

ANÁLISE DO MODELO GENÉTICO ENVOLVIDO NO CONTROLE DE DIAS PARA FLORESCIMENTO EM SOJA¹

JOSÉ FRANCISCO F. DE TOLEDO² e ROMEU AFONSO DE S. KIIHL³

RESUMO - Dados de dias para floração de soja (*Glycine max* (L.) Merrill), das populações P₁, P₂, F₁, F₂, RC₁ e RC₂ foram obtidos sob condições de dias curtos, em experimento de delineamento em blocos completos casualizados, realizado em casa de vegetação. A linhagem D72-7842 e a variedade Santa Maria constituíram P₁ e P₂, respectivamente. Foram realizadas análises com a finalidade de obter informações sobre o modelo genético em controle da característica dias para floração. Concluiu-se pela existência de, pelo menos, três genes controladores do caráter, através de efeitos aditivos e de dominância. A herdabilidade no sentido restrito foi de 84,46%.

Termos para indexação: *Glycine max* (L.) Merrill.

ANALYSIS OF THE GENETIC MODEL CONTROLLING DAYS TO FLOWER IN SOYBEANS

ABSTRACT - Data of days to flower for six populations of soybeans (*Glycine max* (L.) Merrill), P₁, P₂, F₁, F₂, RC₁ and RC₂, were obtained under short day conditions in a green house randomized block design experiment. P₁ and P₂ were represented by the line D72-7842 and the variety Santa Maria respectively. Analysis on the genetic model in control of the characteristic days to flower was performed. It was concluded that at least three genes were in control of the characteristic through additive and dominance effects. The narrow sense heritability was 84,46%.

Index terms: *Glycine max* (L.) Merrill.

INTRODUÇÃO

A dificuldade da compreensão e adaptação das metodologias disponíveis para o estudo dos mecanismos de ação gênica tem causado alguns problemas aos melhoristas de plantas, notadamente autógamias. Estes problemas se traduzem por dificuldades na tomada de decisão quanto aos processos ou métodos mais adequados na condução e ou seleção de seus materiais.

Este trabalho é o primeiro de uma série que visa a apresentação e compatibilização das metodologias de análise genética desenvolvidas para o estudo das plantas autógamias.

Foi desenvolvido sobre dados de dias para floração em soja (sob condições de dias curtos) que incluíam as médias e variâncias das populações P₁, P₂, F₁, F₂, RC₁ e RC₂. P₁ foi a linhagem D-727842 e P₂, a cultivar Santa Maria. Foram discutidos os problemas enfrentados na procura de uma escala adequada aos propósitos desta análise e

feitas considerações sobre a estimativa da variância de ambiente que afeta as populações em questão. A estimativa da herdabilidade no sentido restrito foi realizada segundo a metodologia de Warner (1952). Para as estimativas do número de fatores em controle de caráter, usou-se a fórmula de Mather (1949). As estimativas acima dependem, para sua precisão e validade, de um modelo de ação gênica simplificado, sem epistasia. O ponto crucial da quase totalidade dos trabalhos em genética quantitativa reside na validade ou não de tais simplificações. Sendo assim, neste trabalho, procurou-se reunir algumas metodologias idealizadas para a descrição de efeitos genéticos, a partir de médias das populações disponíveis. Foram considerados os métodos de Mather (1949), Hayman (1958) e Cavalli (1952) que são de aplicação bastante direta.

Para melhor entendimento dos problemas de estimação decorrentes dos estudos de modelagem em genética, introduziu-se a nomenclatura de Cockerham (1973). Tal nomenclatura destaca-se pelo caráter geral e pela flexibilidade nas aplicações. No entanto, a sua complexidade nos processos nos quais a estimativa dos parâmetros genéticos está envolvida, limita a sua utilização prática, por isso, não foi exaustivamente examinada neste trabalho.

¹ Aceito para publicação em 9 de janeiro de 1982.

² Eng^o Agr^o, M.S., Centro Nacional de Pesquisa de Soja (CNPSo) - EMBRAPA, Caixa Postal 1061, CEP 86100 - Londrina, PR.

³ Eng^o Agr^o, Ph.D., CNPSo - EMBRAPA, Londrina, PR.

MATERIAL E MÉTODOS

Dados referentes a dias para floração de soja em dias curtos (fotoperíodo variando de onze horas e dez minutos a doze horas e trinta minutos) foram obtidos para as populações P_1 , P_2 , F_1 , F_2 , RC_1 e RC_2 em experimento realizado em casa de vegetação, durante o inverno 1975/76, no Estado do Mississippi, EUA, P_1 é representada pela linhagem D72-7842; P_2 , pela cultivar Santa Maria; as demais populações resultam dos cruzamentos, autofecundações e retrocruzamentos em esquemas bastante conhecidos. Todas as plântulas foram transplantadas, após germinação em vermiculita, para garantia de uniformidade nos blocos.

O delineamento experimental utilizado foi o de blocos completos, casualizados, com três repetições. As análises de variância foram conduzidas individualmente para cada população, de acordo com a Tabela 1. Estas análises de variância foram utilizadas para uma avaliação dos efeitos de uma possível interação entre genótipos e ambiente.

