

MAPEAMENTO DE QTLS : UMA ABORDAGEM BAYESIANA

Elisabeth Regina de TOLEDO¹
Roseli Aparecida LEANDRO¹
Cláudio Lopes de SOUZA JUNIOR²
Anete Pereira de SOUZA³

- RESUMO: Caracteres vegetais e animais de importância econômica, em sua maioria, podem ser classificados como quantitativos. Caracteres desse tipo são aqueles cuja expressão fenotípica apresenta uma variação contínua, atribuída à segregação simultânea de muitos genes distribuídos pelo genoma, em regiões definidas como QTLs (“Quantitative Trait Loci”). Mapear um QTL significa identificar sua posição no genoma e estimar seus efeitos. Existem, na literatura, vários métodos de mapeamento de QTLs, grande parte deles apresenta uma abordagem clássica. Neste trabalho apresenta-se o mapeamento de QTLs utilizando-se a abordagem Bayesiana. Utilizando-se métodos Bayesianos é possível incorporar a incerteza sobre as quantidades desconhecidas. Nesse trabalho a incerteza com relação ao número de QTLs, também, será incorporada. O mapeamento de QTLs será realizado considerando-se um conjunto de dados de produção de grãos de milho utilizando-se o software QTLCartographer versão 2.5 para Windows.
- PALAVRAS-CHAVE: QTLs; abordagem Bayesiana; fator de Bayes; MCMC com saltos reversíveis.

1 Introdução

Quantitative trait loci designados por QTLs são regiões do genoma responsáveis pela expressão de caracteres fenotípicos, que possuem distribuição contínua, tais como, altura e peso de plantas e de animais; produção de grãos; teor de óleo etc. Com o advento dos marcadores moleculares (Lander; Bolstein, 1989) tornou-se possível mapear regiões cromossômicas (QTLs) que afetam esses caracteres quantitativos.

Mapear um QTL significa identificar sua posição no genoma e estimar seus efeitos genéticos, tais como: o efeito aditivo, efeito de dominância e outros efeitos presentes no modelo adotado. Para realizar o mapeamento são necessárias informações tais como: caractere quantitativo de interesse, dados de marcadores moleculares os quais são obtidos em laboratórios especializados. Uma vez que esses dados estejam disponíveis é

¹ Departamento de Ciências Exatas, Escola Superior de Agricultura “Luiz de Queiroz” - ESALQ, Universidade de São Paulo - USP, CEP: 13418-900, Piracicaba, São Paulo, Brasil. E-mail: erdtoledo@gmail.com / rleandro@esalq.usp.br

² Departamento de Genética, Escola Superior de Agricultura “Luiz de Queiroz” - ESALQ, Universidade de São Paulo - USP, CEP: 13418-900, Piracicaba, São Paulo, Brasil. E-mail: clsouza@esalq.usp.br

³ Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas – UNICAMP, Barão Geraldo, - Caixa-Postal: 6109, CEP: 13083-970 - Campinas, SP – Brasil. E-mail: anete@pq.cnpq.br

necessário: (i) distribuir os marcadores utilizados em grupos, chamados grupos de ligação; (ii) ordená-los no grupo; (iii) dentro de cada grupo estimar a distância entre os marcadores adjacentes ou marcadores flanqueadores. Realizados os passos (i), (ii) e (iii) têm-se o mapa genético. Apresenta-se na Figura 1 um mapa genético dos dez cromossomos de milho com 139 marcadores, neste caso, cada cromossomo é considerado um grupo. Existem várias técnicas para a construção de mapas (Ragonha, 2005; Liu, 1998, Wu et al., 2002). Uma vez construído o mapa pode-se dar continuidade ao mapeamento de QTLs. Existem, na literatura, vários métodos estatísticos para o mapeamento de QTLs (Lander e Bolstein, 1989; Zeng, 1993, 1994, 2002), grande parte deles utiliza a abordagem clássica. Neste trabalho apresenta-se o mapeamento de QTLs utilizando a abordagem Bayesiana. A suposição adotada para a realização desse mapeamento é de que entre dois marcadores flanqueadores (duas marcas adjacentes) existe somente um QTL.

