

ANÁLISE BAYESIANA DO MODELO DE HERANÇA MONOGÊNICA NO MELHORAMENTO VEGETAL: UM EXEMPLO COM ABOBRINHA

Bayesian analysis of monogenic inheritance model in plant breeding: a case study with zucchini

Maria Imaculada de Sousa Silva¹, Eduardo Bearzoti², Júlio Sílvio de Sousa Bueno Filho³

RESUMO

Uma estratégia comum em programas de melhoramento é conduzir estudos básicos de herança para investigar a hipótese de controle do caráter por um ou poucos genes de efeito principal, associados ou não a genes modificadores de pequeno efeito. Neste trabalho, foi utilizada a inferência bayesiana para ajustar modelos de herança genética aditiva-dominante a experimentos de genética vegetal com várias gerações. Densidades normais com médias associadas aos efeitos genéticos das gerações foram consideradas em um modelo linear em que a matriz de delineamento dos efeitos genéticos tinha coeficientes indeterminados (precisando ser estimada para cada indivíduo). A metodologia foi ilustrada com um conjunto de dados de um estudo de herança da partenocarpia em abobrinha (*Cucurbita pepo* L.). Tal ajuste permitiu explicitar a distribuição a posteriori das probabilidades genotípicas. A análise corrobora resultados anteriores da literatura, porém foi mais eficiente que alternativas prévias que supunham a matriz de delineamento conhecida para as gerações. Conclui-se que a partenocarpia em abobrinha é governada por um gene principal com dominância parcial.

Termos para indexação: Algoritmo Metropolis, Amostragem Gibbs, *Cucurbita pepo*, Gene Principal, Partenocarpia.

ABSTRACT

A common breeding strategy is to carry out basic studies to investigate the hypothesis of a single gene controlling the trait (major gene) with or without polygenes of minor effect. In this study we used Bayesian inference to fit genetic additive-dominance models of inheritance to plant breeding experiments with multiple generations. Normal densities with different means, according to the major gene genotype, were considered in a linear model in which the design matrix of the genetic effects had unknown coefficients (which were estimated in individual basis). An actual data set from an inheritance study of partenocarpia in zucchini (*Cucurbita pepo* L.) was used for illustration. Model fitting included posterior probabilities for all individual genotypes. Analysis agrees with results in the literature but this approach was far more efficient than previous alternatives assuming that design matrix was known for the generations. Partenocarpia in zucchini is controlled by a major gene with important additive effect and partial dominance.

Index terms: *Cucurbita pepo*, Gibbs sampler, major gene, Metropolis algorithm, partenocarpia.

(Recebido em 18 de junho de 2007 e aprovado em 12 de dezembro de 2008)

INTRODUÇÃO

Em estudos de herança genética de uma característica contínua, a hipótese de herança monogênica frequentemente é testada por meio da aderência à distribuição de qui-quadrado (Arias et al., 1994; Souza Sobrinho, 1998; Freitas et al., 2002). Supondo a ação de polígenes, em geral a inferência é feita estimando componentes de média por meio do teste de escala conjunto, e os componentes de variância pela análise de variância (Mather & Jinks, 1984; Ramalho et al., 1993). Recentemente, estudos de herança propõem investigar simultaneamente a ação de um único gene (gene principal) e ou de vários genes de pequeno efeito (polígenes) (Changjian et al., 1994; Lou & Zhu, 2002; Silva, 2003). Com este enfoque, a inferência bayesiana também tem sido usada para inferir sobre tal herança, porém, no melhoramento animal (Janss et al., 1995, 1997; Sorensen, 1996).

Neste trabalho, objetivou-se apresentar estimadores bayesianos de parâmetros referentes ao gene principal, contemplando as particularidades dos experimentos com plantas, tais como a existência de várias gerações e a consideração de efeitos de dominância. A técnica é ilustrada com um conjunto de dados reais, com a finalidade de estudar a herança do fenômeno da partenocarpia em abobrinha (*Cucurbita pepo* L.). A partenocarpia, que consiste na formação de frutos sem que ocorra fertilização, é um caráter de crescente importância no melhoramento de abobrinha, pois permite a boa produção de frutos, mesmo na ausência de agentes polinizadores, ou em situações em que a polinização ou a fecundação seja deficiente por incompatibilidades morfológicas ou outras (Nogueira, 2007).

