

LUCAS KOSHY NAOE

**SELEÇÃO PARA ALTO TEOR PROTÉICO EM SEMENTES DE SOJA
EM POPULAÇÕES DE RETROCRUZAMENTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2004

LUCAS KOSHY NAOE

**SELEÇÃO PARA ALTO TEOR PROTÉICO EM SEMENTES DE SOJA
EM POPULAÇÕES DE RETROCRUZAMENTOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 10 de fevereiro de 2004.

Prof. Sebastião Martins Filho

Prof. Paulo Sávio Lopes

Prof. Everaldo Gonçalves de Barros
(Conselheiro)

Prof. Maurílio Alves Moreira
(Conselheiro)

Prof. Carlos Sigueyuki Sedyama
(Orientador)

Esta dissertação é proibida para quem sabe conceituar gene;
É proibido para quem sabe melhoramento;
É proibida para quem entende da cultura da soja;
É proibido para quem não quer ler aqui o que já sabe.

Não escrevi pesquisando;
Parti das experiências de alguns professores e dos amigos;
Nestas páginas não escrevi uma tese e nem sobre a soja;
Tese é algo científico e a soja tem gosto de mato.

Não escrevi ou descrevi sobre coisa alguma;
Confesso: ‘O que eu almejei é impossível’;
Mas, fundamentei algo na prática;
Aos familiares *in memoriam*, dedico;
Aos tios, tias, primos e primas, ofereço.

AGRADECIMENTO

À Universidade Federal de Viçosa, pela oportunidade da realização do Curso de Doutorado em Genética e Melhoramento.

Ao Professor Carlos Siguelyuki Sedyama, pela orientação circunspeta e amigável, que auxiliaram não só nos estudos como também na vida.

Ao Professor Maurílio Alves Moreira, este que persegui durante toda formação acadêmica, por seus conselhos, opiniões e sugestões concisas.

Ao Professor Everaldo Gonçalves Barros, pela grande reciprocidade e agudeza de espírito, pois ânimo nunca faltou.

Ao Professor Paulo Sávio Lopes, pela paciência e compreensão, sem a sua ajuda muitas dúvidas persistiriam.

À Professora Carmen Silva Pereira e ao Professor Tuneo Sedyama pelos seus exemplos, a minha deferência e homenagem.

Aos Professores Múcio Silva Reis e Pedro Crescêncio Souza Carneiro, pela pronta disposição, assegurando o sucesso do trabalho.

Ao Cupertino, José Carlos, Marcos, Paulinho, Jander, Aloísio e Inês, por muitas vezes recônditos, pela importância de seus trabalhos.

À Célia, Daniel, Germano e Poliane, Leandro, Marcio, e Rodrigo companheiros de república, digo lar, agradeço.

A Newton Deniz Piovesan pelos conselhos e ajuda, pela amizade construída no trabalho e na república.

Aos amigos Takahashi, Oikawa, Mauren, Rodrigo Fanny, Dirlane, Gisele e Lucy que apesar da distância, nunca me esqueceram.

Aos discípulos Fabio, Lucimara, Fernanda, Lucinete, Vagner, Inês, Isaura, Dirce, Juliana, Gustavo, Vilmar, Tiago, Ronaldo, Willian, Geraldo, Adriano e Priscila, Júpiter e Alessandra pela convivência tão agradável.

Aos companheiros Odilon Lemos de Mello Filho e Marcelo Antônio Tomaz pela amizade e alegria.

À Elisangela de Oliveira, Adeniz Macedo e Alessandra Maria Lima pela compreensão e carinho.

Aos familiares por compreenderem que a minha persistente ausência, não foi indiferença.

À você, esquecido agora, peço minhas sinceras desculpas, a sua ausência me extenua

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais pela concessão da bolsa de estudo.

BIOGRAFIA

Lucas Koshy Naoe, filho de lavradores, Mamoru Naoe e Izabel Hideko Naoe, nasceu no dia 21 de setembro de 1974, em Lucélia – SP.

Em abril de 1992, ingressou no curso de engenharia agrônômica da Universidade Federal de Viçosa, graduando-se em fevereiro de 1997.

Em outubro de 1994, foi contemplado com uma bolsa de iniciação científica, para desenvolver trabalhos na área de melhoramento da soja para consumo humano, no Instituto de Biotecnologia Aplicada à Agropecuária – BIOAGRO / UFV, permanecendo até janeiro de 1997.

Iniciou o curso de mestrado em Genética e Melhoramento nesta mesma Universidade em março de 1997, defendendo tese em 22 de setembro de 1999.

Em fevereiro de 2000, ingressou no programa de doutorado em Genética e Melhoramento nessa mesma instituição, defendendo tese em 10 de fevereiro de 2004.

Em março de 2004, foi contratado como professor titular pela Fundação Universidade do Tocantins – UNITINS para colaborar na formação da Organização Estadual de Pesquisa Agropecuária do Tocantins.

CONTEÚDO

RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	4
2.1. Melhoramento da soja e o método de retrocruzamento	4
2.2. Melhor preditor não viesado	8

CAPÍTULO 1

SELEÇÃO PARA TEOR PROTÉICO EM SOJA BASEANDO-SE NA MELHOR PREDIÇÃO LINEAR NÃO-VIESADA

1. RESUMO	12
2. INTRODUÇÃO	13
3. MATERIAL E MÉTODOS	15
3.1. Campos experimentais	15
3.2. Material genético	16
3.2.1. Progenitores	16
3.2.2. Gerações de autofecundação	16
3.2.2.1. Geração F1	16

3.2.2.2.	Geração F2	17
3.2.2.3.	Geração F3 e RC1F1	18
3.2.2.4.	Geração RC1F2	19
3.2.2.5.	Geração RC1F3 e RC2F1	19
3.2.2.6.	Geração RC2F2	20
3.3.	Análises genético-estatísticas	21
3.3.1.	Equações de modelos mistos	21
3.3.2.	Herdabilidade no sentido restrito	25
3.3.3.	Respostas esperadas	27
4.	RESULTADOS E DISCUSSÃO	29
5.	CONCLUSÕES	45

CAPÍTULO 2

DISTRIBUIÇÃO DE FREQUÊNCIA DO TEOR DE PROTEÍNA NAS SEMENTES DE POPULAÇÕES DE RETROCRUZAMENTO EM SOJA

1.	RESUMO	46
2.	INTRODUÇÃO	47
3.	MATERIAL E MÉTODOS	49
3.1.	Material genético e município	49
3.2.	Análise dos dados	50
4.	RESULTADOS E DISCUSSÃO	53
5.	CONCLUSÕES	56

REFERÊNCIAS BIBLIOGRÁFICAS	57
----------------------------------	----

APÊNDICE	64
----------------	----

RESUMO

NAOE, Lucas Koshy, D. S., Universidade Federal de Viçosa, fevereiro de 2004.
Seleção para alto teor protéico em sementes de soja em populações de retrocruzamentos. Orientador: Carlos Siqueyuki Sedyama. Conselheiros: Everaldo Gonçalves de Barros e Maurílio Alves Moreira.

