossa inferência sobre a evidência a favor de um modelo de pleiotropia se restringe à menor probabilidade de se observar QTLs muito próximos e com efeitos independentes sobre os caracteres.

Nesse estudo pode-se verificar a grande capacidade da análise bayesiana de múltiplos intervalos em descrever a arquitetura genética de caracteres quantitativos, como já observado também por Xu (2003), Wang et al. (2005) e Xu et al. (2009). Contudo, alguns autores têm sugerido outras abordagens para o mapeamento de múltiplos caracteres (Jiang e Zeng, 1995; Weller et al., 1996; Gilbert e Leroy, 2004; Masoletti et al., 2008). Embora algumas dessas metodologias apresentem grande vantagem computacional, não são capazes de extrair a quantidade de informações contidas nos dados a respeito dos

parâmetros, pois são restritas em lidar com modelos superparametrizados e por isso lançam mão da técnica de seleção de modelos.

Recentes estudos têm sugerido a modelagem mista de múltiplos caracteres e múltiplos ambientes em um único modelo (Masoletti et al., 2008; Van Eeuwijk et al., 2010). Esses autores consideram o efeito de QTL como fixo e os efeitos marginais das marcas como aleatórios (mapeamento por intervalo composto). Embora a modelagem conjunta de múltiplos ambientes e múltiplos caracteres seja desejável, a utilização de análise por intervalo simples pode contrabalancear esse possível ganho. Além disso, nessas abordagens, a covariância do QTL é confundida com a covariância residual, tanto em direção quanto em magnitude, não explorando - de forma direta - as (co)variâncias específicas de cada QTL ao longo dos caracteres. Fora isto, no nosso entendimento, conceitualmente tomar um efeito de QTL como fixo aparentemente contrasta com toda a incerteza presente na arquitetura genética de um caráter (Buchanan et al. 2009).

Recentes estudos tentam descrever o controle genético da heterose utilizando o mapeamento em caracteres complexos em milho (Lu, et al., 2003, Frascaroli et al. 2007, Garcia et al. 2008; Shon et al. 2010). A maioria desses trabalhos tem observado forte efeito de dominância/sobredominância no controle genético da heterose para produção de grãos e efeito marginal de epistasia. Nossos resultados sugerem que, conjuntamente, o controle da heterose apresenta dominância parcial, porém com distribuição equitativa de QTLs com efeitos aditivos, com dominância parcial e sobredominância. Embora a epistasia não tenha sido investigada neste trabalho, observa-se, através da literatura, que esse efeito tem sido marginal no controle genético da heterose (Garcia et al. 2008; Shon et al. 2010) e, a partir desses resultados, duas questões ficam em aberto: (i) o ajuste dos efeitos epistáticos apenas em QTLs de grande efeito pode não desenhar a real contribuição desse fenômeno no controle genético da heterose,

requerendo dessa maneira a aplicação de modelos complexos para o estudo da distribuição da epistasia ao longo do genoma; (ii) a epistasia pode não ser caráter-específica e poderia ser melhor estudada dentro do conceito da multiplicidade de QTLs pleiotrópicos uma vez que esses dois fenômenos (pleiotropia e epistasia) podem ser indissociáveis (Schnell e Cockerham, 1992; Cheverud, et al., 2004; Wolf et al., 2005; Wolf et al., 2006; Pavlicev et al., 2007; Mackay et al., 2009).

Nossos resultados reforçam as discussões levantadas por Mackay (2001), Mackay et al. (2009) e Buchanan et al. (2009), e sugerem que a herança dos caracteres é descrita de forma limitada se analisada isoladamente, porém pode ser melhor compreendida através de redes pleiotrópicas que permitam visualizar a complexidade da herança de caracteres quantitativos e, a partir disto, traçar novas estratégias de seleção. Também foi possível confrontar a ideia de que é possível identificar QTLs “para” caracteres complexos tais como a produção de grãos, visto que a pleiotropia atua de forma destacada em seus subcaracteres e que esse “caráter” pode ser decomposto e predito quase que na sua totalidade pelos QTLs dos seus componentes. Além disso, a pleiotropia de QTLs não necessariamente significa pleiotropia das interações alélicas e sendo assim, a ampla distribuição dos efeitos pleiotrópicos não restringe os ganhos com o melhoramento genético.