¹Doutoranda em Estatística e Experimentação Agropecuária. DEX-UFLA. Caixa postal 3037, 37200-000, Lavras, MG. Bolsista do CNPq-Brasil – maissilva@gmail.com

²Dr., IA – UNICAMP, Caixa Postal 6159, 13083-854, Campinas, SP – edbearzoti@yahoo.com.br

³Prof. Dr. DEX-UFLA. Caixa postal 3037, 37200-000, Lavras, MG. Bolsista do CNPq-Brasil – jssbueno@ufla.br

MATERIALE MÉTODOS

Foi analisado um conjunto de dados oriundos de um experimento desenvolvido no setor de Olericultura da Universidade Federal de Lavras, sobre a herança do fenômeno da partenocarpia em abobrinha. O método empregado foi o da condução de um cruzamento segregante (Mather & Jinks, 1984). Foram consideradas seis gerações, sendo que os genitores contrastantes considerados foram a variedade Caserta (P_1) e a variedade Whitaker (P_2); o cruzamento entre elas (geração “ F_1 ”), o cruzamento entre indivíduos P_1 e F_1 (“ RC_{11} ”); o cruzamento entre indivíduos P_2 e F_1 (“ RC_{12} ”) e o cruzamento entre indivíduos F_1 (geração “ F_2 ”). O número de plantas amostradas foi de 562, sendo destas, 94 da geração P_1 , 51 de P_2 , 53 de F_1 , 86 do RC_{11} , 75 do RC_{12} e 203 de F_2 . As plantas amostradas foram avaliadas com notas de 1 a 5, conforme menor ou maior ocorrência de partenocarpia, respectivamente, tomando-se as médias das notas por planta atribuídas por três avaliadores. O modelo linear proposto tem a seguinte representação:

$$Y|\beta, W, m, \sigma^2 \sim N(X\beta + Wm, I\sigma^2); \quad \varepsilon \sim N(\Phi, I\sigma^2),$$

em que $\sim N(a, b)$ representa a distribuição multivariada normal com média “a” e covariância “b”. Nesse modelo, Y é o vetor dos dados (notas médias), b contém uma constante de referência m , m é um vetor de efeitos fixos, aditivos (A) e de dominância (D), do gene principal, de forma que $m' = [A, D]$, X é o vetor de incidência de b (suposto conhecido) e ε é o vetor de erros aleatórios. A matriz de delineamento dos efeitos genéticos W (desconhecida) contém vetores aleatórios w_i para cada observação, referentes à incidência de m , de forma que $W' = (w_1, w_2, \dots, w_n)$, sendo n o número de plantas observadas. Os vetores w_i podem assumir os

valores $[1, 0]$, $[-1, 0]$ ou $[0, 1]$, associados respectivamente ao homozigoto cujos alelos são favoráveis à incidência da partenocarpia, ao homozigoto cujos alelos são desfavoráveis, ou ao heterozigoto. Portanto, para cada uma das gerações P_1 , P_2 e F_1 , por apresentarem apenas um genótipo, o vetor w_i pode assumir apenas um valor, diferente para cada uma delas, enquanto que para as gerações RC_{11} , RC_{12} e F_2 existem duas ou três possibilidades, conforme a Tabela 1, sendo que em tais casos, os vetores w_i constituem um parâmetro a ser estimado, com distribuições a priori $p(w_i)$ independentes, apresentadas na própria Tabela 1.

Para os vetores b e m foram atribuídas distribuições a priori não informativas e degeneradas, ou seja, proporcionais a constantes. Para a componente da variância σ^2 (variância ambiental), considerou-se como priori uma distribuição qui-quadrado escalada invertida com $\nu=6$ graus de liberdade e parâmetro de escala $S_e=7$.

Tabela 1 – Possíveis valores para o vetor w_i e suas probabilidades a priori.