O melhoramento vegetal é um processo longo e caro, de modo que quaisquer procedimentos metodológicos que auxiliem na tomada de decisão durante a seleção das linhagens mais promissoras são extremamente importantes. Este trabalho teve como objetivo avaliar as distribuições das frequências das populações de retrocruzamento de soja, bem como selecionar as prováveis linhagens promissoras para alto teor de proteína nas sementes. Utilizando-se de um modelo linear misto há possibilidade de se fazer a predição de efeitos aleatórios, na presença de efeitos fixos, por meio dos BLUP's (best linear unbiased prediction) que são de grande valia no melhoramento. Os experimentos foram conduzidos em três municípios do Estado de Minas Gerais, Capinópolis, Oratórios e Viçosa entre os anos de 1999 e 2001. Linhagens de soja foram obtidas pelo método de retrocruzamentos utilizando como progenitores recorrentes as variedades OCEPAR 13, COODETEC 201, COODETEC 202, COODETEC 205, como progenitores doadores para alto teor de proteína nas sementes, as linhagens BARC 8 e BR 80 14887. O método utilizado para

determinação do teor de proteína das sementes foi o Kjeldahl. Concluiu-se que as estimativas de componentes de variância, assim como as estimativas de herdabilidade, baseada na máxima verossimilhança restrita, apresentaram-se com magnitudes baixas; selecionando-se somente entre as linhagens que apresentaram maiores valores preditos pelo BLUP, houve aumento do parentesco na população e; a utilização do BLUP para auxiliar seleção de linhagens promissoras mostrou-se satisfatória. As distribuições das observações e, em consequência, a natureza da variação genética, foi avaliada por meio da análise da curtose e da assimetria das distribuições nas populações RC_2F_2 . Observou-se que a média do teor protéico foi deslocada positivamente, indicando potencial para o desenvolvimento de linhagens promissoras, com maior teor de proteína nas sementes; a variabilidade do teor de proteína dentro das populações depende somente da presença de poucos alelos nas diferentes populações; no segundo retrocruzamento, as populações tendem a apresentar distribuição mesocúrtica ou leptocúrtica e, por meio da estimativa de assimetria, foi possível constatar efeito de dominância em algumas das populações avaliadas.

Abstract

NAOE, Lucas Koshy, D. S., Universidade Federal de Viçosa, February 2004.
High-protein soybean seeds selection in Backcross populations. Advisor: Carlos Sigueyuki Sedyama. Committee Members: Everaldo Gonçalves de Barros and Maurílio Alves Moreira.

The plant breeding is a long and expensive process, so that any methodological procedures that aid taking decision during the selection of the most promising lines is extremely important. This work had as objective to evaluate the frequencies distributions of soybean backcross populations, as well as to select the probable promising lines for high protein content in the seeds. By using a mixed linear model, there is the possibility to predict the aleatory effects, in the presence of fixed effects, through BLUP's (best lineal unbiased prediction) that are of great value in breeding programs. The experiments were accomplished in three locations of the Minas Gerais State: Capinópolis, Oratórios and Viçosa from 1999 to 2001. The soybean lines were obtained by the backcross method, using as recurrent ancestors the varieties OCEPAR 13, COODETEC 201, COODETEC 202, and COODETEC 205, as donors ancestors for high protein content in the seeds, the lines BARC 8 and BR 80 14887. The method used for determination of the seed protein contents was the Kjeldahl. It was concluded

that the estimates of variance components, as well as the heritabilities estimates, based on the maxim restrict likelihood, presented low magnitude; selecting only among the lines that presented larger predicted values by BLUP, it was increased the relationship in the population and; the use of BLUP to aid selection of promising lines was shown satisfactory. The distributions of the observations and, in consequence, the nature of the genetic variation, were evaluated through the analysis of the kurtosis and of the asymmetry of the distributions in the RC_2F_2 populations. It was observed that the protein content average moved positively, indicating potential for the development of promising lines, with higher seed protein content; the variability of the protein content inside the populations depends only upon the presence of few alleles in the different populations; in the second backcross, the populations tend to present mesokurtic or leptokurtic distribution and; through the asymmetry estimate, it was possible to verify dominance effect in some of the appraised populations.

1. INTRODUÇÃO

Um dos meios empregados no melhoramento de plantas é a introdução de genes de linhagens exóticas. Os programas de melhoramento de soja do Brasil, embora o país seja o segundo maior produtor mundial dessa leguminosa, apresentam grande dependência de materiais exóticos uma vez que a base genética do germoplasma brasileiro é estreita. Atualmente, as introduções de material genético exótico estão voltadas para o desenvolvimento de variedades com características especiais para atender demandas específicas.

O método de retrocruzamento é utilizado quando se deseja traspasar apenas um ou poucos genes de uma variedade ou linhagem para uma variedade comercial. É muito usado para a incorporação de resistência a doenças nas variedades em que a suscetibilidade pode comprometer a estabilidade da produção. A variedade em que se deseja incorporar a resistência (recorrente) é cruzada com a variedade ou linhagem doadora (não recorrente). A seguir, a variedade recorrente é utilizada em retrocruzamentos sucessivos com sua descendência, sendo mantida a característica ou o gene desejado do não recorrente.

Na maioria dos programas de retrocruzamento de soja, a seleção em cada ciclo é baseada no rastreamento da característica desejada a ser incorporada em outra variedade, embora em alguns poucos casos sejam utilizados marcadores moleculares para acelerar a recomposição genotípica do recorrente na

descendência e/ou rastreamento da característica alvo (GUIMARÃES e MOREIRA, 1999).

Atualmente, na maioria dos programas de retrocruzamentos, a seleção é baseada apenas em características qualitativas, ou seja, que são determinadas por poucos genes, e tem sido muito eficiente no melhoramento de plantas (JENSEN, 1988). Exemplos recentes são a obtenção das variedades de soja DOKO RC a partir de DOKO (KIIHL et al., 1999) e Cristalina RCH a partir de Cristalina. Entretanto, para características quantitativas, poucos trabalhos têm sido desenvolvidos, sendo que a maior dificuldade reside no estabelecimento de critérios de seleção durante os ciclos de retrocruzamentos.