TABELA 1. Modelo de análise de variância utilizado para cálculo das variâncias fenotípicas de cada população.

Fonte de variação	G.L.	Q.M.
Blocos	(b-1)	SQB/(b-1)
Resíduo	(GL total-b+ 1)	(SQT-SQB)/ (GL total-b+ 1)

A partir da Tabela 2, seguindo-se a metodologia proposta por Warner (1952), calculou-se a herdabilidade no sentido restrito. Esta metodologia possui, pelo menos, dois pontos questionáveis: um por assumir um modelo genético livre de epistasia e outro por assumir que os componentes de ambiente das variâncias de F_2 e ($RC_1 + RC_2$) são comparáveis.

Sucintamente:

$$\sigma_{FF_2}^2 = \sigma_A^2 + \sigma_D^2 + \sigma_E^2$$

$$\sigma_{FRC_1}^2 + \sigma_{FRC_2}^2 = \sigma_A^2 + 2\sigma_D^2 + 2\sigma_E^2$$

onde:

$$\sigma_A^2 = \text{variância aditiva}$$

$$\sigma_D^2 = \text{variância de dominância}$$

$$\sigma_E^2 = \text{variância de ambiente}$$

Das equações descritas anteriormente, estimativas de σ_E^2 , σ_A^2 e σ_D^2 podem ser obtidas, se σ_E^2 for estimada a

TABELA 2. Médias, variâncias e número de plantas (símbolos) por geração.

População	Média	Var. fenotípica	Nº plantas observadas
P_1	\bar{P}_1	$\sigma_{FP_1}^2$	n_1
P_2	\bar{P}_2	$\sigma_{FP_2}^2$	n_2
F_1	\bar{F}_1	$\sigma_{FF_1}^2$	n_3
F_2	\bar{F}_2	$\sigma_{FF_2}^2$	n_4
RC_1	\bar{RC}_1	$\sigma_{FRC_1}^2$	n_5
RC_2	\bar{RC}_2	$\sigma_{FRC_2}^2$	n_6

partir da variância das gerações não segregantes. O cálculo da herdabilidade no sentido restrito foi diretamente realizado, como mostrado abaixo (Warner 1952):

$$\frac{2\sigma_{FF_2}^2 - \sigma_{FRC_1}^2 - \sigma_{FRC_2}^2}{\sigma_{FF_2}^2} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}$$

O número de fatores (n) de Mather (1949) e o grau de dominância (g.d.) dos genes foram estimados a partir das fórmulas seguintes:

$$n = \frac{1}{8} \frac{(\bar{P}_1 - \bar{P}_2)}{(\sigma_{FF_2}^2 - \sigma_{FF_1}^2)} \quad \text{g.d.} = \left[\frac{2\sigma_D^2}{\sigma_A^2} \right]^{\frac{1}{2}}$$

Os testes de validade do uso do modelo genético simples podem ser obtidos a partir das relações observadas abaixo (Mather & Jinks 1971), válidas em ausência de epistasia.

$$\bar{RC}_1 = \frac{1}{2} (\bar{P}_1 + \bar{F}_1)$$

$$\bar{RC}_2 = \frac{1}{2} (\bar{P}_2 + \bar{F}_1)$$

$$\bar{F}_2 = \frac{1}{4} (\bar{P}_1 + \bar{P}_2 + 2\bar{F}_1) = \frac{1}{2} (\bar{RC}_1 + \bar{RC}_2)$$

Dessa maneira, se escrevermos:

$$A = 2\bar{RC}_1 - \bar{P}_1 - \bar{F}_1$$

$$B = 2\overline{RC}_2 - \overline{P}_2 - \overline{F}_1$$

$$C = 4\overline{F}_2 - \overline{P}_1 - \overline{P}_2 - 2\overline{F}_1$$

poderemos testar o modelo genético envolvendo efeitos aditivos e de dominância, somente. Se o modelo simples realmente se aplicar, o valor esperado de A, B e C será zero. Na prática, entretanto, as relações acima devem ser analisadas dentro do contexto experimental e só são válidas dentro dos limites impostos pela precisão com a qual as várias médias são obtidas.

Seguindo-se este raciocínio, os testes de significância de A, B e C devem ser realizados com base nos respectivos erros calculados a partir das variâncias das análises de variância, como se mostra abaixo:

$$\text{Var (A)} = 4 \text{ Var } (\overline{RC}_1) + \text{Var } (\overline{P}_1) + \text{Var } (\overline{F}_1)$$

$$\text{Var (B)} = 4 \text{ Var } (\overline{RC}_2) + \text{Var } (\overline{P}_2) + \text{Var } (\overline{F}_1)$$

$$\text{Var (C)} = 16 \text{ Var } (\overline{F}_2) + \text{Var } (\overline{P}_1) + \text{Var } (\overline{P}_2) + 4 \text{ Var } (\overline{F}_1)$$

onde:

$$\text{Var } (\overline{F}_2) = \sigma_{FF_2}^2 / n_4, \text{ sendo } n_4 \text{ o número de plantas da geração } F_2$$

$$\text{Var } (\overline{RC}_1) = \sigma_{FRC_1}^2 / n_6, \text{ sendo } n_6 \text{ o número de plantas da geração } RC_1$$

$$\text{Var } (\overline{P}_1) = \text{etc...}$$

então:

$$t = \frac{(A)}{\sqrt{\text{Var (A)}}}, \text{ com tantos graus de liberdade quantos forem os graus de liberdade da Var (A).}$$

O cálculo dos graus de liberdade para o teste t, proposto acima, foi desenvolvido por Satterthwaite (1946). Exemplos da aplicação dessa metodologia estão apresentados a seguir.