Geração	$w_i = [-1, 0]$	$w_i = [1, 0]$	$w_i = [0, 1]$
	Probabilidade		
P_1	1	0	0
P_2	0	1	0
F_1	0	0	1
RC_{11}	1/2	0	1/2
RC_{12}	0	1/2	1/2
F_2	1/4	1/4	1/2

Supondo distribuição normal, a função de verossimilhança é dada pela expressão (2.1), e a distribuição conjunta de todos os parâmetros, supondo independência entre eles, é dada pela expressão (2.2), nas quais o símbolo μ representa proporcionalidade.

$$L(Y) = p(Y|\beta, W, m, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}(Y - (X\beta + Wm))'(I\sigma^2)^{-1}(Y - (X\beta + Wm))\right\} \quad (2.1)$$

$$p(\beta, W, m, \sigma^2 | Y) \propto L(Y)p(W)p(\beta)p(m)p(\sigma^2) \propto L(Y)p(W) \left(\frac{1}{\sigma^2}\right)^{\frac{\nu}{2}+1} e^{-\frac{\nu S_e}{2\sigma^2}} \quad (2.2)$$

As distribuições condicionais completas para cada parâmetro são obtidas a partir da distribuição conjunta pela solução da seguinte identidade:

$$p(c/a, b) = \frac{p(b, c/a)}{p(b/a)} = \frac{p(b, c/a)}{\int p(b, c/a)dc}$$

em que $p(c/a,b)$ é a condicional completa para c , $p(b,c/a)$ é distribuição conjunta para b e c , e $p(b|a)$ é a distribuição marginal para b . Isso pode ser obtido com métodos práticos de simplificação das condicionais para distribuições proporcionais que dispensam o cálculo explícito das integrais do denominador (Gelfand & Smith, 1990; Sørensen, 1996).

Assim, obteve-se que a distribuição condicional completa para \mathbf{b} é normal com média $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y}-\mathbf{W}\mathbf{m})$ e variância σ^2 . A distribuição condicional completa para \mathbf{m} é normal com vetor de médias $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'(\mathbf{Y}-\mathbf{X}\mathbf{b})$ e matriz de covariâncias igual a $\mathbf{I}\sigma^2$.

$$p(\sigma^2 | \mathbf{W}, \mathbf{b}, \mathbf{m}, \mathbf{Y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n+y}{2}+1} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{(\mathbf{Y}-(\mathbf{X}\mathbf{b}+\mathbf{W}\mathbf{m}))'(\mathbf{Y}-(\mathbf{X}\mathbf{b}+\mathbf{W}\mathbf{m}))+S_e}{2}\right]\right\} \quad (2.4)$$

As condicionais completas dos parâmetros \mathbf{b} , \mathbf{m} e σ^2 são todas densidades conhecidas, a partir das quais as amostras podem ser obtidas, utilizando-se o algoritmo do amostrador de Gibbs. Para a matriz \mathbf{W} , deve-se utilizar o algoritmo Metropolis-Hastings, um método de simulação Monte Carlo via cadeias de Markov usado para aproximar uma distribuição de interesse, quando esta não apresenta a forma de uma densidade conhecida.

Para a análise, foi desenvolvido um programa utilizando o software R, gerando duas cadeias com 10.000 iterações, descartando as 2.000 iniciais e tomando amostras a cada duas iterações, totalizando duas sequências de 4.000 amostras da densidade de cada parâmetro.

RESULTADOS E DISCUSSÃO

As cadeias foram submetidas aos critérios de convergência de Gelman e Rubin e de Geweke, os quais apontaram a convergência para os parâmetros \mathbf{b} , A , D , e σ^2 , concluindo-se que a amostra gerada aproxima-se da distribuição desejada. A distribuição a posteriori dos parâmetros estão representadas graficamente pelas Figuras 1 e 2. Tais figuras sugerem que as distribuições dos parâmetros \mathbf{m} , A e D apresentam uma tendência à simetria, porém a variância (σ^2) apresentou distribuição a posteriori ligeiramente assimétrica (Figura 1.b).

Para exemplificar a utilidade de estimar a matriz \mathbf{W} , foram escolhidos aleatoriamente alguns indivíduos de cada geração segregante, e analisada a probabilidade a posteriori que cada um tem de corresponder a cada vetor \mathbf{w}_i e, os resultados estão apresentados na tabela 2.

Na Tabela 2, estão apresentados também os fenótipos (\mathbf{Y}) dos indivíduos em questão. Percebe-se uma clara coerência entre estes e as probabilidades a posteriori de cada vetor \mathbf{w}_i .

A distribuição condicional completa para os vetores \mathbf{w}_i da matriz \mathbf{W} é dada por:

$$p(\mathbf{w}_i | \beta, \mathbf{m}, \mathbf{w}_{-i}, s^2, \mathbf{Y}) \propto L(\mathbf{Y}) p(\mathbf{w}_i), \quad (2.3)$$

sendo que \mathbf{w}_{-i} representa todos os vetores da matriz \mathbf{W} , com exceção de \mathbf{w}_i .