O problema fundamental na seleção de genitores em programas de melhoramento é obter uma predição do valor genético tão acurada quanto possível, por meio da observação do valor fenotípico. A seleção é eficiente desde que o indivíduo seja comparado com outros indivíduos que foram avaliados nas mesmas condições ambientais, sendo este posteriormente indicado especificamente para o ambiente avaliado.

Na maioria dos casos, os dados utilizados em estudos de melhoramento vegetal são referentes a experimentos que estão localizados em diversas regiões e, portanto, sujeitos a diferentes condições. Nesse contexto, os dados devem ser ajustados para uma base comum em termos de efeitos ambientais identificáveis, como altitude, manejo e solo em que as observações foram coletadas. Esses ajustes são necessários para melhor comparação entre os indivíduos candidatos à seleção, ou seja, uma melhor avaliação genética. Outro problema comum é a perda de tratamentos, repetições ou localidades nos ensaios experimentais, este fato gera desbalanceamentos dos dados.

O uso de metodologias adequadas para a avaliação genética possibilita identificar com maior acurácia as linhagens geneticamente superiores que deverão ser utilizadas no processo de hibridação, permitindo alcançar, em menor espaço de tempo, o almejado progresso genético. A seleção de indivíduos promissores pode ser baseada na melhor predição linear não-viesada do valor genético, a qual é obtida por meio da metodologia de modelos mistos. Essa

metodologia, em substituição aos modelos de análise de variância, foi desenvolvida por HENDERSON (1959, 1963) para analisar dados desbalanceados no melhoramento de gado de leite. A teoria de modelos mistos na seleção de linhagens de soja, para aumentar a produção, foi empregada por PANTER e ALLEN (1995a, 1995b), entretanto, esses autores não utilizaram a matriz de parentesco.

A aplicação desta metodologia requer o conhecimento das estimativas de componentes de variância (SEARLE, 1971), os quais podem ser estimados pelo método da Máxima Verossimilhança Restrita (REML), proposto por PATERSON e THOMPSON (1971), descrito em LOPES et al. (1998).

Esta dissertação está estruturada em dois capítulos. O primeiro capítulo descreve e aplica a metodologia de modelos mistos para a obtenção dos valores genéticos e avalia o ganho obtido. No segundo capítulo, avalia-se a natureza da variação genética das populações RC2F2.

2. REVISÃO DE LITERATURA

2.1. Melhoramento da soja e o método de retrocruzamento

Durante muito tempo, o desenvolvimento de variedades de soja foi baseado na produtividade e teor de óleo. A torta residual era um subproduto que tinha uso restrito. Com o desenvolvimento de grandes indústrias, maior ênfase foi dada à obtenção de variedades altamente produtivas e com alto teor de óleo. Os melhoristas possuem linhagens disponíveis com mais de 50% de proteína, mas com produtividade 10 a 20% abaixo das variedades convencionais (SHIBLES et al., 1975).

Atualmente existem variedades contendo 6 a 8 % a mais de proteína do que variedades antigas e com rendimento de grãos 9 % mais elevado, isto indica a possibilidade de isolar linhagens com alto teor de proteína e elevado rendimento de grãos (SEDIYAMA et al., 1989).

As fontes de genes para alto teor de proteína têm sido de baixas características agrônômicas, tornando difícil o desenvolvimento de linhagens de alto teor de proteína e superiores em rendimento de grãos (HARTWIG, 1973). HANSON (1959), mostrou que na soja blocos gênicos de tamanho considerável permanecem intactos, e concluiu que os métodos de melhoramento da soja têm propiciado poucas possibilidades para a quebra dos blocos gênicos. O melhor

progenitor não recorrente é aquele que, além de ser portador dos alelos desejáveis, não seja seriamente deficiente em outras características agrônômicas desejáveis (FEHR, 1987).

Com o desenvolvimento dos produtos protéicos oriundos da soja para alimentação humana e animais; os programas de melhoramento têm dado maior atenção ao teor protéico na semente da soja. No entanto são relatadas correlações negativas para produtividade e alto teor de proteína. WILCOX e CAVINS (1995), utilizando uma série crescente de retrocruzamentos para teor de proteína, reportaram uma correlação negativa decrescente entre teor de proteína e produção de grãos. Em 1998, Wilcox pesquisando seleção recorrente para aumentar o teor de proteína nos grãos de soja reportou um aumento de variabilidade para teor de proteína à medida que se aumentava o número de ciclos de seleção recorrente.

JOHNSON et al. (1955), estudando populações F4 de soja em diferentes ambientes, reportou correlação fenotípica entre $-0,12$ à $-0,80$ para teor de proteína e produção de grão. KNOW e TORRIE (1964), obtiveram correlações fenotípicas de $-0,58$ e $-0,42$ para teor de proteína e produção. WEHRMANN et al. (1987), reportaram correlação fenotípica entre teor de proteína e produção de $-0,86$ à $-0,56$ em populações obtidas por retrocruzamentos, na tentativa de traspassar genes que conferem alto teor de proteína nas sementes de linhagens exóticas à variedades de produção comercial. HELMS e ORF (1998) concluíram com base em 10 populações diferentes que a seleção para incremento do teor de proteína resulta em decréscimo na produção de grãos.

As estimativas de herdabilidade são importantes para orientar e planejar programas de melhoramento, sendo de fundamental importância para estimar ganhos genéticos e orientar a escolha dos métodos de seleção a serem aplicados. Para que a herdabilidade seja estimada há necessidade que exista variabilidade genética, quando não há variabilidade na população não é possível promover mudança na frequência gênica nem aumentar a média populacional para a característica avaliada (ALLARD, 1960, FALCONER, 1987).

SHORTER et al. (1976) estudaram a herdabilidade no sentido restrito na soja, com base na regressão pai – filho de F₃ em F₄, tendo encontrado estimativas para teor de proteína variando de 21 a 57 %. BRIM e BURTON (1979) utilizando seleção recorrente com o objetivo de aumentar o teor de proteína nas sementes de soja, estimaram as herdabilidades com base na regressão em resposta ao diferencial de seleção e reportaram valores que variaram de 20 a 37 %. ERICKSON et al. (1981) estimaram a herdabilidade, pela regressão pai-filho, para teor de proteína, em quatro populações de soja, utilizando plantas F₂ e médias de famílias F₃ e obtiveram herdabilidade de 27%.