Agradecimentos

À professora Anete Pereira de Souza do Departamento de Genética e Evolução da Universidade de Campinas (UNICAMP), pela genotipagem das populações, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo financiamento da pesquisa.

REFERÊNCIAS

BUCHANAN, A.V.; SHOLTIS, S.; RICHTSMEIER, J; WEISS, K.M. What are genes “for” or where are traits “from”? What is the question? **Bioessays**, Hoboken, v. 31, n. 2, p. 198–208, Feb. (2009).

CARBONE, M.A.; JORDAN, K.W.; LYMAN, R.F.; HARBISON, S.T.; LEIPS, J.; MORGAN, T.J.; DeLUCA, M.; AWADALLA, P.; MACKAY, T.F. Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. **Current Biology**. Cambridge v. 16, n. 9, p. 912-919, May 2006.

CHEVERUD, J.; EHRICH, T.H.; VAUGHN, T.T.; KOREISHI, S.F.; LINSEY, R.B.; PLETSCHER, L.S. Pleiotropic effects on mandibular morphology II: differential epistasis and genetic variation in morphological integration. **Journal of Experimental Zoology**, Hoboken, v.302, n. 5, p.424–435. Sep. 2004

FLINT, J.; MACKAY, T.F.C. Genetic architecture of quantitative traits in mice, flies, and humans. **Genome Research**, Woodbury, v.19, n. 5, p.723– 73, May. 2009

FRASCAROLI, E.; CANÈ, M.A.; LANDI, P.; PEA, G.; GIANFRANCESCHI, L.; VILLA, M.;MORGANTE, M.; PÈ M.E. Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. **Genetics**, Baltimore, v. 176, n. 1, p. 625–644. May. 2007.

GARCIA, A.A; WANG, S.; MELCHINGER, A.E.; ZENG, Z.B.; **Quantitative trait mapping and the genetic basis of heterosis in maize and rice**. **Genetics**, Baltimore, v.180, n.3, p. 1707 -1724, Nov. 2008.

GILBERT, H.; LE ROY, P. Comparison of three multitrait methods for QTL detection. **Genetics, Selection, Evolution**, London, v.35, n. 3, p. 281–304, May. 2003

GOMMANS, W.M.; MULLEN, S.P.; MAAS, S. RNA editing: a driving force for adaptive evolution? **Bioessays**. Hoboken, v. 31, n. 10, p. 1137-1145, Oct. 2009.

GRUNEBERG, H. An analysis of the "pleiotropic" effects of a new lethal mutation in the rat (*mus norvegicus*). **Proceeding of the Royal Society**. London, v.125, n.3, p. 123-144, Mar.1938.

HASTINGS, W.K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, Feb. 1970.

HUBERTS, D.H.; VAN DER KLEI, I.J. Moonlighting proteins: an intriguing mode of multitasking. **Biochimica et Biophysica Acta**, Amsterdam, v.1803, n.4, p.520-525, Apr. 2010.

JIANG, C.; ZENG, Z. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Baltimore, v. 140, n. 3, p. 1111-1127, jul. 1995.

LINCOLN, S.E.; DALY, M.J.; LANDER, E.S. **Constructing genetic maps with Mapmaker Exp 3.0**. 3rd ed. Cambridge: Whitehead Institute for Biometrical Research, 1992. 230p.

LIU, J.; LIU, Y.; LIU, X; DENG, H.W. Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. **American Journal of Human Genetics**, Cambridge, v. 81, n. 2, p. 304-320, Aug. 2007.

LU, H.; ROMERO-SEVERSON, J.; BERNARDO, R. Genetics basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. **Theoretical and Applied Genetics**, Berlin, v. 107, n.3,p. 494-502, Ago. 2003.

MACKAY, T.F.C. The genetic architecture of quantitative traits. **Annual Review of Genetics**. Palo Alto, v.35, p. 336-339, 2001.