Exemplo: cálculo do número de graus de liberdade da Var (A)

$$\text{Var (A)} = 4 \text{ Var } (\overline{RC}_1) + \text{Var } (\overline{P}_1) + \text{Var } (\overline{F}_1)$$

define-se: ϕ_A = graus de liberdade da variância de A
 $\phi_{\overline{RC}_1}$ = g.l. da variância de \overline{RC}_1 (1)
 $\phi_{\overline{P}_1}$ = g.l. da variância de \overline{P}_1 (1)
 $\phi_{\overline{F}_1}$ = g.l. da variância de \overline{F}_1 (1)
 (1) estes graus de liberdade foram obtidos anteriormente nas análises de variância (Tabela 1).

Aplica-se então a fórmula de Satterthwaite (1946):

$$\frac{[\text{Var (A)}]^2}{\phi_A} = \frac{[4 \text{ Var } (\overline{RC}_1)]^2}{\phi_{\overline{RC}_1}} + \frac{[\text{Var } (\overline{P}_1)]^2}{\phi_{\overline{P}_1}} + \frac{[\text{Var } (\overline{F}_1)]^2}{\phi_{\overline{F}_1}}$$

e obtém-se ϕ_A .

A metodologia de Cavalli (1952) consiste em estimar os parâmetros m, [d] e [h], definidos como $\frac{1}{2}(\overline{P}_1 + \overline{P}_2)$, somas ponderadas dos efeitos aditivos e de dominância, respectivamente, a partir das médias das populações disponíveis. Estas estimativas são então utilizadas para o cálculo do valor esperado de cada população e, por comparação às respectivas médias experimentalmente obtidas, permitem testes qui-quadrados de exatidão para o modelo genético assumido.

Para estimar m, [d] e [h], são necessárias, no mínimo, três populações distintas. No entanto, para que os testes qui-quadrados sejam realizados, pelo menos, quatro populações são necessárias. Os três parâmetros são então estimados pelo método dos quadrados mínimos ponderados, servindo como pesos a razão inversa das variâncias das médias de cada população.

Da Tabela 3 extraem-se as seguintes matrizes:

TABELA 3. "Joint scaling test", Cavalli (1952).

População	Pesos	m	[d]	[h]	Média valor observado
P ₁	w ₁	1	1	0	\overline{P}_1
P ₂	w ₂	1	-1	0	\overline{P}_2
F ₁	w ₃	1	0	1	\overline{F}_1
F ₂	w ₄	1	0	.5	\overline{F}_2
RC ₁	w ₅	1	.5	.5	\overline{RC}_1
RC ₂	w ₆	1	-.5	.5	\overline{RC}_2

a. uma matriz X, chamada matriz de informações,

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0,5 \\ 1 & 0,5 & 0,5 \\ 1 & -0,5 & 0,5 \end{bmatrix}$$

b. uma matriz Y, chamada matriz dos resultados,

$$Y = \begin{bmatrix} \bar{P}_1 \\ \bar{P}_2 \\ \bar{F}_1 \\ \bar{F}_2 \\ \bar{RC}_1 \\ \bar{RC}_2 \end{bmatrix}$$

c. uma matriz W, chamada matriz dos pesos,

$$W = \begin{bmatrix} w_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & w_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & w_6 \end{bmatrix}$$

Das matrizes descritas anteriormente, obtêm-se as equações normais da forma:

$$(X'WX)\hat{\beta} = X'WY$$

onde X' representa a matriz transposta de X e $\hat{\beta}$ representa o vetor dos parâmetros a estimar e tem a seguinte forma:

$$\hat{\beta} = \begin{bmatrix} \hat{m} \\ \hat{d} \\ \hat{h} \end{bmatrix}$$

A teoria da estimação dos quadrados mínimos ponderados nos diz que a solução das equações normais, escritas acima, em notação de matriz, tem a seguinte forma:

$$\hat{\beta} = (X'WX)^{-1} X'WY$$

onde $(X'WX)^{-1}$ representa a matriz inversa de $(X'WX)$. De $(X'WX)^{-1}$ e $(Y'WY - \hat{\beta}'X'WY)$, obtêm-se a variância e os desvios padrões dos estimadores de m, \hat{d} e \hat{h} :

$$(Y'WY - \hat{\beta}'X'WY)(X'WY)^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

A variância de \hat{m} é (a), a variância de \hat{d} é a variância de \hat{g} é (i). Se se representarem os desvios padrões de cada estimador da seguinte maneira, $\sqrt{a} = \pm s_1$, $\sqrt{e} = \pm s_2$ e $\sqrt{i} = \pm s_3$, pode-se escrever $\hat{m} \pm s_1$, $\hat{d} \pm s_2$ e $\hat{h} \pm s_3$, ou seja, tem-se uma medida da precisão dos estimadores.