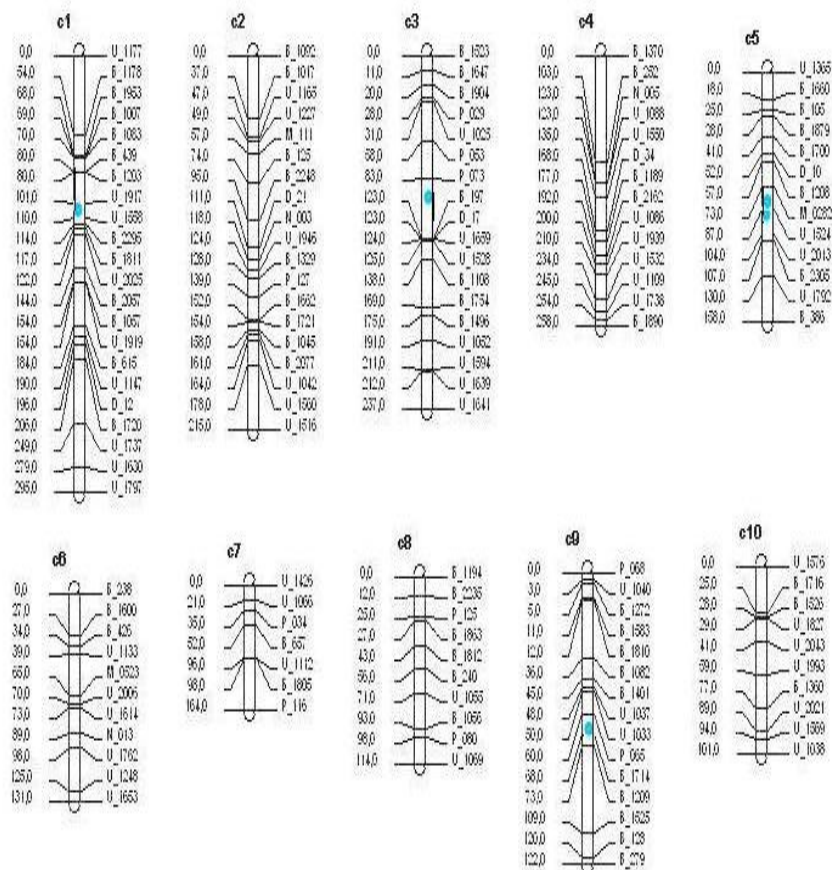


Figura 1 - Mapa genético dos dez cromossomos de milho com 139 marcas com os cinco QTLs obtidos pelo mapeamento por intervalo Bayesiano, representados pelos círculos.

2 Material e Métodos

2.1 Material

Os dados utilizados neste trabalho são provenientes de 256 progênies $F_{2:3}$ de milho, obtidas a partir do cruzamento entre as linhagens L02-03D e L20-01F contrastantes para vários caracteres agronômicos. Ambas as linhagens foram desenvolvidas no Departamento de Genética da ESALQ/USP. Os dados genotípicos de marcadores moleculares para todas as progênies utilizadas no experimento foram obtidos no Laboratório de Biologia Molecular e Análises Genéticas, situado no Centro de Biologia Molecular e Engenharia Genética (CBMEG), na Universidade Estadual de Campinas/UNICAMP. As progênies foram avaliadas em experimentos instalados em látices simples 16 x 16 com duas repetições em nove ambientes, sendo cada ambiente correspondente à combinação local x ano agrícola. O caractere estudado neste trabalho foi produção de grãos de milho, em gramas planta⁻¹. Os dados analisados são referentes à média dos nove ambientes.