PIOVESAN (2000), objetivando estudar os componentes genéticos envolvidos no controle de características agronômicas na soja por meio de análise dialélica, com base no modelo proposto por Griffing, utilizando genitores e progênies F₁ e F₂, avaliou a herdabilidade para teor protéico que variou de 22 a 81%. SOARES (2000), trabalhando com uma população de linhagens recombinantes endogâmicas (RIL's) de soja na geração F₆, sendo um dos progenitores a linhagem BARC-8, observou que para teor protéico nas sementes os valores de herdabilidades no sentido amplo variaram de 73,16 a 83,11 %, sendo que o coeficiente de variação experimental variou de 2,93 a 3,10 %.

Os métodos de melhoramento conhecidos como retrocruzamento e método genealógico, são os mais indicados para espécies autógamas para traspassar características de interesse. Nos dois métodos, um parental é escolhido por suas boas qualidades agronômicas e o outro é selecionado com base na característica que se deseja traspassar. Dentre os métodos clássicos (genealógico, bulk, SSD, etc) geralmente a seleção é realizada para muitos caracteres durante as gerações de autofecundação em um só local e nem sempre são eficientes, em razão da interação genótipo x ambiente (BORÉM, 1997; DESTRO e MONTALVÁN, 1999). Quando a herança da resistência é poligênica, a seleção de plantas durante o avanço de gerações tem a desvantagem de perder linhagens promissoras que não são mais recuperadas.

O método de retrocruzamento é utilizado quando se deseja traspassar apenas um ou poucos genes de uma variedade ou linhagem para uma variedade

comercial. É muito usado para a incorporação de resistência a doenças e características qualitativas nas variedades. A variedade em que se deseja incorporar a característica de interesse (recorrente) é cruzada com a variedade ou linhagem doadora (não recorrente). A seguir, a variedade recorrente é utilizada em retrocruzamentos sucessivos com sua descendência, sendo mantida apenas a característica ou o gene desejado do não recorrente (JENSEN, 1988). No final do processo de retrocruzamento considerando espécies autógamas, os alelos traspassados estarão em sua maioria na condição heterozigota, o mesmo não ocorrendo com os demais locos. A cada retrocruzamento as linhagens tornam-se cada vez mais homozigotos e a constituição genotípica das progênes aproxima-se gradativamente do progenitor recorrente. Exemplos recentes de variedades de soja obtidas por retrocruzamento são a Garimpo RCH e a COODETEC 201.

O processo tradicional de melhoramento vem sendo cada vez mais auxiliado por técnicas como a transformação de plantas e hibridação somática, permitindo aos melhoristas novas variantes genéticas (BORÉM, 1997; JENSEN, 1988).

A técnica de transformação, ou seja, criação de plantas transgênicas, é uma importante ferramenta para o melhoramento das variedades de soja, pois permite a introdução na planta de genes clonados da mesma ou de outras espécies, incorporando às novas variedades, características agrônomicas desejáveis (BORÉM e SANTOS, 2003). A associação do método de retrocruzamento e transformação de plantas é importantíssimo, pois após conseguido traspassar a característica para uma linhagem específica, este potencial deve ser traspassado rapidamente para as variedades elites.

Trabalhos recentes mostram que a divergência assim estimada pode ser utilizada para substituir a matriz de parentesco tradicionalmente utilizada para a metodologia de modelos mistos (CARNEIRO, 2002). ABDELNOOR et al. (1995), utilizando marcadores moleculares do tipo RAPD em soja, determinou a divergência genética de 36 potenciais genitores. Apesar da possibilidade de aplicação de marcadores moleculares para uma variedade de estudos, o uso da

tecnologia hoje disponível na rotina de programas de melhoramento é pontual e deve ser avaliado criteriosamente para cada situação.

2.2. Melhor preditor linear não viesado

O critério de predição consiste em minimizar o quadrado médio do erro de predição; para o critério de estimação minimiza-se a soma de quadrados dos resíduos. Estimação consiste em obter medidas representativas da população calculadas a partir de uma amostra. Predição está associado à eventos futuros ou à estimação de valores realizados de uma variável aleatória que foi amostrada de uma população com uma estrutura de matriz de variância e covariância conhecidas. Para minimizar o erro de predição, com a restrição de não tendenciosidade do preditor, combina-se a variância do erro de predição com um multiplicador de LaGrange (SCHAFFER, 1993). O termo predição refere-se a fatores aleatórios, enquanto que estimação para fatores fixos.

Considerando-se um vetor de variáveis aleatórios $\mathbf{u} = [u_1, u_2, \dots, u_n]$, conjuntamente distribuído com um vetor $\mathbf{y} = [y_1, y_2, \dots, y_n]$. O Melhor Preditor requer que se conheça $f(\mathbf{u}, \mathbf{y})$ (função de distribuição conjunta das observações e dos valores genéticos) e todos os momentos da distribuição. É não viesado e, de todos os preditores é o que tem o menor quadrado médio do erro de predição.

Sob normalidade, sendo a média e a variância, parâmetros conhecidos, a média condicional de \mathbf{u} é linear em \mathbf{y} . Portanto, sob normalidade, o Melhor Preditor é idêntico ao Melhor Preditor Linear, sendo que o Melhor Preditor Linear não requer conhecimento prévio da $f(\mathbf{u}, \mathbf{y})$, mas exige que se conheça o primeiro e segundo momento (média e variância).

De todos os preditores lineares não viesados, o melhor preditor linear não viesado (BLUP) tem o menor quadrado médio de erro de predição, este preditor assume somente os segundos momentos conhecidos, sendo a média estimada. Se \mathbf{y} não tem distribuição normal, podem existir preditores não lineares que tenham menor quadrado médio do erro do que o BLUP.

O BLUP não prediz um valor de observação (y) de um indivíduo, mas de variáveis aleatórias que têm um papel explicativo. O BLUP pode ser, resumidamente, definido como o resultado da regressão dos efeitos de um fator aleatório em função das observações corrigidas para o efeito fixos (MARTINS et al., 1997). O BLUP não é um método de seleção, é um procedimento para se obter o valor genético dos indivíduos.