Obtidos os estimadores anteriormente descritos, pode ser testado, o valor do modelo genético simples em representar os mecanismos de ação gênica nos materiais em estudo. Os valores esperados de cada população em estudo devem ser obtidos usando-se as proporções de m, \hat{d} e \hat{h} respectivas constantes da Tabela 3.

O teste qui-quadrado que envolve as seis populações consideradas, tem a seguinte forma:

$$\chi^2_{g.l.} = \sum_{i=1}^6 [O_i - E_i]^2 w_i$$

onde:

O_i = média observada da i-ésima população

E_i = média esperada da i-ésima população

w_i = peso da i-ésima população

Se as indicações dadas pelos testes propostos acima forem no sentido da rejeição do modelo genético simplificado, duas alternativas podem ser seguidas. Ou o modelo deve ser estendido no sentido de incluir interações não alélicas, ou uma escala apropriada, em que o modelo simples seja adequado, deve ser procurada.

No que diz respeito à avaliação dos efeitos genéticos, a escala deve ser escolhida de modo a não causar vieses na análise, porque, embora seja desejável trabalhar com um modelo genético simplificado, o uso de uma escala apropriada somente irá eliminar interações não alélicas satisfatoriamente se os genes mantiverem a mesma relação uns com os outros. Caso as interações de pares de genes difiram umas das outras, tal mudança de escala pode, no máximo, causar a utilização de uma escala na qual as interações, em média, estão ausentes. Não se deve esperar, no entanto, que este último caso afete os resultados de forma tão sensível quanto o uso de uma escala não apropriada que causaria vieses consideráveis. Em outras palavras, tal escala pode diminuir a precisão das análises genéticas nas quais o modelo simples foi aplicado, mas não deve alterar significativamente as interpretações das estimativas dos parâmetros (Fisher et al. 1932).

Neste trabalho, foram usadas as transformações logarítmica e raiz quadrada sobre os valores individuais, ou seja, $F_1(x) = \log(x)$ e $F_2(x) = \sqrt{x}$, respectivamente.

Outra alternativa que se coloca quando o modelo simples se mostra inadequado em representar os dados relacionados é a inclusão de outros parâmetros no modelo genético, como sugerem Hayman (1958) e Cockerham (1973). O processo de Hayman (1958), pela sua simplicidade na estimativa dos parâmetros a partir das populações disponíveis neste estudo, será apresentado a seguir.

Seis parâmetros que estendem o modelo simplificado anterior para incluir os três tipos possíveis de epistasia digênica, são considerados. Desses seis, três, \hat{m} , \hat{d} e \hat{h} , foram anteriormente considerados e \hat{i} , \hat{j} e \hat{l} são respectivamente estimadores dos efeitos genéticos aditivos x aditivos, aditivo x dominância e dominância x dominância. De acordo com o modelo de Hayman (1958), entretanto, \hat{m} difere do apresentado anteriormente, representando aqui a média da população F_2 . Infelizmente, os valores esperados dos parâmetros não são isentos de interações (epistasia) de ordem outra do que a representada por eles em cada caso. Mesmo d e h não são livres de efeitos epistáticos. Tal fato pode ser consta-

tado a partir das relações apresentadas na Tabela 4, às quais se aplicaria a nomenclatura de Cockerham (1973).

TABELA 4. Estimadores dos parâmetros genéticos, Hayman (1958).

\hat{m}	\bar{F}_2	\bar{RC}_1	\bar{RC}_2
$[\hat{d}]$			
$[\hat{h}]$	$= -\frac{1}{2}\bar{P}_1 - \frac{1}{2}\bar{P}_2 + \bar{F}_1 - 4\bar{F}_2 + 2\bar{RC}_1 + 2\bar{RC}_2$		
$[\hat{i}]$	$= -4\bar{F}_2 + 2\bar{RC}_1 + 2\bar{RC}_2$		
$[\hat{j}]$	$= -\frac{1}{2}\bar{P}_1 + \frac{1}{2}\bar{P}_2 + \bar{RC}_1 - \bar{RC}_2$		
$[\hat{l}]$	$= \bar{P}_1 + \bar{P}_2 + 2\bar{F}_1 + 4\bar{F}_2 - 4\bar{RC}_1 - 4\bar{RC}_2$		

Para os testes de significância dos vários estimadores, a variância de cada um deles deve ser calculada. Assim,

Var (\hat{m}) = Var (\bar{F}_2)

Var ($[\hat{d}]$) = Var (\bar{RC}_1) + Var (\bar{RC}_2)

Var ($[\hat{h}]$) = $\frac{1}{4}$ Var (\bar{P}_1) + $\frac{1}{4}$ Var (\bar{P}_2) + Var (\bar{F}_1) + 16 Var (\bar{F}_2) + 4 Var (\bar{RC}_1) + 4 Var (\bar{RC}_2)