MACKAY, T.F.; STONE, E.A.; AYROLES, J.F. The genetics of quantitative traits: challenges and prospects. **Nature Reviews. Genetics**, London, v. 10, n. 8, p. 565-577, Aug. 2009.

MALOSETTI, M.; RIBAUT, J.M.; VARGAS, M.; CROSSA, J.; VAN EEUWIJK, F.A. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). **Euphytica**, Dordrecht, v. 161, n.1, p.241–257, 2008.

MAYDANOVYCH, O.; BEAL, P.A. Breaking the central dogma by RNA editing. **Chemical Reviews**. Washington, v. 106, n. 8, p. 3397-3411, Aug. 2006.

MELCHINGER, A. E.; UTZ, H. F.; PIEPHO, H.P.; ZENG, Z.B.; SCHÖN, C. C. The role of epistasis in the manifestation of heterosis: A systems-oriented approach. **Genetics**, Baltimore, v.177, n. 3 p. 1815–1825, Nov. 2007.

METROPOLIS, N.; ROSEMBLUT, A.W.; ROSEMBLUT, M.N.; TELLER, A.H.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics, New York**, v. 21, 1087-1092, 1953.

PAVLICEV, M.; KENNEY-HUNT, J.P.; NORGARD, E.A.; ROSEMAN, C.C.; WOLF, J.B.; CHEVERUD, J.M. Genetic variation in pleiotropy: differential epistasis as a source of variation in the allometric relationship between long bone lengths and body weight. **Evolution**. Malden, v. 62, n. 1, p. 199-213, Jan. 2008.

SCHÖN, C.C.; DHILLON, B.S.; UTZ, H.F.; MELCHINGER, A.E. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 321-332, Jan. 2010.

SCHNELL, F.W.; COCKERHAM, C.C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics**, Baltimore, v. 131, n. 2, p. 461-469, Jun.1992.

SIBOV, S.T.; SOUZA JÚNIOR, C.L.; GARCIA, A.A.F.; GARCIA, A.F.; SILVA, A.R.; MANGOLIN, C.A.; BENCHIMOL, L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and

localization of loci showing distorted segregation. **Hereditas**, Oxford, v. 139, p. 96-106, 2003.

SILVA, A.R. Análise genética de caracteres quantitativos em milho com o delineamento III e marcadores moleculares. Piracicaba, 2002. 143p. Tese (Doutorado) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N.S.; WANG, J.T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Research**. New York, v.13, n.11, p. 2498-2504, Nov. 2003.

STEARNS, F.W. One hundred years of pleiotropy: a retrospective. **Genetics**, Baltimore, v. 186, n.3, p.767-773, Nov. 2010.

SU, Z.; ZENG, Y.; GU, X. A preliminary analysis of gene pleiotropy estimated from protein sequences. **Journal of Experimental Zoology**, Hoboken, v. 314, n. 2, p. 115-122, Mar. 2010

VAN EEUWIJK, F.A.; BOER, M.; TOTIR, R. L.; BINK, M.; WRIGHT, D.; WINKLER, C, R.; PODLICH, D.; BOLDMAN, K.; BAUMGARTEN A.; SMALLEY M.; ARBELBIDE, M.; BRAAK, C.J.F.T.; COOPER, MARK. Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. **Theoretical and Applied Genetics**, Berlin, v. 120, n. 2, p. 429-440, Jan. 2010.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Baltimore, v. 163, n.2, p. 789–801, Feb. 2003.

XU, C.; WANG, X.; LI, Z.; XU, S.; Mapping QTL for multiple traits using Bayesian statistics. **Genetics Research**, New York, v. 91, n. 1, p. 23-37, Feb. 2009.

WAGNER, G.P.; KENNEY-HUNT, J.P.; PAVLICEV, M.; PECK, J.R.; WAXMAN, D.; CHEVERUD JM. Pleiotropic scaling of gene effects and the 'cost of complexity'. **Nature**, London, v.27, n. 7186, p. 470-472. Mar. 2008.