O BLUP é assim conhecido porque *best*, melhor, tem o sentido de erro quadrático médio mínimo entre os preditores lineares; *linear*, linear, porque é função linear dos valores de y ; *unbiased*, não viesado, no sentido que o valor esperado do preditor obtido é igual ao valor esperado da variável aleatória que se deseja prever; *predictor*, preditor, profetiza os valores genéticos dos indivíduos. Um método preciso é quando consegue resultados cuja flutuação em torno de um valor médio qualquer é pequena, e um método é acurado quando sua discrepância em relação ao valor verdadeiro é muito pequena. Fatores que influenciam na precisão e acurácia são a herdabilidade, perda parcial da genealogia, modelo usado na avaliação genética, correlação genética, número de gerações (Van der WERF, 1999).

Segundo HENDERSON (1974), a metodologia do modelo misto apresenta as seguintes características: a avaliação do valor genético estimado é não tendenciosa; a avaliação apresenta a propriedade de variância mínima; é fácil de ser modificado quando as condições de avaliação mudam; proporciona estimativas do valor genético, bem como os respectivos erros de estimação; em determinadas situações, elimina erros introduzidos pela ausência de casualização, pela seleção e/ou descarte dos indivíduos; evita em certos casos, o ajuste prévio dos dados, uma vez que as constantes para certos efeitos fixos podem ser obtidas simultaneamente no processo de estimar os valores genéticos.

A metodologia de modelos mistos para obter os preditores dos valores genéticos (“Best Linear Unbiased Prediction traduzido como Melhor Predição Linear Não-Viesada - BLUP), em substituição aos modelos de análise de variância, foi desenvolvida por HENDERSON (1959, 1963) para lidar com dados desbalanceados no melhoramento de gado de leite. A teoria de modelos

mistos na seleção de linhagens de soja, para aumentar a produção, foi empregada por PANTER e ALLEN (1995a, 1995b), entretanto estes autores não utilizaram a matriz de parentesco.

A aplicação da metodologia requer o conhecimento das estimativas de componentes de variância (SEARLE, 1971). O primeiro estudo dos componentes de variância relacionada à genética quantitativa está no trabalho de FISHER (1918) e, posteriormente, em seu livro em 1925, contribuiu para o método de estimação de componentes de variância por meio da análise de variância. Os estimadores de máxima verossimilhança foram utilizados por HERBACH (1959), para dados balanceados; entretanto, apenas em 1967, HARTLEY e RAO reportaram a derivação do método da máxima verossimilhança para estimação de componentes de variância.

Dentre as diversas metodologias disponíveis, destacam-se os métodos I, II e III de Henderson (HENDERSON, 1953); o método de estimação quadrática não-viesada de norma mínima – MINQUE (RAO, 1970), o método de estimação quadrática não-viesada de mínima variância – MIVQUE (RAO, 1971), o método da máxima verossimilhança – ML (HARTLEY e RAO, 1967) e o método da máxima verossimilhança restrita – REML (PATTERSON e THOMPSON, 1971).

Os componentes de variância estimados pela REML coincidem com os obtidos na análise de variância para dados balanceados, além disso, pelo método REML não ocorre o ‘viés’ resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo (ANDERSON, 1984). Este ainda ressalta que o método de estimação elimina o viés atribuído às mudanças nas frequências gênicas, se o parentesco entre os indivíduos for considerado. BOLDMAN et al. (1995) desenvolveram um conjunto de programas para estimar os componentes de variância e covariância baseados em modelos animais, denominado MTDFREML (“Multiple trait derivative-free restricted maximum likelihood”).

Segundo PATTERSON e THOMPSON (1971), o REML baseia-se na maximização da função densidade de probabilidade da parte referente aos efeitos aleatórios em relação aos componentes de variância. Desse modo, cada observação é dividida em duas partes independentes, uma referente aos efeitos

fixos e a outra aos efeitos aleatórios. Portanto, as estimativas desses efeitos são independentes. BOLDMAN e VAN VLECK (1991) reportaram que a estimação de componentes de (co)variância com a utilização do REML, geralmente é considerado melhor método no caso de dados desbalanceados.

PANTER e ALLEN (1995a) utilizaram o BLUP na identificação de linhagens superiores em cruzamentos de soja. Estes estudaram a eficiência dos métodos de mínimos quadrados ordinários e o BLUP, do qual ainda simularam situações de desbalanceamento e concluíram que o BLUP apresenta menor erro de predição e maiores correlações entre os valores preditos e os valores fenotípicos.

BERNARDO (1996) salientou que o BLUP no melhoramento vegetal pode prever o desempenho genético de híbridos que não estão presentes ou que tenham sido perdidos. Em seu trabalho com híbridos de milho, comparou os resultados obtidos via BLUP empregando a genealogia e dados de marcadores moleculares do tipo RFLP (restriction fragment length polymorphism). Observou que não foram verificadas alterações significativas nos resultados obtidos usando as genealogias ou os dados de marcadores moleculares. IEMMA (2003), também trabalhando com milho e marcador do tipo RFLP, obteve correlações moderadas, observando algumas imprecisões.

ANDRÉ (1999), trabalhando com dialelo de feijão, utilizando dados reais e com simulações e ainda auxiliado por marcador molecular do tipo RAPD (random amplified polymorphic DNA), concluiu que o BLUP seria mais recomendado que o método dos mínimos quadrados ordinários para predição das capacidades de combinação e do efeito do cruzamento.

CAPÍTULO 1

SELEÇÃO PARA TEOR PROTÉICO EM SOJA BASEANDO-SE NA MELHOR PREDIÇÃO LINEAR NÃO-VIESADA

1. RESUMO

Foram utilizadas, neste trabalho, linhagens de soja oriundas de populações originadas de retrocruzamentos de quatro variedades COODETEC 201, COODETEC 202, COODETEC 205 e OCEPAR 13 (progenitores recorrentes), tendo sido utilizados como progenitores doadores as linhagens BARC-8 e a BR80-14887 e avaliadas as gerações F2, F3, RC1F2, RC1F3, RC2F2 e RC2F3. O objetivo foi o de testar a metodologia de modelos mistos para obtenção dos valores genéticos e avaliar o ganho predito e realizado. O modelo considerou como fixos, as diferentes épocas e ambientes de plantio, e como aleatório, os efeitos genéticos. As estimativas de componentes de variância, assim como as estimativas de herdabilidade, baseada na máxima verossimilhança restrita, apresentaram-se condizentes com a literatura; selecionando-se somente nos maiores valores preditos pelo BLUP, haverá aumento do parentesco na população. A utilização do BLUP foi considerada eficiente para seleção de linhagens promissoras.