Var ($[\hat{i}]$) = 16 Var (\bar{F}_2) + 4 Var (\bar{RC}_1) + 4 Var (\bar{RC}_2)

Var ($[\hat{j}]$) = $\frac{1}{4}$ Var (\bar{P}_1) + $\frac{1}{4}$ Var (\bar{P}_2) + Var (\bar{RC}_1) + Var (\bar{RC}_2)

Var ($[\hat{l}]$) = Var (\bar{P}_1) + Var (\bar{P}_2) + 4 Var (\bar{F}_1) + 16 Var (\bar{F}_2) + 16 Var (\bar{RC}_1) + 16 Var (\bar{RC}_2)

Os testes obedecem ao modelo abaixo:

$t_{k_1} = \frac{\hat{m}}{\sqrt{\text{Var}(\hat{m})}}$, onde k_1 são os graus de liberdade do teste t.

$t_{k_2} = \frac{\hat{d}}{\sqrt{\text{Var}(\hat{d})}}$, etc...

Os cálculos de graus de liberdade para o teste t, proposto acima, foram desenvolvidos anteriormente.

Outra metodologia que versa sobre estimativa de efeitos genéticos e que, por seu caráter geral, ocupa lugar de destaque na literatura, é a desenvolvida por Cockerham (1973).

Cockerham (1973) generalizou o modelo fatorial para

dois locos de Kempthorne (1957), aplicando-o a um número qualquer de locos. Tal modelo é aplicável a indivíduos, cruzamentos entre indivíduos, retrocruzamentos, autofecundações etc. e sumariza os efeitos genéticos, mantendo a descrição quanto à origem.

Modelo:

$$G = \mu + \sum \alpha_i A_i + \sum \delta_{ij} D_{ij} + (\sum \alpha_i A_i)^2 + (\sum \alpha_i A_i) (\sum \delta_{ij} D_{ij}) + (\sum \delta_{ij} D_{ij})^2 + \dots$$

onde:

- G = genótipo dos indivíduos em consideração
- μ = média
- $\sum \alpha_i A_i$ = efeito aditivo, onde α_i é a proporção, nos indivíduos considerados, dos genes provenientes do i-ésimo pai e A_i é a soma dos efeitos aditivos dos genes do i-ésimo pai.
- $\sum \delta_{ij} D_{ij}$ = efeito de dominância, onde δ_{ij} (δ_{ij}) é a proporção de genótipos, por loco, com alelos dos pais i-ésimo e j-ésimo
- $(\sum \alpha_i A_i)^2$ = $\sum \alpha_i^2 (AA)_{ii} + 2 \sum_{i < j} \alpha_i \alpha_j (AA)_{ij}$ = efeito epistático, aditivo x aditivo
- $(\sum \alpha_i A_i) (\sum \delta_{ij} D_{ij})$ = efeito epistático, aditivo x dominância
- $(\sum \delta_{ij} D_{ij})^2$ = efeito epistático, dominância x dominância.

Segundo o modelo descrito acima, as populações deste trabalho podem ser representadas como se segue. O modelo foi expandido para envolver as interações epistáticas entre dois locos, mas não os de ordem mais alta.

$F_1 = G_1 = \mu + 2A_1 + D_{11} + 4(AA)_{11} + 2(AD)_{(11)}$

$F_2 = G_2 = \mu + 2A_2 + D_{22} + 4(AA)_{22} + 2(AD)_{(22)}$

$F_1 = G_{12} = \mu + A_1 + A_2 + D_{12} + 2(AA)_{11} + 2(AA)_{22} + (AD)_{(12)}$

$F_2 = (G_{12})^{AB} = \mu + A_1 + A_2 + \frac{1}{4}D_{11} + \frac{1}{2}D_{12} + \frac{1}{4}D_{22} + (AA)_{11} + 2(AA)_{12} + (AA)_{22} + \frac{1}{4}(AD)_{(11)} + \frac{1}{2}(AD)_{(12)} + \frac{1}{4}(AD)_{(22)} + \frac{1}{4}(AD)_{(11)(12)} + \frac{1}{4}(AD)_{(12)(11)} + \frac{1}{16}(DD)_{(11)(11)} + \frac{1}{4}(DD)_{(11)(12)} + \frac{1}{4}(DD)_{(12)(11)} + \frac{1}{16}(DD)_{(12)(12)} + \frac{1}{8}(DD)_{(11)(12)}$

$RC_1 = (G_{12})_1 = \mu + \frac{3}{2}A_1 + \frac{1}{2}A_2 + \frac{1}{2}D_{11} + \frac{1}{2}D_{12} + \frac{9}{4}(AA)_{11} + \frac{3}{2}(AA)_{12} + \frac{1}{4}(AA)_{22} + \frac{3}{4}(AD)_{(11)} + \frac{1}{4}(AD)_{(12)} + \frac{3}{4}(AD)_{(22)} + \frac{1}{4}(DD)_{(11)(11)} + \frac{1}{2}(DD)_{(11)(12)} + \frac{1}{4}(DD)_{(12)(11)} + \frac{1}{4}(DD)_{(12)(12)}$