WAGNER, G.P.; ZHANG, J. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. **Nature Reviews, Genetics**, London, v.12, p. 204-213, Mar. 2011.

WANG, H.; ZHANG, Y.M.; LI, X.; MASINDE, G.L.; MOHAN, S.; BAYLINK, D.J.; XU, S. Bayesian shrinkage estimation of quantitative trait loci parameters. **Genetics**, Baltimore, v. 170, n. 1, p. 465-480, May 2005.

WELLER, J.I.; WIGGANS, G.R.; VANRADEN, P.M.; RON, M. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment. **Theoretical Applied Genetics**, Berlin, v. 92, n.4, p. 998–1002, Oct.1996

WOLF, J. B.; LEAMY, L.L.; ROUTMAN, E. J.; CHEVERUD, J. M. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. **Genetics**, Baltimore, v. 171, n.2, p.1–12, Oct. 2005.

WOLF, J. B.; POMP, D.; EISEN, E. J.; CHEVERUD, J. M.; LEAMY, L. L. The contribution of epistatic pleiotropy to the genetic architecture of covariation among polygenic traits in mice. **Evolution Development**, Malden, v.8, n.5, p. 468–476, Oct. 2006.

WRIGHT, S. **Evolution and the genetics of populations**. Chicago, University of Chicago Press. 1968. 480 p.

YAN, W.; HUNT, L. A. Interpretation of Genotype x Environment Interaction for Winter Wheat Yield in Ontario. **Crop Science**.Madison, v. 41, n.1 p. 19-25, Jan-Feb. 200.

YANG, R.; XU, S. Bayesian Shrinkage Analysis of Quantitative Trait Loci for Dynamic Traits. **Genetics**, Baltimore, v.2, n.176, p. 1169–1175. Jul. 2007.

YANG, R.C. Epistasis of quantitative trait loci under different gene action models. **Genetics**, Baltimore, v.167, n.3, p. 1493–1505. Jul. 2004.

ZENG, Z.B.; WANG, T.; ZOU W. Modeling quantitative trait loci and interpretation of models. **Genetics**, Baltimore, v.169, n.3, p.1711-1725, Mar. 2005.

Tabela 1 – Probabilidade condicional de segregação do QTL (Q) dado a marca adjacente (M). A probabilidade condicional do intervalo é dada pelo produto de Kronecker das matrizes de retrocruzamentos ($RC_1 \otimes RC_1$ e $RC_2 \otimes RC_2$)

	RC_1			RC_2		
	Q_1Q_1	Q_1Q_2	Q_2Q_2	Q_1Q_1	Q_1Q_2	Q_2Q_2
M_1M_1	$(1-r)$	r	0	0	$(1-r)$	r
M_1M_2	$1/2$	$1/2$	0	0	$1/2$	$1/2$
M_2M_2	r	$(1-r)$	0	0	r	$(1-r)$

Tabela 2 – Evidência a favor do modelo de pleiotropia obtida pelo logaritmo do fator de Bayes $\ln(\text{BF})$ contra o modelo de ligação.