2. INTRODUÇÃO

O Brasil é o segundo maior produtor mundial de soja, respondendo por cerca de 22% da soja mundial. Atualmente, a soja é semeada em todas as regiões brasileiras e há programas de melhoramento em todo o território nacional com o objetivo de desenvolver variedades até mesmo para os pontos mais distantes do país. Destaca-se atualmente a entrada de grandes empresas internacionais no mercado de sementes brasileiro, desenvolvendo programas próprios de melhoramento e aumentando a competitividade no setor (CÂMARA, 2000). Deste modo, empresas detentoras de genes capazes de gerar algum benefício para a soja, sejam na produção, indústria ou comercialização, levam considerável vantagem nessa corrida tecnológica.

Uma característica que tem sido valorizada na comercialização é o teor protéico nos grãos de soja, em face da exportação da soja ser para a produção de farelo para consumo animal. Alguns programas de melhoramento no exterior têm dado ênfase a esta característica como reportaram WEHRMANN et al. (1987); WILCOX e CAVINS (1995).

Atualmente, na maioria dos programas de retrocruzamentos, a seleção é baseada apenas em características qualitativas, ou seja, que são determinadas por poucos genes, e tem sido muito eficiente no melhoramento de plantas. Entretanto, quando se quer traspasar uma característica quantitativa, poucos trabalhos têm sido desenvolvidos, sendo que a maior dificuldade reside no estabelecimento de critérios de seleção durante os ciclos de retrocruzamentos.

A obtenção do melhor estimador linear não-viesado e do o melhor preditor linear não-viesado, considerando, em separado, efeitos fixos e aleatórios, aproveitando, ainda, a relação de parentesco entre as linhagens avaliadas para a predição do valor genético, pode contribuir de maneira significativa para auxiliar nos programas de melhoramento de soja.

O objetivo desse trabalho foi comparar, por meio da correlação de Spearman, e predição de ganhos, o método tradicional e a metodologia de modelos mistos, na avaliação de linhagens de soja, utilizando dados de teor protéico nas sementes de populações de retrocruzamentos, do programa de melhoramento da qualidade da soja, conduzido no BIOAGRO/ UFV.

3. MATERIAL E MÉTODOS

3.1. Campos experimentais

Os experimentos foram conduzidos em três municípios do Estado de Minas Gerais que possuem as seguintes características:

Capinópolis – situado no Triângulo Mineiro, cujas coordenadas geográficas são 18°41'05" S e 49°34'51" W e, aproximadamente, 564 metros de altitude. A área experimental utilizada foi da Central de Pesquisa e Extensão do Triângulo Mineiro (CEPET) pertencente a Universidade Federal de Viçosa. O clima é classificado como Aw – clima tropical úmido (megatérmico) de savana, com inverno seco e verão chuvoso. O solo onde foi instalado o ensaio é classificado como Latossolo vermelho-escuro distrófico, textura argilo-arenoso.

Oratórios – situado na Zona da Mata Mineira, cujas coordenadas geográficas são 20°24'17" S e 42°48'13" W e, aproximadamente, 494 metros de altitude. A área experimental utilizada foi a do Centro de Pesquisa e Melhoramento da Cana-de-açúcar (CECA) pertencente à Universidade Federal de Viçosa. O clima é Cwa – clima temperado chuvoso (mesotérmico) com inverno seco e verão chuvoso. O solo é classificado como Latossolo Vermelho-Amarelo distrófico, textura argilosa.

Viçosa – situado na Zona da Mata Mineira, cujas coordenadas geográficas são 20°45'20" S e 42°52'40" W e, aproximadamente, 650 metros de altitude. A área experimental utilizada foi a do Campo Experimental Diogo Alves de Mello,

localizado no “campus” da Universidade Federal de Viçosa. O clima é também Cwa. O solo é classificado como Latossolo Vermelho-Amarelo distrófico, textura argilosa.

3.2. Material genético

3.2.1. Progenitores

Foram utilizadas neste trabalho linhagens de soja de ciclo precoce/semiprecoce oriundas de retrocruzamentos das variedades indicadas para plantio comercial no Brasil, utilizadas como progenitores recorrentes COODETEC 201, COODETEC 202, COODETEC 205 e OCEPAR 13 (Apêndice – Quadro 1). Como doadores foram utilizados progenitores de alto teor de proteína, as linhagens BARC 8 (LEFFEL, 1992) e BR 80 14887 (Apêndice – Quadro 2). Os progenitores recorrentes foram desenvolvidos pela Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda (COODETEC) e os progenitores doadores foram desenvolvidos pelo USDA-ARS, nos Estados Unidos, e pela EMBRAPA/Soja, respectivamente. Este trabalho foi desenvolvido junto ao Programa de Melhoramento da soja desenvolvido no Instituto de Biotecnologia de Aplicada à Agropecuária (BIOAGRO), da Universidade Federal de Viçosa.

3.2.2. Gerações de Autofecundação

3.2.2.1. Geração F1

Os cruzamentos iniciais foram realizados em casa de vegetação do BIOAGRO, no mês de fevereiro de 1999. As sementes F1's têm a seguinte genealogia:

- a) PTN I12 x OCEPAR 13;
- b) BR 80 14887 x COODETEC 201;
- c) BR 80 14887 x COODETEC 202;
- d) PTN I12 x COODETEC 205;

O progenitor à esquerda do esquema de cruzamentos sempre foi o genitor materno. A linhagem PTN I12 foi oriunda do cruzamento de (BARC 8 x IAC 12) x IAC 12 . A variedade IAC 12 não apresenta alto teor protéico nas sementes e foi desenvolvida pelo Instituto Agrônomo de Campinas.

As sementes F1 foram plantadas em maio de 1999, sendo o cultivo realizado em vaso, com uma planta. Os cruzamentos oriundos de COODETEC 201, COODETEC 202 e COODETEC 205 foram confirmados por meio de análise de marcadores RAPD, pois não se dispunha de marcador morfológico. A colheita e a debulha foram efetuadas, manualmente, para cada planta, quando 100% das vagens estavam secas para colheita.

3.2.2.2. Geração F2

As sementes F2 obtidas foram utilizadas para análise, não destrutiva, do teor protéico individual das sementes, pelo método de quantificação de proteínas do ácido bicinconínico (BCA), descrito por SMITH et al. (1985).

As sementes F2 foram cultivadas em casa de vegetação do BIOAGRO, no mês de setembro de 1999. Entretanto, somente foram cultivadas 25% das sementes de maior teor protéico de cada cruzamento, em média 50 sementes.

O cultivo foi realizado em vaso, com duas plantas. Durante o florescimento foram realizados cruzamentos em todas as plantas F2 com os

respectivos progenitores recorrentes (retrocruzamento). A colheita e a debulha foram efetuadas, manualmente, para cada planta, quando 100% das vagens estavam secas para colheita. Os cruzamentos foram colhidos individualmente e identificados.