$RC_2 = (G_{12})_2 = \mu + \frac{1}{2}A_1 + \frac{3}{2}A_2 + \frac{1}{2}D_{11} + \frac{1}{2}D_{12} + \frac{1}{4}(AA)_{11} + \frac{3}{2}(AA)_{12} + \frac{9}{4}(AA)_{22} + \frac{1}{4}(AD)_{(11)} + \frac{3}{4}(AD)_{(12)} + \frac{1}{4}(AD)_{(22)} + \frac{1}{4}(DD)_{(11)(11)} + \frac{1}{2}(DD)_{(11)(12)} + \frac{1}{4}(DD)_{(12)(11)} + \frac{1}{4}(DD)_{(12)(12)}$

O problema de estimação, decorrente do enorme número de parâmetros em comparação ao número de populações usualmente disponíveis, pode ser facilmente visualizado. Cockerham (1973) identificou algumas restrições a serem impostas ao seu modelo, com o intuito de possibilitar a estimativa de alguns parâmetros.

O caráter geral deste modelo torna a sua apresentação obrigatória num trabalho desta natureza. Nos casos em que o modelo genético simples explica satisfatoriamente os mecanismos de ação gênica do caráter em estudo, este modelo também tem aplicação direta. Testes de escala são facilmente construídos para o teste de tais modelos simples. As aplicações práticas do modelo de Cockerham serão objeto de análise em outro trabalho.

RESULTADOS E DISCUSSÃO

As médias e variâncias, referentes a dias para floração das populações analisadas neste trabalho, assim como o número de plantas consideradas em cada caso, encontram-se na Tabela 5.

Considerou-se, a princípio, que seria necessário proceder à transformação dos dados, uma vez que associações do tipo média-variância nas populações P_1 e P_2 estariam ocorrendo (Curtiss 1943). A primeira mudança de escala tentada foi a transformação para logaritmos que, embora satisfatória do ponto de vista de homogeneidade das variâncias de P_1 , P_2 e F_1 , gerações não segregantes, não permitiu a interpretação dos efeitos genéticos de acordo com um modelo de ação gênica simplificado. Os testes de escala de Mather (1949), indicaram a presença de interações não alélicas que complicariam os estudos. Embora uma escala logarítmica, na maioria das vezes, seja, pela lógica, adaptável a dados obtidos a partir de seres vivos, o presente conhecimento da ação dos genes não permite a confirmação de tal hipótese.

Partindo desta última premissa, diversos autores preconizaram o uso de transformações em dados biológicos a partir de testes estatísticos e genéticos específicos. A escala na qual as medidas são tomadas para os propósitos de análise genética, é obtida, então, por métodos empíricos. Com base na opinião desses autores, utilizou-se a transformação raiz quadrada. Esta nova transformação também trouxe problemas que tornaram difícil a análise dos resultados. Dessa maneira, embora conscientes da não-homogeneidade das variâncias de P_1 , P_2 e F_1 , resolveu-se manter, para efeito destas análises, a escala original.

TABELA 5. Médias e variâncias, por população, para dias para florescimento, sob condições de dias curtos.

População	Média	Variância fenotípica	Número de plantas
P_1 (D72-7842)	28,00	2,0842	60
P_2 (Santa Maria)	51,03	5,4789	60
F_1	33,21	1,9821	24
F_2	36,45	20,7105	225
RC_1	30,66	4,3878	120
RC_2	42,38	19,9553	120

As populações não segregantes, mencionadas anteriormente, possuem individualmente distribuição normal. Para contornar, pelo menos parcialmente, o problema da não-homogeneidade das variâncias, propôs-se a utilização, para cálculo da variância de ambiente, de uma média ponderada das gerações descritas anteriormente. A ponderação que melhores resultados apresentou foi a que considerou a proporção dos genótipos homocigotos e heterocigotos nas gerações segregantes. Por exemplo, para uso da população F_2 na estimativa de parâmetros genéticos, a ponderação utilizada para obter uma estimativa de sua variância de ambiente envolveu pesos 1:2:1 para as variâncias de P_1 , F_1 e P_2 , respectivamente. Esta ponderação está implícita na metodologia de Warner (1952), usada para o cálculo da herdabilidade em sentido restrito.

Os resultados dos testes de avaliação do modelo genético simples, que envolvem somente efeitos aditivos e de dominância, acham-se apresentados na Tabela 6. Os parâmetros A, B e C são funções unicamente de interações não alélicas, epistasia. Analisando-se a Tabela 6, ficou evidente que A, B e C não diferiram significativamente de zero, uma vez que as médias são menores que os respectivos desvios padrões. Estes testes mostraram não haver razões para a rejeição do modelo genético simplificado assumido. Este fato permite, então, o uso da metodologia de Warner (1952) para cálculo da herdabilidade.

A herdabilidade no sentido restrito foi de 86,46%. Este resultado veio confirmar algumas observações anteriores de Kiihl⁴ e Brim (1973).