2.2 Métodos

O mapeamento por intervalo Bayesiano (BIM, “Bayesian Interval Mapping”) considera o seguinte modelo estatístico:

$$y_i = \mu + \sum_{j=1}^P a_j x_{ij}^* + \sum_{j=1}^P d_j z_{ij}^* + e_i$$

em que y_i se refere ao valor fenotípico do i -ésimo indivíduo ($i = 1, \dots, n$), μ é uma constante, $e_i \sim N(0, \sigma^2)$ o erro aleatório, a_j e d_j é, respectivamente, o efeito de aditividade e dominância para do QTL j , $j = 1, \dots, p$. A contribuição do j -ésimo QTL para o indivíduo i , (Q_{ij}) é definida pelas variáveis indicadoras x_{ij}^* e z_{ij}^* . A variável indicadora x_{ij}^* está associada ao efeito aditivo e corresponde aos genótipos homocigóticos: assume valores -1 e 1 para qq e QQ , respectivamente; z_{ij}^* está associada ao efeito de dominância e corresponde ao genótipo heterocigóticos, Qq ; $z_{ij}^* = 1 - |x_{ij}^*|$ será igual a 1 quando $x_{ij}^* = 0$ e, quando $x_{ij}^* = \pm 1$, $z_{ij}^* = 0$. Os valores das variáveis indicadoras, que são os possíveis genótipos do QTL j , são quantidades desconhecidas que possuem distribuição massa de probabilidade que são estimadas utilizando-se a informações dos marcadores flaqueadores (Satagopan et al., 1996).

Uma notação equivalente é dada por:

$$\mathbf{y} = \mathbf{X}^{(p)} \boldsymbol{\Theta}^{(p)} + \boldsymbol{\varepsilon} \quad (1)$$

sendo \mathbf{y} o vetor de valores fenotípicos de n indivíduos; $\boldsymbol{\Theta}^{(p)} = [\mu, \sigma^2, a_1, d_1, a_2, d_2, \dots, a_p, d_p]^T$ o vetor de parâmetros desconhecidos: a_j e d_j associados aos efeitos aditivo e dominância

do QTL j , com $2p+1$ colunas (se o número de QTLs for considerado aleatório, o comprimento do vetor de parâmetros, $\theta^{(p)}$, também o será); ε é o vetor de erros, normalmente distribuídos, $\varepsilon \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$; a matriz de delineamento $\mathbf{X}^{(p)} = \begin{bmatrix} 1 & x_1 & z_1 & x_2 & z_2 & \dots & x_p & z_p \end{bmatrix}$, com n linhas e $2p+1$ colunas incluirá todas as informações genótípicas dos QTLs, sendo $x_j = [x_{1j}, x_{2j}, \dots, x_{nj}]^T$ e $z_j = [z_{1j}, z_{2j}, \dots, z_{nj}]^T$ vetores indicadores para os genótipos dos marcadores flanqueados ao QTL. (Gaffney, 2001; Satagopan et al., 1996). O número de QTLs, e portanto o comprimento do vetor dos efeitos de aditividade e de dominância, não é uma quantidade observável.

Para utilização da abordagem Bayesiana é necessário: (i) construir a verossimilhança; (ii) atribuir distribuições *a priori* às quantidades não observáveis, isto é, é necessário especificar distribuições *a priori* para a constante μ ; para os vetores de aditividade, dominância, posição e a variância σ^2 . Além disso, considerando-se que o número de QTLs (P) é desconhecido, também, especifica-se uma distribuição *a priori* para P .

Visto que a distribuição conjunta *a posteriori* para os parâmetros de interesse do modelo possui forma analítica complexa, resumos *a posteriori* para estes parâmetros apresentam cálculos inviáveis. Desse modo, empregam-se métodos computacionalmente intensivos (algoritmo Monte Carlo via Cadeias de Markov, MCMC), para a obtenção de uma amostra da distribuição conjunta *a posteriori*.

A suposição de que o número de QTLs é desconhecido, torna a obtenção da amostra da distribuição conjunta *a posteriori* ligeiramente mais complicada. Uma solução para esse problema está na utilização de algoritmos trans-dimensionais projetados para construir cadeias de Markov onde a dimensão do espaço de estados muda ao longo das iterações, tais como algoritmos MCMC com saltos reversíveis (RJCMC, “Reversible Jump MCMC”) (Green, 1995; Sorensen e Gianola, 2002).