	DE	CE	PRO	NFI	NGF
DE		615.61	550.93	453.54	562.22
CE			424.21	453.54	440.32
PRO				362.08	552.91
NFI					557.00

DE- Diâmetro da espiga

CE- Comprimento da espiga

PRO- Produção de grãos

NFI- Número de fileiras

NGF- Número de grãos por fileira

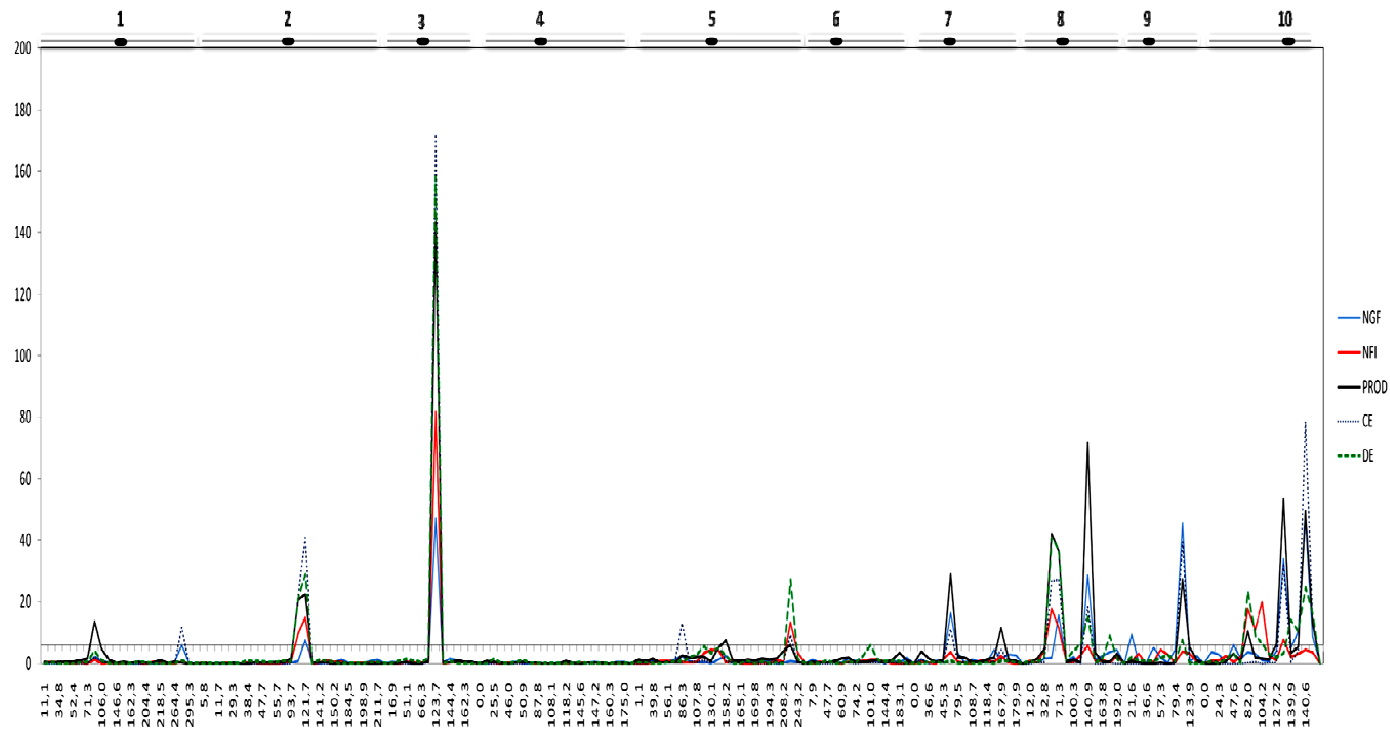


Figura 1- Perfil da localização de QTLs no genoma responsáveis pelo caráter produção de grãos (PROD) em milho e seus componentes definidos por número de grãos fileira (NGF), número de fileiras (NFI), comprimento da espiga (CE) e diâmetro de espiga (DE). A linha horizontal pontilhada representa o valor crítico para o teste univariado de Wald ($\chi^2_{2,1-0.05} = 5.99$).