⁴ Dados não publicados.

TABELA 6. Médias, desvios padrões e teste do modelo genético (teste de escala), dias para florescimento, sob condições de dias curtos.

População	Média	Desvio padrão
P ₁	28,00	0,1864
P ₂	51,03	0,3022
F ₁	33,21	0,2874
F ₂	36,45	0,3034
RC ₁	30,66	0,1912
RC ₂	42,38	0,4078
A	0,1083 ns	0,5134
B	0,5084 ns	0,9160
C	0,3633 ns	1,3889

A estimativa do número de fatores (método de Mather 1949) indicou que, pelo menos, três pares de genes controlam o caráter dias para floração em soja.

O grau de dominância dos genes em ação indicou haver dominância parcial dos genes para floração precoce sobre os genes para floração tardia (g.d. = 0,30).

A Tabela 7 traz os resultados de outro teste para o modelo genético em questão, o teste de Cavalli (1952), que também possibilitou uma estimativa dos parâmetros genéticos envolvidos no modelo. Os valores observados para as médias das populações não diferiram significativamente dos valores estimados e resultaram num qui-quadrado de valor 0,3386. Desta maneira, o teste de Cavalli (1952) confirmou os resultados do teste anterior de Mather (1949), e os parâmetros m , $[d]$ e $[h]$ puderam ser satisfatoriamente estimados. Outra estimativa do grau de dominância dos genes em ação foi obtida, em seguida, e também indicou existir dominância parcial dos genes para floração precoce sobre os genes para floração tardia $\left[\frac{[h]}{[d]} = 0,54 \right]$. As duas estimativas obtidas diferem bastante entre si. Cumpre ressaltar que esta última, baseada em estatísticas de 1º grau, pode ser afetada por cancelamento de efeitos positivos e negativos dos alelos dos genes envolvidos; por isso, a primeira estimativa deve ser preferida.

Com o intuito de ilustrar os casos em que uma escala adequada (para a descrição dos fenômenos genéticos de acordo com o modelo simplificado) não seja encontrada, ensaiou-se, com as mesmas

TABELA 7. Teste do modelo genético ("joint scaling test"), estimativa de parâmetros e médias de populações e do grau de dominância, dias para florescimento, sob condições de dias curtos.

População	Média (valor observado)	Média (valor estimado)	Desvio padrão
P ₁	28,00	28,01	
P ₂	51,03	51,10	
F ₁	33,21	33,28	
F ₂	36,45	36,42	
RC ₁	30,66	30,65	
RC ₂	42,38	42,19	
\hat{m}		39,5534 ^{ns}	0,1657
$[\hat{d}]$		- 11,5472 ^{ns}	0,1633
$[\hat{h}]$		- 6,2694 ^{ns}	0,3103
$\frac{1}{2}(\bar{P}_1 + \bar{P}_2)$	39,52		
χ^2 3g.l.			0,3386 ns
$\frac{[\hat{h}]}{[\hat{d}]}$			0,54

gerações disponíveis, o uso da metodologia proposta por Hayman (1958). Tal modelo estende a gama dos efeitos genéticos considerados até interações entre dois locos. Assume-se, na construção de tal modelo, a livre recombinação entre os dois locos considerados.

As estimativas \hat{m} , $[\hat{d}]$, $[\hat{h}]$, $[\hat{i}]$, $[\hat{j}]$ e $[\hat{l}]$, realizadas de acordo com a metodologia de Hayman (1958), encontram-se na Tabela 8, juntamente com os respectivos desvios padrões. Os estimadores \hat{m} , $[\hat{d}]$ e $[\hat{h}]$ diferiram significativamente de zero, ao nível de 1%, conforme indicam os testes t:

$$t_{m, 223} = 120,15^{**}$$

$$t_{[\hat{d}], 166} = - 26,01^{**}$$

$$t_{[\hat{h}], 2493} = - 3,91^{**}$$

onde m , $[d]$, $[h]$ indicam os parâmetros testados e os números que os seguem, o número de graus de liberdade para cada teste.

Os estimadores dos efeitos epistáticos, $[i]$, $[j]$ e $[l]$ são menores que os respectivos desvios pa-

TABELA 8. Estimativas dos parâmetros do modelo genético epistático e respectivos desvios padrões, dias para florescimento, sob condições de dias curtos.