A condicional completa para o parâmetro σ^2 é dada por (2.4) a seguir. Pode-se notar que a forma dessa distribuição é de uma gama invertida.

Por exemplo, na geração RC_{11} , os dois indivíduos com valor fenotípico 1,00 tiveram alta probabilidade de serem homozigóticos, ou seja, de estarem associados ao vetor \mathbf{w}_i igual a $[-1,0]$. Como as plantas foram avaliadas com notas de 1 a 5, de acordo com uma menor ou maior incidência de partenocarpia, fica evidente a coerência do resultado, já que uma planta com a menor nota (1) é aquela em que a característica não se manifestou e assim seria mais lógico que ela apresentasse apenas os alelos que diminuem a expressão do caráter, como ocorre no indivíduo homozigótico dessa geração. Na geração RC_{12} , o indivíduo com valor fenotípico igual a 1,00 apresentou probabilidade nula de ser homozigoto associado ao vetor $\mathbf{w}_i = [0,1]$, o que também foi perfeitamente coerente, pois, com uma nota tão baixa seria improvável que ele apresentasse dois alelos que aumentassem a expressão do caráter, como ocorre com os indivíduos homozigóticos dessa geração. Ainda na geração RC_{12} , um indivíduo com nota igual a 4,33 apresentou uma alta probabilidade de ser homozigoto contra uma baixa probabilidade de ser heterozigoto. Portanto, percebe-se claramente pela Tabela 2 que, na geração RC_{11} , quanto maior o valor fenotípico, maior a probabilidade de o indivíduo ser heterozigoto, e na geração RC_{12} , maior a probabilidade de o indivíduo ser homozigoto dominante com o aumento do fenótipo. Comportamentos coerentes com a escala de valores fenotípicos também foram observados na geração F_2 . Exemplos disso são as plantas avaliadas com notas mais altas (4,00 ou mais), que apresentaram probabilidades nulas de serem homozigotas recessivas ($\mathbf{w}_i = [-1, 0]$) e as plantas avaliadas com notas baixas (1,667), que apresentaram probabilidade nula de serem homozigotas dominantes ($\mathbf{w}_i = [1, 0]$).

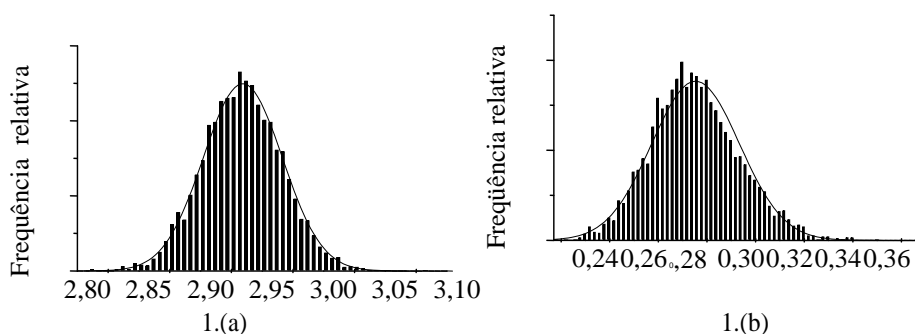


Figura 1 – Representação gráfica da distribuição a posteriori dos parâmetros μ (1.a) e s^2 (1.b) e da distribuição normal correspondente, em cada caso.

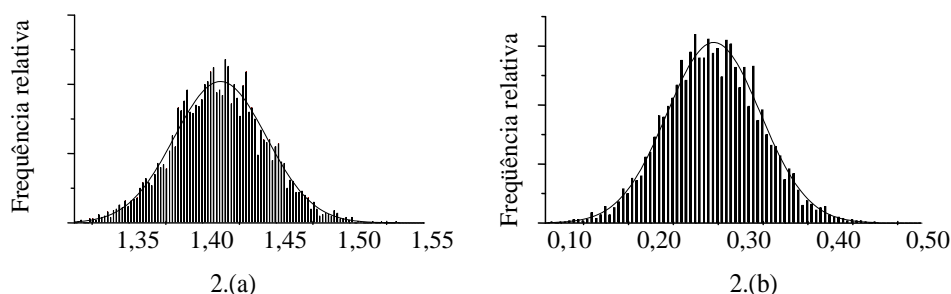


Figura 2 – Representação gráfica da distribuição a posteriori dos parâmetros A (2.a) e D (2.b) e da distribuição normal correspondente, em cada caso.