Para construção da amostra da distribuição conjunta *a posteriori* foi utilizado o módulo BIM do programa QTLCartographer, versão 2.5 para Windows. Foi gerada uma cadeia de Markov com 400.000 iterações, “burn-in” de 2000, “pré burn-in” de 1000 e “thin” de 100 iterações sendo obtida uma amostra de tamanho 15.315.

Para a seleção do número de QTLs foi utilizado o Fator de Bayes (BF) definido por

$$BF(M_i, M_j) = \frac{\left(\frac{P(M_i | y)}{P(M_j | y)} \right)}{\left(\frac{P(M_i)}{P(M_j)} \right)},$$

sendo $P(M_i | y)$ e $P(M_j | y)$ as probabilidades *a posteriori* condicionadas às observações para os modelos M_i e M_j respectivamente e, $P(M_i)$ e $P(M_j)$ as probabilidades *a priori* para os modelos com j e com $j+1$ QTLs, ou seja, M_j e $M_{(j-1)}$, M_i e M_j . A Tabela 1 apresenta uma calibração para o Fator de Bayes baseado no valor de $BF(M_j; M_{(j-1)})$ (Raftery, 1995):

Table 1 - Decisões sobre a evidência de M_j em relação a $M_{(j-1)}$

Valores de $B(M_j, M_{(j-1)})$	Conclusão
$1 \leq B(M_j, M_{(j+1)}) \leq 3$	Evidência a favor de M_j
$3 < B(M_j, M_{(j+1)}) \leq 10$	Evidência positiva a favor de M_j
$10 < B(M_j, M_{(j+1)}) \leq 100$	Forte evidência a favor de M_j
$B(M_j, M_{(j+1)}) > 100$	Evidência decisiva a favor de M_j

Fonte: Jeffreys, 1935, apud Raftery, 1995.

3 Resultados de Discussão

Considerando-se amostra aleatória obtida para a quantidade P , conclui-se que o modelo com cinco QTLs ajusta-se melhor aos dados. O número de QTLs presente no genoma foi escolhido utilizando-se o Fator de Bayes.

Considerando-se o modelo com cinco QTLs (M_5) e considerando-se que o genoma do milho possui dez cromossomos foi definida a configuração cromossômica dos cinco QTLs, ou seja, foi escolhido através do Fator de Bayes a possível distribuição dos QTLs nos cromossomos. A seguir, foram estimadas: sua posição (λ) no cromossomo escolhido e seus os efeitos: efeito aditivo (a) e de dominância (d) e o tipo de interação alélica (GD) ocorrida (LIU, 1998). A Tabela 2 apresenta o resumo dos resultados obtidos na análise. Os nomes dos QTLs são compostos por código indicando o caractere correspondente seguido por número indicando o cromossomo no qual ele se encontra o QTL e por letra identificando os vários QTLs localizados no mesmo cromossomo. Dessa forma, observa-se que QTL nomeado por qprod5 significa presença de QTL para o caractere produção de grãos localizado no cromossomo cinco. As letras a, b identificam os vários QTLs localizados no mesmo cromossomo.

A maior parte dos alelos que contribui de forma favorável para o caractere é proveniente da linhagem parental L02-03D, que apresentou maior produção de grãos. A contribuição através de alelos favoráveis para essa linhagem foi efetiva para quatro dos cinco QTLs mapeados. Os QTLs detectados pelo mapeamento por intervalo Bayesiano, localizados no genoma do milho podem ser vistos na Figura 1. As progênies $F_{2:3}$ avaliadas apresentaram produção média de 43,84 gramas planta⁻¹. O coeficiente de herdabilidade (\hat{h}^2) foi de 15%, indicando baixa herdabilidade para o caractere.