Parâmetro estimado	Média	Desvio padrão
m	36,4533	0,3034
[d]	- 11,7167	0,4504
[h]	- 6,0550	1,5486
[i]	0,2534	1,5114
[j]	- 0,2001	0,4841
[l]	- 0,8001	2,2749

drões e, portanto, não significativos. Os resultados traduzidos pelas estimativas descritas anteriormente estão de pleno acordo com os obtidos através das metodologias anteriores. O modelo de Hayman (1958), entretanto, na ausência de epistasia, resulta em estimativas de m, [d] e [h] que são menos precisas que as anteriormente obtidas segundo o método de Cavalli (1952). Para os casos em que os efeitos epistáticos estejam presentes, a metodologia de Hayman (1958) fornecerá resultados mais confiáveis. A metodologia de Cavalli (1952), embora ainda possa ser usada, deve ter seus resultados examinados com maior cautela, uma vez que o uso dos parâmetros m, [d] e [h], obtidos pelo método de Cavalli (1952), em situações em que efeitos epistáticos estão presentes, pode acarretar alguns problemas de estimação. Isto ocorre porque os efeitos d ou h de cada gene não serão aditivos, como em situações nas quais o modelo genético simplificado se aplica. Note-se que a diferença conceitual entre os parâmetros m de Cavalli (1952)

$$\left[m = \frac{1}{2} (\bar{P}_1 + \bar{P}_2) \right] \text{ e } m \text{ de Hayman (1958)} \\ \left[m = F_2 \right]^2.$$

Outra metodologia que versa sobre estimativa de efeitos genéticos e que, por seu caráter geral, ocupa lugar de destaque na literatura, é a desenvolvida por Cockerham (1973).

Cockerham (1973) generalizou o modelo fatorial para dois locos de Kempthorne (1957), aplicando-o a um número qualquer de locos. Tal modelo é aplicável a indivíduos, cruzamentos entre indivíduos, retrocruzamentos, autofecundações etc. e sumariza os efeitos genéticos, mantendo a descrição quanto à origem.

CONCLUSÕES

1. As tentativas de mudança de escala não levaram a uma solução satisfatória para o problema da desigualdade de variância entre populações não segregantes. Resolvido o problema da homogeneidade de variância entre P_1 , P_2 e F_1 , apareciam efeitos genéticos epistáticos, por exemplo. A escolha da escala original para as análises foi consequência de diversos fatores, mas ocorreu principalmente devido à possibilidade de adaptar um modelo genético simplificado ao fenômeno em estudo, o que viabilizou a aplicação de fórmulas do cálculo de herdabilidade, do número de fatores, do grau de dominância etc. O modelo encontrado adaptou-se bem à realidade, e seu valor de predição foi indisputável.

2. Para o cruzamento da linhagem D 72-7842 com a cultivar Santa Maria, ficou evidente a importância dos efeitos aditivos como principais componentes dos efeitos genéticos e a ocorrência de dominância parcial no sentido da floração mais precoce.

3. Quanto à aplicação dos resultados obtidos neste trabalho, vale notar que o modelo genético simples, embora usado com sucesso no caso de cruzamento da linhagem D 72-7842 com a cultivar Santa Maria, não pode ser usado para outros cruzamentos que envolvem diferentes linhagens e ou variedades de soja, sem que se proceda novamente aos testes mostrados neste trabalho. Mather & Jinks (1971) apresentam vários exemplos de cruzamentos entre linhagens de tomate que mostram mecanismos de ação gênica diferentes para o mesmo caráter em estudo. Este fato reforça a necessidade da existência de trabalhos de revisão e adaptação das metodologias existentes para estudos genéticos que as tornem de aplicação fácil e direta. Outra metodologia a ser estudada e possivelmente tornada mais acessível, será a análise de experimentos com cruzamentos em dialelo. O objetivo final seria o de conseguir, através de diversos métodos e experimentos, informação suficiente para chegar a conclusões válidas quanto aos mecanismos genéticos em controle de característica em estudo.

4. No caso específico deste estudo do modelo de ação gênica em controle de dias para floração em soja, sob condições de dias curtos, as informa-

ções, aqui obtidas, serão, em futuro próximo, co-
tejadas com outras provenientes de novos experi-
mentos, para que, no final, seja possível obter uma
compreensão razoável deste mecanismo.

REFERÊNCIAS

- BRIM, C.A. Quantitative genetics and breeding. Madison, Wisc., Am. Soc. of Agr. Inc., 1973. v.16. p.155-83.
- CAVALLI, L.L. An analysis of linkage in quantitative inheritance. In: RIEVE, E.C. & WADDINGTON, C.H. Quantitative inheritance. London, HMSO, 1952. p.135-44.
- COCKERHAM, C.C. Random vs. fixed effects in plant genetics. Raleigh, N.C., Dep. of Statistics, N.C.S.U. Paper Unpublished. 1973.
- CURTISS, J.H. On transformations used in the analysis of variance. *Ann. Math. Stat.*, 14:107-22, 1943.
- FISHER, R.A.; IMMER, F.R. & TEDIN, U. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics*, 17: 107-24, 1932.
- HAYMAN, B.I. The separation of epistatic from additive and dominance variation in generation means. *Heredity*, 12:371-90, 1958.
- KEMPTHORNE, O. An introduction to genetics statistics. New York, John Wiley and Sons, 1957.
- MATHER, K. Biometrical genetics. London, Methuen, 1949. n.p.
- MATHER, K. & JINKS, L.L. Biometrical genetics; the study of continuous variations. Ithaca, NY, Cornell University, 1971. n.p.
- SATTERTHWAITE, F. An approximate distribution of estimates of variance components. *Biom. Bull.*, 2: 110-4, 1946.
- WARNER, J.N. A method of estimating heritability. *Agron. J.*, 44:427-30, 1952.