Na Tabela 2 ilustra-se uma vantagem da inferência bayesiana, no presente contexto, pois possibilita explicitar claramente probabilidades de cada indivíduo ser de um ou outro genótipo de um gene principal. Isso auxiliaria o geneticista de plantas a selecionar aqueles indivíduos com maiores probabilidades de terem fixado o alelo favorável ao caráter sob estudo.

Na Tabela 3, estão dispostas as estimativas a posteriori por ponto e por intervalo dos parâmetros \mathbf{b} , A, D, e σ^2 , e do erro de Monte Carlo. Os valores estimados de A e D permitem concluir que o fenômeno da partenocarpia possui herança monogênica com efeito de dominância parcial e efeito aditivo um pouco mais expressivo. Esses resultados estão de acordo com Menezes et al. (2005), que também registram a presença de um gene principal controlando a característica.

Verifica-se na Tabela 3 que, apesar de o modelo aqui sugerido levar em conta fatores antes não considerados, as estimativas a posteriori para os parâmetros \mathbf{b} , A e D tiveram valores semelhantes àqueles obtidos por Silva (2003), com esse mesmo conjunto de dados, pelo método da máxima verossimilhança. As medidas de qualidade de ajuste do modelo,

no entanto, são melhores. A amplitude dos IC para as estimativas diminuiu e, principalmente, houve redução significativa na estimativa da variância não explicada pelo delineamento (variância ambiental). Para a variância ambiental (σ^2), o valor obtido (0,29) mostrou-se abaixo da estimativa de máxima verossimilhança ($\sigma^2 = 0,8$). A rigor, a diferença não é apenas nas estimativas, mas nos modelos, pois o autor supôs a matriz \mathbf{W} como sendo conhecida, com os vetores que no modelo apresentado constituem apenas as probabilidades a priori para \mathbf{w}_i . Assim, por exemplo, um indivíduo fenotipicamente extremo na geração RC_{11} tem probabilidades diferenciadas de ser homocigoto ou heterocigoto, no presente modelo, enquanto que no modelo de Silva (2003) essas chances são iguais. É possível que essa seja a razão para a maior precisão na forma de análise ora apresentada, pois parte da variação que é essencialmente genética dentro de cada classe genotípica é confundida com variação ambiental no modelo com \mathbf{W} constante. É improvável que a análise bayesiana tenha introduzido *per se* a maior precisão, pois as distribuições a priori empregadas para \mathbf{b} , A, D foram constantes e para σ^2 , pouco informativa, ou seja, o resultado da análise bayesiana deveria ser muito próximo à análise de

verossimilhança pura. No modelo apresentado a estimação da matriz **W** permite levar em conta pesos mais adequados para cada indivíduo, levando a uma estimativa mais realista de δ^2 . Pode-se assim concluir que o ajuste do modelo com **W** aleatória é mais eficiente que o do modelo com **W** fixa, conforme Silva (2003).

Tabela 2 – Probabilidades a posteriori de cada vetor w_i para indivíduos de cada geração.

Geração	Vetor w_i			Y
	[1, 0]	[-1, 0]	[0, 1]	
	Probabilidade			
RC ₁₁	0	0,9997	0,0003	1,000
	0	0,9995	0,0005	1,000
	0	0,9820	0,0180	1,667
	0	0,5326	0,4674	2,333
	0	0,0003	0,9997	3,667
RC ₁₂	0,9105	0	0,0895	4,333
	0,1537	0	0,8463	3,333
	0,0000	0	1,0000	1,000
	0,0003	0	0,9997	2,000
	0,0005	0	0,9995	2,000
F ₂	0,7133	0,0000	0,2867	4,000
	0,0230	0,0100	0,9670	3,000
	0,0000	0,9788	0,0212	1,667
	0,0000	0,9730	0,0270	1,667
	0,0000	0,9635	0,0365	1,667
	0,9828	0,0000	0,0172	5,000
	0,5404	0,0000	0,4596	4,000
	0,0038	0,0812	0,9150	2,667
	0,9848	0,0000	0,0152	5,000
	0,8155	0,0000	0,1845	4,333

Tabela 3 – Estimativas a posteriori dos parâmetros do modelo, e erro de Monte Carlo (MC); IC-HPD: intervalo de credibilidade de máxima densidade de probabilidade (*highest probability density*), LI e LS: limites inferior e superior do IC-HPD.