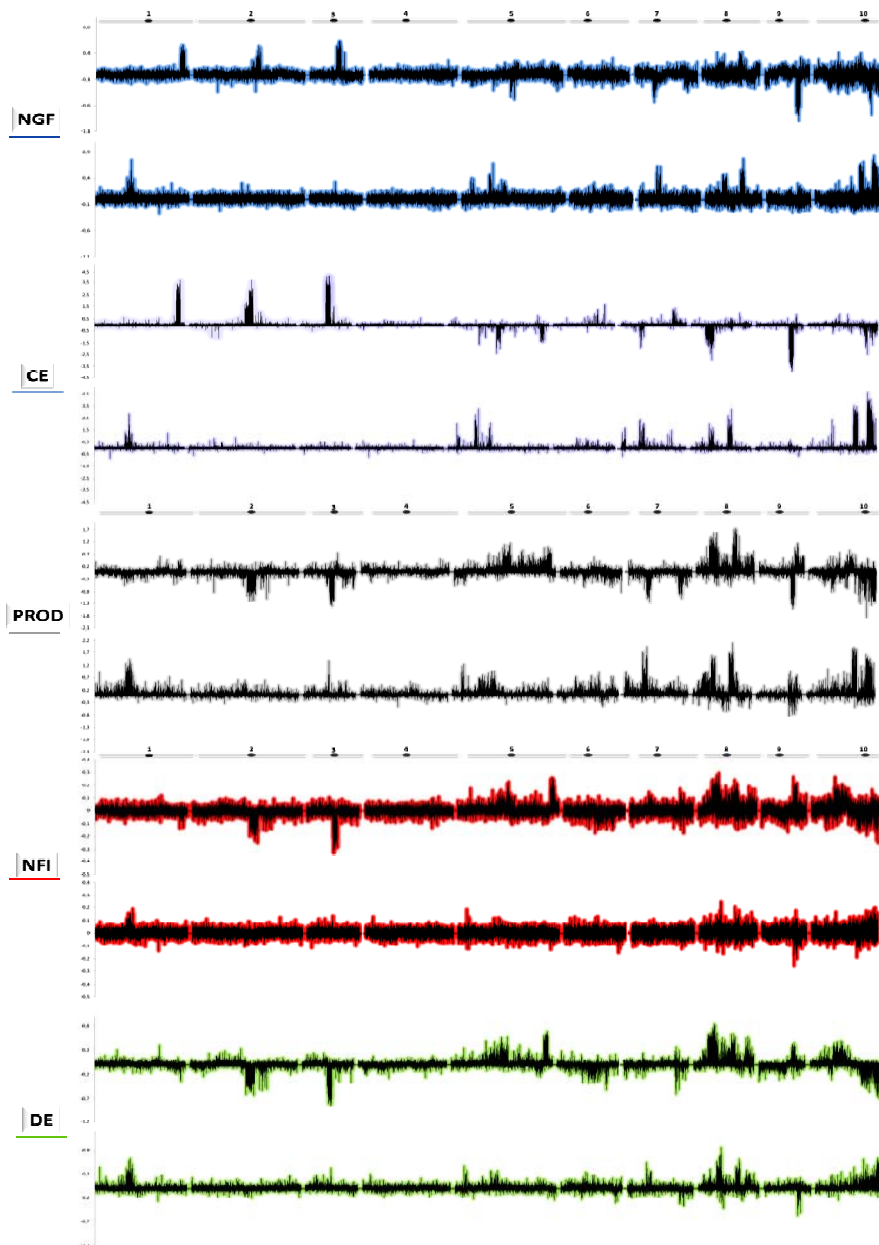


Figura 2- Traços da distribuição posteriori referentes aos efeitos aditivos (traço superior) e dominantes (traço inferior) considerando a produção de grãos (PROD) em milho e seus componentes definidos por número de grãos fileira (NGF), número de fileiras (NFI), comprimento da espiga (CE) e diâmetro da espiga(DE)..