Os cinco QTLs associados a produção de grãos, mapeados através da abordagem Bayesiana, estão distribuídos nos cromossomos 1 (um QTL), 3 (um QTL), 5 (dois QTLs) e 9 (um QTL). Os efeitos aditivos (a) dos QTLs para produção de grãos variaram de -0,23 a 0,92, enquanto os de dominância (d) variaram de -0,16 a 0,34, para os QTLs qprod5b* e qprod9a* respectivamente. As estimativas dos efeitos genéticos e dos valores de $|d|/|a|$ para cada QTL indicaram a ocorrência de diferentes tipos de interações alélicas, mostrando que as interações gênicas predominantes foram de dominância parcial (2 QTLs) e aditiva (2 QTLs), seguidas por sobredominância (1 QTL). O grau médio de dominância estimado para o caractere produtividade de grãos considerando todos os QTLs

mapeados ponderados por seus respectivos $\hat{R}^2_{G^*}$ foi de 0,28, evidenciando que o caractere apresenta em média dominância parcial.

Table 2 - Resumo do mapeamento de QTLs para o caráter produção de grãos

QTL	Crom ^a	Posição cM	Intervalo	Efeitos		
				a	d	
qprod1a*	1	106,02	U1917-U1558	0,51	0,19	
qprod3a*	3	107,34	P073-B197	0,81	0,09	
qprod5a*	5	75,52	M0282-U1524	0,66	0,25	
qprod5b*	5	98,05	U1524-U2013	0,92	-0,16	
qprod9a*	9	64,84	P065-B1714	-0,23	0,34	
					<i>GMD</i> ^e	

Tabela 2 - (continuação)

QTL	Crom ^a	$ \hat{d} / \hat{a} $	GD ^b	Direção	$\hat{R}^2_{F^*}(\%)$ ^c	$\hat{R}^2_{G^*}(\%)$ ^d
qprod1a*	1	0,38	DP	L02-03D	0,19	1,31
qprod3a*	3	0,11	A	L02-03D	0,47	3,15
qprod5a*	5	0,37	DP	L02-03D	0,33	2,20
qprod5b*	5	0,18	A	L02-03D	0,60	4,04
qprod9a*	9	1,45	SD	L20-01F	0,08	0,53
		0,28	DP	R ² Total	1,66	11,23

(a) Cromossomo em que foi detectado o QTL; (b) GD (grau de dominância): A=aditiva ($|\hat{d}|/|\hat{a}| < 0,2$), DP=dominância parcial ($0,21 \leq |\hat{d}|/|\hat{a}| < 0,8$), DC=dominância completa ($0,81 \leq |\hat{d}|/|\hat{a}| < 1,2$), SD=sobredominância ($|\hat{d}|/|\hat{a}| > 1,2$); (c) $\hat{R}^2_{F^*}(\%)$: Porcentagem da variação fenotípica explicada pelo QTL; (d) $\hat{R}^2_{G^*}(\%)$: Porcentagem da variação genotípica explicada pelo QTL; (e) *GMD* (grau médio de dominância)

$$GMD = \frac{\sum_i GD_i \hat{R}_{G_i}^2}{\sum_i \hat{R}_{G_i}^2}$$

A proporção da variância fenotípica ($\hat{R}^2_{F^*}$) explicada pelos QTLs variou de 0,08% a 0,6%, enquanto a proporção da variância genética ($\hat{R}^2_{G^*}$) variou de 0,53% a 4,04%, explicada, respectivamente, pelos QTLs qprod9a* e qprod5b*. Em conjunto, os cinco QTLs detectados para produção de grãos, explicaram 1,66% da variância fenotípica e 11,23% da variância genética do caractere.

A Tabela 3 apresenta os intervalos de credibilidade de 95% para a localização e os efeitos aditivo (a) e de dominância (d),

Tabela 3 - Intervalos de Credibilidade para os efeitos genéticos e localização (λ) dos QTLs controladores do caráter produção de grãos

QTL	λ	Efeitos	
		a	d
qprod1a*	(12,56; 276,47)	(-1,95; 3,10)	(-1,61; 2,17)
qprod3a*	(11,028; 275,27)	(-2,10; 3,70)	(-1,93; 2,36)
qprod5a*	(4,61; 153,48)	(-2,36; 3,84)	(-1,93; 2,56)
qprod5b*	(3,64; 240,11)	(-2,21; 3,76)	(-2,64; 1,96)
qprod9a*	(2,95; 141,29)	(-2,64; 3,37)	(-2,03; 2,73)