Parâmetro	Média	IC-HPD		Erro de MC
		LS	LI	
β	2,93	2,87	2,99	0,016
A	1,41	1,34	1,47	0,016
D	0,27	0,17	0,37	0,016
s^2	0,29	0,25	0,32	0,016

CONCLUSÕES

Corrobora-se a hipótese de que a partenocarpia em abobrinha seja governada por um gene com efeito aditivo pronunciado e dominância parcial. O cálculo de probabilidades genotípicas para cada indivíduo aumentou a precisão das estimativas de efeitos aditivos e dominantes e diminuiu a estimativa do erro associado ao modelo.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pela bolsa de estudos concedida à primeira autora no estágio inicial de realização deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARIAS, C.A.A.; TOLEDO, J.F.F.; YORINORI, J.T. An improved procedure for testing theoretical segregation in qualitative genetic studies of soybeans. **Revista Brasileira de Genética**, Ribeirão Preto, v.17, n.3, p.291-297, 1994.
- CHANGJIAN, J.; XUEBIAO, P.; MINGHONG, G. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. **Genetics**, Berlin, v.136, p.383-394, 1994.
- FREITAS, J.A.; MALUF, W.R.; CARDOSO, M.G.; GOMES, L.A.A.; BEARZOTI, E. Inheritance of foliar zingiberene contents and their relationship to trichome densities and whitefly resistance in tomatoes. **Euphytica**, Wageningen, v.127, n.2, p.275-287, 2002.
- GELFAND, A.E.; SMITH, A.F.M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Madison, v.85, n.410, p.127-131, 1990.
- JANSS, L.L.G.; ARENDONK, J.A.M. van; BRASCAMP, E.W. Bayesian statistical analyses for presence of single genes affecting meat quality traits in a crossed pig population. **Genetics**, Berlin, v.145, p.395-408, 1997.
- JANSS, L.L.G.; THOMPSON, R.; ARENDONK, J.A.M. van. Application of Gibbs sampling for inference in a major gene-polygenic inheritance model in animal populations. **Theoretical and Applied Genetics**, Berlin, v.91, p.1137-1147, 1995.
- LOU, X.Y.; ZHU, J. Analysis of genetics effects of major genes and polygenes on quantitative traits. **Theoretical and Applied Genetics**, Berlin, v.104, p.414-421, 2002.

MATHER, K.; JINKS, J.L. **Introdução à genética biométrica**. Ribeirão Preto: Sociedade Brasileira de Genética, 1984. 242p.

MENEZES, C.B.; MALUF, W.R.; AZEVEDO, S.M.; FARIA, M.V.; NASCIMENTO, I.R.; NOGUEIRA, D.W.; GOMES, L.A.A.; BEARZOTI, E. Inheritance of parthenocarpy in summer squash (*Cucurbita pepo* L.). **Genetics and Molecular Research**, Cambridge, v.4, p.39-46, 2005.

NOGUEIRA, D.W. **Heterose e capacidade combinatória de linhagens de abobrinha partenocárpicas e resistentes ao Papaya ringspot vírus (PRSV-W)**. 2007. 25p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras, 2007.

RAMALHO, M.A.P.; SANTOS, J.B. dos; ZIMMERMANN, M.J. de O. **Genética quantitativa em plantas autógamas**: aplicação ao melhoramento do feijoeiro. Goiânia: UFG, 1993. 271p.

SILVA, W.P. **Estimadores de máxima verossimilhança em misturas de densidades normais: uma aplicação em genética**. 2003. 60p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras, 2003.

SORENSEN, D. **Gibbs Sampling in quantitative genetics**. Denmark: Danish Institute of Animal Science, 1996. 186p.

SOUZA SOBRINHO, F. **Herança da reação de resistência à raça 2 de *Meloidogyne incognita* na pimenta *Capsicum annum* L. cv *Carolina Cayenne***. 1998. 57p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)-Universidade Federal de Lavras, Lavras, 1998.