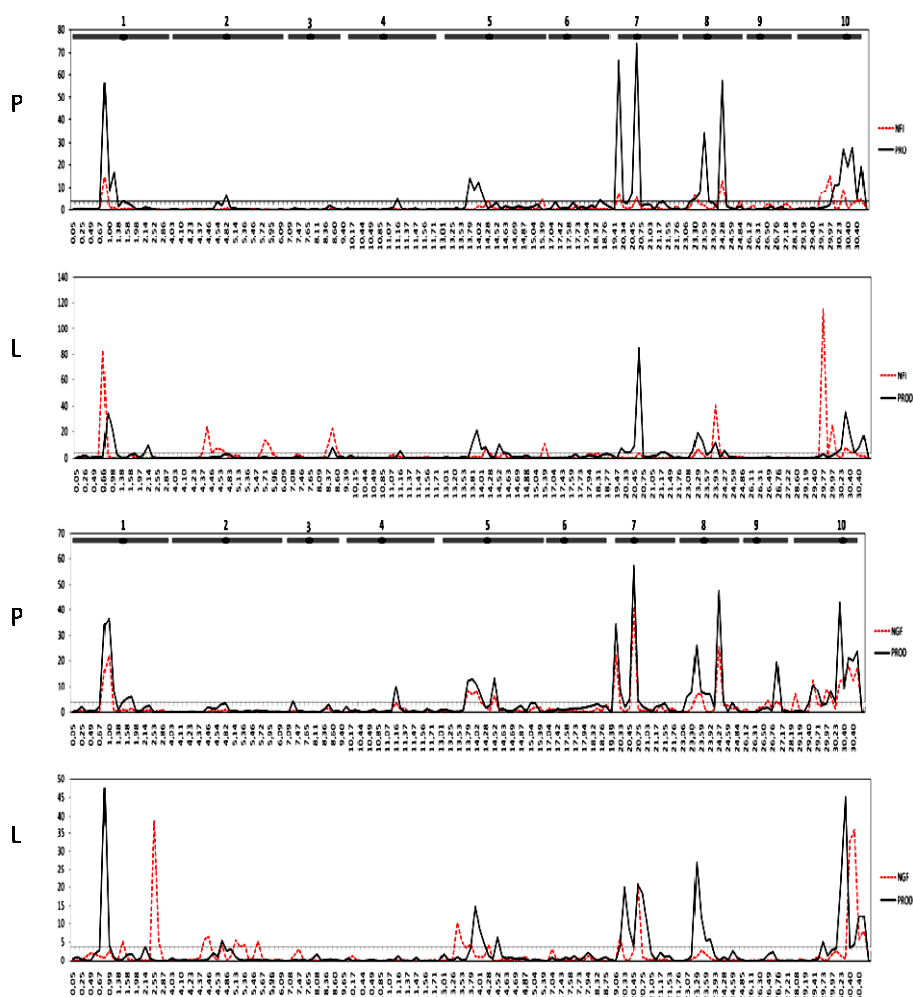


Figura 3- Perfil da localização de QTL no genoma do milho aplicando o modelo pleiotrópico (P) e o modelo de ligação (L) para os caracteres produção de grãos (PRO), número de fileiras por espiga (NFI) e número de grãos por fileira (NGF).