3.2.2.3. Geração F3 e RC1F1

As sementes das plantas F2 formaram as famílias F3. Foi coletada uma amostra de sementes F3 de cada família para análise de proteína pelo método Kjeldahl (IAL, 1985). As amostras de sementes F3 que apresentaram teor de proteína inferior ao progenitor recorrente, foram eliminadas no plantio.

Cada cruzamento foi representado por uma parcela de pelo menos 23 famílias F3, a cada cinco famílias foram intercaladas os progenitores recorrentes correspondentes. O experimento foi instalado no mês de janeiro de 2000, no Campo Experimental Diogo Alves de Mello, localizado no Campus da Universidade Federal de Viçosa. Concomitantemente, foi realizados o cultivo das sementes RC1F1 em casa de vegetação do BIOAGRO, em vaso com plantas individuais. Os retrocruzamentos oriundos de COODETEC 201, COODETEC 202 e COODETEC 205 foram confirmados por meio de análise de marcadores RAPD, pois não se dispunha de marcador morfológico.

As famílias F3 foram colhidas em bulk e uma amostra das sementes F4 de cada bulk foi submetida à análise de teor de proteína pelo método kjeldahl (IAL, 1985).

Folhas das plantas F2 foram coletadas e utilizadas para o calculado da divergência genética entre plantas por meio do método de coeficiente de coincidência simples, baseados nos dados de marcadores moleculares do tipo microsatélites. Os seis bulks das famílias F3 que apresentaram maior teor de proteína e ainda, oriunda de plantas F2 que apresentaram menor divergência em relação ao progenitor recorrente, indicaram as seis plantas F2 selecionadas e conseqüentemente as plantas RC1F1 que continuarão o retrocruzamento.

3.2.2.4. Geração RC1F2

As sementes RC1F2 obtidas foram utilizadas para análise, não destrutiva, do teor protéico individual das sementes, pelo método de quantificação de proteínas do ácido bicinconínico (BCA), descrito por SMITH et al. (1985).

As sementes RC1F2 foram cultivadas em casa de vegetação do BIOAGRO, no mês de abril de 2000. Entretanto, somente foram cultivadas 25% das sementes de maior teor protéico de cada cruzamento, em média 50 sementes.

O cultivo foi realizado em vaso, com duas plantas. Durante o florescimento foram realizados cruzamentos em todas as plantas com os respectivos progenitores recorrentes (retrocruzamento). A colheita e a debulha foram efetuadas, manualmente, para cada planta, quando 100% das vagens estavam secas para colheita. Os cruzamentos foram colhidos individualmente e identificados.

3.2.2.5. Geração RC1F3 e RC2F1

As sementes das plantas RC1F2 formaram as famílias RC1F3. Foi coletada uma amostra de sementes RC1F3 de cada família para análise de teor de proteína pelo método kjeldahl (IAL, 1985). As famílias RC1F3 que apresentaram teor de proteína inferior ao progenitor recorrente, foram eliminadas.

Cada cruzamento foi representado por uma parcela de pelo menos 23 famílias RC1F3, a cada cinco famílias foram intercaladas os progenitores recorrentes correspondentes. O experimento foi instalado no dia 2 de outubro de 2000, no Campo Experimental Diogo Alves de Mello. Anteriormente, no dia 10 de setembro de 2000 foram realizados o cultivo das sementes RC2F1 em casa de vegetação do BIOAGRO, em vaso com plantas individuais.

As famílias RC1F3 foram colhidas em bulk e uma amostra das sementes RC1F4 de cada bulk foi submetida à análise de teor de proteína pelo método kjeldahl (IAL, 1985).

Folhas das plantas RC1F2 foram coletadas e utilizadas para o calculado da divergência genética entre plantas por meio do método de coeficiente de coincidência simples, baseados nos dados de marcadores moleculares do tipo microsatélites. Os seis bulks das famílias RC1F3 que apresentaram maior teor de proteína e ainda, oriunda de plantas RC1F2 que apresentaram menor divergência em relação ao progenitor, indicaram as seis plantas RC1F2 selecionadas e conseqüentemente as plantas RC2F1 que continuarão o retrocruzamento.

3.2.2.6. Geração RC2F2

As sementes RC2F2 foram semeadas em casa de vegetação no dia 3 de janeiro de 2001, para obtenção das plantas RC2F2. Cada vaso continha duas plantas. O número de vasos variou de acordo com a disponibilidade de sementes de cada população. A adubação foi a comumente recomendada para a cultura, sendo comum a todas as populações e progenitores. A colheita de cada planta foi efetuada manualmente, quando 100% das vagens estavam secas. O teor protéico das sementes foi determinado em amostras de cinco sementes de cada planta RC2F2, pelo método Kjeldhal (IAL, 1985), no laboratório de análises físico-químicas de alimentos do Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO/UFV).

As sementes RC2F3 foram cultivadas em 10 de maio de 2001 na CEPET e no dia 7 de junho de 2001 na CECA. O preparo do solo foi aquele recomendado para a cultura. A adubação foi realizada no sulco de plantio, conforme a recomendação de cada local. Utilizou-se o delineamento em blocos casualizados, com duas ou três repetições dependendo do número de sementes disponíveis, cultivadas em fileiras de 3 metros com 0,75 metros entre linhas com 12 plantas

por metro linear. O controle de doenças foi de caráter preventivo, com a aplicação do fungicida Captan 75 TS nas sementes, na dose de 160 g por 100 kg de sementes. O controle de plantas daninhas foi realizado manualmente e o controle de pragas obedeceu a um controle conjunto com outros experimentos dos locais. Não houve necessidade de utilização de inoculante em nenhum local. Os tratos culturais foram comuns a todos os cruzamentos e progenitores. As covas foram identificadas após o plantio. A colheita foi realizada manualmente, à medida que as vagens se apresentaram completamente maduras e secas, aproximadamente uma semana após o estágio R8 (FEHR et al., 1971). A trilhadura foi realizada com trilhadeira de plantas individuais. O teor protéico das famílias RC2F3 foi determinado pelo método Kjeldhal (IAL, 1985), no laboratório de análises físico-químicas de alimentos do Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO/UFV).

3.3. Análises genético-estatísticas

Para se fazer a predição dos efeitos aleatórios (valores genéticos), na presença de efeitos fixos, foi utilizada a metodologia de modelos mistos. Apesar da utilização do delineamento em blocos na geração F3, RC1F3 e RC2F3, para o cálculo do BLUP não foi considerado o delineamento. Nessas gerações as fileiras foram colhidas em “bulk” e a cada fileira foi considerada como se fosse um indivíduo.