Conclusões

Os resultados obtidos nesse trabalho possibilitaram as seguintes conclusões:

- A vantagem de se utilizar informações adicionais sobre as quantidades desconhecidas (distribuições *a priori*) que conectadas às informações dos dados observados (verossimilhança) são atualizadas segundo a abordagem Bayesiana;
- Existência de cinco QTLs associados à produção de grãos sendo 1 QTL nos cromossomos um, três e nove; e dois QTLs no cromossomo cinco.
- Na abordagem clássica é necessário determinar o limiar crítico para a presença de QTLs enquanto que na abordagem Bayesiana não existe essa necessidade;
- É possível calcular intervalos de credibilidade $(1 - \alpha)\%$ que garantem que o parâmetro de interesse pertence a esse intervalo.
- Uma fragilidade verificada, na metodologia Bayesiana de mapeamento por intervalo, através da análise de sensibilidade foi com relação a especificação da *priori* para a o número de QTLs presente no modelo, P . Estudos estão sendo realizados para solucionar essa fragilidade.
- Ainda, verifica-se a necessidade de implementação de programas para obtenção de amostras da distribuição *a posteriori* conjunta visto que o *software* utilizado não permite interações com o pesquisador.

Agradecimentos

Agradecemos ao departamento de Genética da ESALQ/USP pelo fornecimento dos dados analisados neste trabalho e à CAPES pela concessão do suporte financeiro.

TOLEDO, E. R.; LEANDRO, R. A.; SOUZA JUNIOR, C. L.; SOUZA, A. P. Mapping QTL: a Bayesian approach. *Rev. Bras. Biom.*, São Paulo, v.26, n.2, p.107-1114, 2008.

- **ABSTRACT:** Characters of agronomics importance, in its majority, can be classified as quantitative, whose phenotypic expression presents continuous variation, attributed to the

simultaneous segregation of many genes, in definite regions as QTLs (Quantitative Trait Loci), Mapping QTL is identify its position in the genoma and esteem its effect, Some methods of mapping exist, great part of them present classic boarding, This work will present Bayesian approach,

- **KEYWORDS:** *Quantitative trait loci; Bayesian approach; Bayes factor; reversible jump MCMC.*

Referências

- GAFFNEY, P. J. *An efficient reversible jump Markov chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses.* 2001. 194f. Ph. D. (Doctor of Philosophy in Statistics) - University of Wisconsin, Madison, 2001.
- GREEN, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, Oxford, v.82, p.711-732, 1995.
- LIU, B. H. *Statistical genomics: linkage, mapping and QTL analysis.* New York: CRC Press, 1998. 611p.
- RAFTERY, A. E. Bayesian model selection in social research. *Sociol. Methodol.*, Washington, v.25, p.111-163, 1995.
- RAGONHA, F. Construção de mapas genéticos em espécies de polinização aberta: Uma abordagem Bayesiana com o uso de uma priori informativa. 2005. 149f. Dissertação (Mestrado em Experimentação Agrícola) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, São Paulo, 2005.
- SATAGOPAN, J. M. et al. A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics*, Pennsylvania, v.144, p.805-816, 1996.
- SILLANPÄÄ, M. J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics*, Pennsylvania, v.148, p.1373-1388, 1998.
- SORENSEN, D.; GIANOLA, D. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer-Verlag, 2002. 740p.
- ZENG, Z. B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Nat. Acad. Sci.*, v.90, p.10972-10976, 1993.
- ZENG, Z. B. Precision mapping of quantitative trait loci. *Genetics*, Pennsylvania, v.136, n.4, p. 1457-1468, 1994.
- ZENG, Z. B. Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics*, Pennsylvania, v.160, p.1243-11261, 2002.
- WU, R.; MA, C. X.; PAINTER, I.; ZENG, Z. B. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor. Popul. Biol.*, Saint Louis, v.61, p.349-363, 2002.

Recebido 12.07.2007.

Aprovado após revisão 18.03.2008.