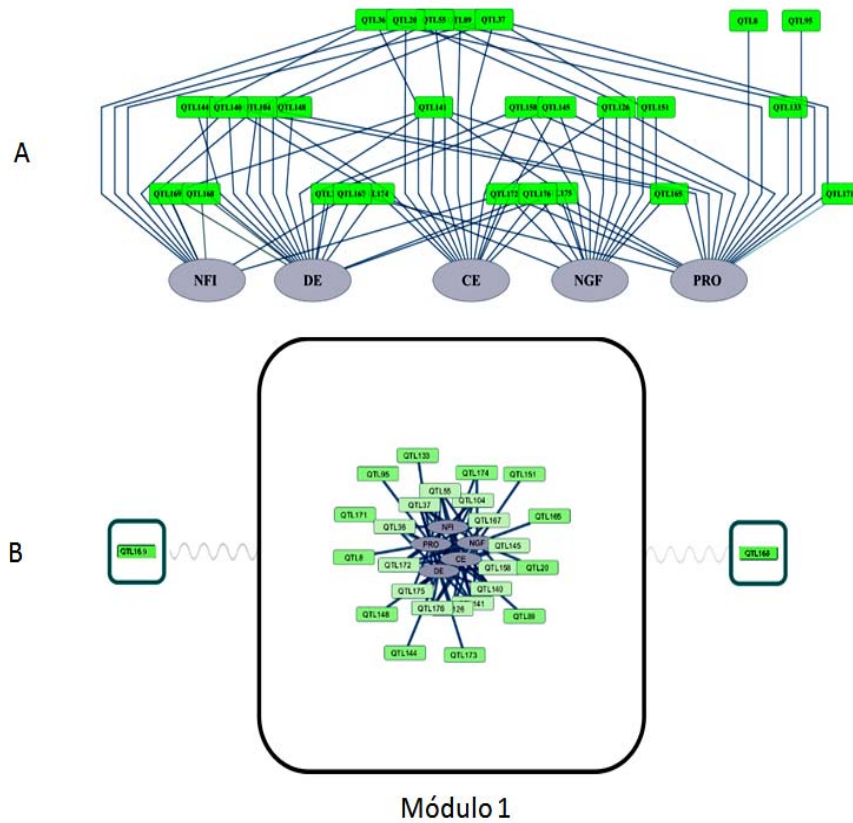


Figura 4- Rede pleiotrópica hierárquica (A) e análise modular (B) dos QTLs atuantes na produção de grãos (PROD) em milho e seus componentes definidos por número de grãos fileira (NGF), número de fileiras (NFI), comprimento da espiga (CE) e diâmetro da espiga (DE).

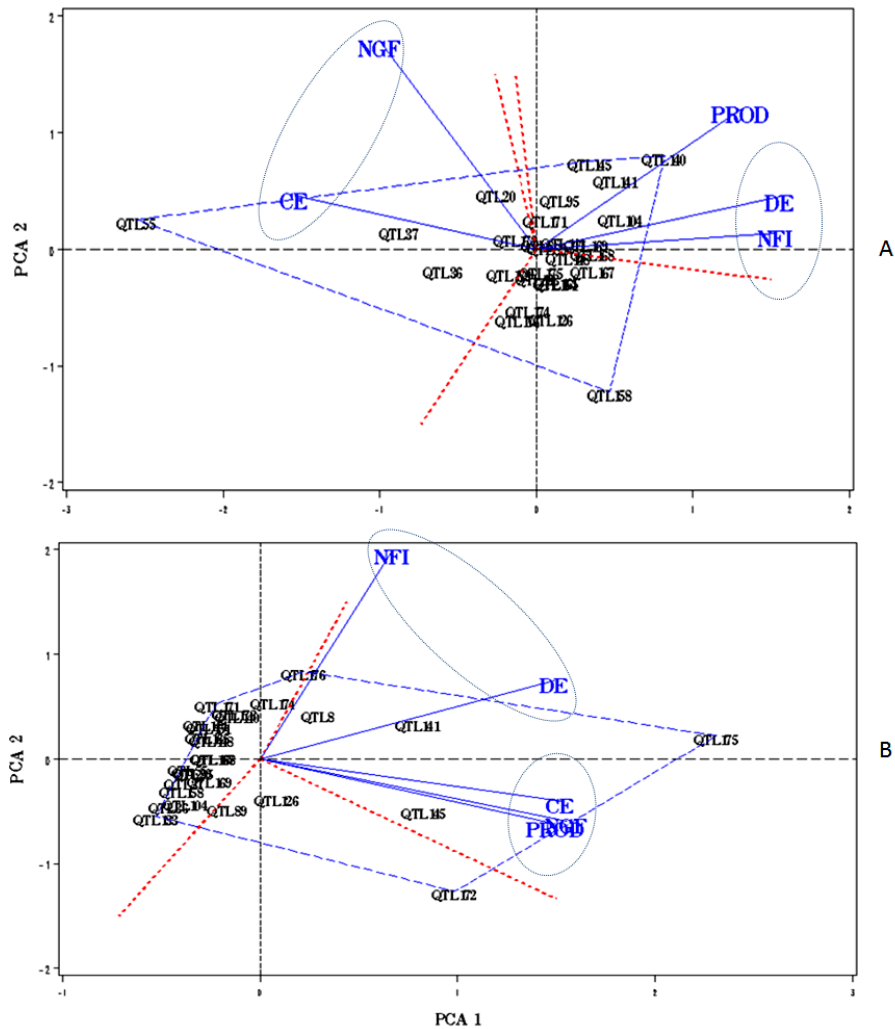


Figura 5- Análise de componentes principais dos efeitos aditivos (A) e dominantes (B) dos QTLs sobre a produção de grãos (PROD) em milho e seus componentes definidos por número de grãos fileira (NGF), número de fileiras (NFI), comprimento da espiga (CE) e diâmetro da espiga (DE).