3.3.1. Equações de modelos mistos

Para a obtenção de preditores BLUP, as matrizes de variâncias e covariâncias genotípica (G) e residual (R) precisam ser conhecidas sem erro; no entanto, esta condição, em geral, não ocorre e elas precisam ser estimadas. O modelo utilizado para estimação de componentes de variância foi uma adaptação

do modelo conhecido como modelo animal (HENDERSON, 1974). O modelo animal assume que somente os efeitos aditivos são de maior importância e que todos os parentescos entre indivíduos são conhecidos.

Nas análises, foi considerado o seguinte modelo estatístico:

$$y_{ij} = \mu + a_i + t_{ij} + e_{ij}$$

em que:

y_{ij} = conteúdo de proteína da j-ésima genitora no i-ésimo ambiente;

μ = média geral;

a_i = efeito fixo do i-ésimo ambiente;

t_{ij} = efeito aleatório da j-ésima genitora, no ambiente i;

e_{ij} = erro experimental.

Em termos matriciais HENDERSON (1963) desenvolveu um conjunto de equações conhecidas como Equações do Modelo Misto de Henderson, apresentado a seguir:

$$\tilde{y} = X\tilde{\beta} + Z\tilde{u} + \tilde{\varepsilon},$$

em que:

\tilde{y} = vetor de observações;

X e Z = matrizes de incidência de efeitos fixos e aleatórios, respectivamente;

$\tilde{\beta}$ = vetor dos efeitos fixos desconhecidos;

\tilde{u} = vetor dos efeitos aleatórios do indivíduo; e

$\tilde{\varepsilon}$ = vetor de erros aleatórios.

Assumindo-se que \tilde{u} e $\tilde{\varepsilon}$ apresentam valor esperado nulo e que a variância de $\tilde{u} = G = A \otimes G_0$ (A = matriz do numerador dos coeficientes de parentesco de Wright, \otimes denota produto direto e G_0 = matriz de variâncias e

covariâncias genéticas aditivas), a variância de $\tilde{\varepsilon} = R = I \otimes R_0$ ($I =$ matriz identidade e $R_0 =$ matriz de variância e covariâncias residuais) e a covariância de \tilde{u} e $\tilde{\varepsilon}$ é nula, então

$$E(\tilde{u}) = E(\tilde{\varepsilon}) = 0 \Rightarrow E(\tilde{y}) = X\tilde{\beta}$$

$$V(\tilde{u}) = G, V(\tilde{\varepsilon}) = R \Rightarrow V(\tilde{y}) = ZGZ' + R = V$$

Admitindo-se que y segue distribuição normal, tem-se

$$\tilde{y} = X\tilde{\beta} + Z\tilde{u} + \tilde{\varepsilon} \stackrel{\text{iid}}{\sim} N(X\tilde{\beta}, V)$$

Desta forma, o sistema de equações de modelos mistos a ser solucionado para obtenção das estimativas dos efeitos fixos e das predições dos valores genéticos é:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\tilde{\beta}} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}\tilde{y} \\ Z'R^{-1}\tilde{y} \end{bmatrix}$$

3.3.1.2. Cálculo da matriz de parentesco e os componentes de variância

A matriz A é chamada de matriz do numerador dos coeficientes de parentesco (numerator relationship matrix). WRIGHT (1922) definiu a relação de parentesco entre indivíduos como correlação entre os valores gênicos, mas A é essencialmente definida como a covariância (the numerator of the correlation

coefficients). Assim ela somente representa a relação genética aditiva entre indivíduos. Assumindo apenas os efeitos aditivos, a variância e covariância podem ser definidos como:

$$\sigma_{G_x}^2 = (1 + F_x) \sigma_A^2;$$

$$C_{G_{xy}} = 2 \theta_{xy} \sigma_A^2.$$

em que:

$\sigma_{G_x}^2$ = variância genética do indivíduo X;

$C_{G_{xy}}$ = covariância genética entre o indivíduo X e Y;

F_x = probabilidade de um indivíduo X ter dois alelos idênticos por descendência em um determinado loco, coeficiente de endogamia de WRIGHT (1922);

θ_{xy} = é a probabilidade em média de X e Y terem alelos idênticos por descendência, coeficiente de parentesco de MALÉCOT (1948);

Portanto, a forma geral da matriz A tem a seguinte estrutura:

$$A = \begin{bmatrix} (1 + F_1) & 2\theta_{12} & \dots & 2\theta_{1n} \\ 2\theta_{21} & (1 + F_2) & \dots & 2\theta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2\theta_{n1} & 2\theta_{n2} & \dots & (1 + F_n) \end{bmatrix}$$

Para predição dos valores genéticos, foi utilizado o procedimento IML do SAS[®] System (SAS, 2003). Na estimação dos componentes de variância foi utilizado o programa MTDFREML (“Multiple trait derivative-free restricted maximum likelihood” – BOLDMAN et al. (1995)) que utiliza o método da máxima verossimilhança restrita. Os componentes de variância iniciais foram calculados com base na população F2 cultivada juntamente com o progenitor

recorrente, em que a variância das plantas foi uma estimativa da variância genética e a variância do progenitor recorrente foi uma estimativa da variância ambiental. O critério de convergência foi a variância dos valores do simplex ($-2\log$) inferior a 10^{-8} .

3.3.2. Herdabilidade no sentido restrito

As herdabilidades, no sentido restrito, foram estimadas segundo a expressão abaixo:

$$h^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_p^2}$$

em que:

h^2 = herdabilidade no sentido restrito;

$\hat{\sigma}_a^2$ = estimador da variância genética aditiva; e

$\hat{\sigma}_p^2$ = estimador da variância genética fenotípica.

Foram estabelecidas quatro situações para obtenção das estimativas de herdabilidade, descritas a seguir.

Situação A As estimativas são baseadas na máxima verossimilhança restrita, utilizando todas as informações disponíveis (progenitores + F2 + F3 + RC1F2 + RC1F3 + RC2F2 + RC2F3).

Situação B As estimativas são baseada na máxima verossimilhança restrita, utilizando todas as informações disponíveis exceto a geração RC2F3 (progenitores + F2 + F3 + RC1F2 + RC1F3 + RC2F2).

Situação C As estimativas baseada na máxima verossimilhança restrita, utilizando das gerações RC2F2 e RC2F3.

Situação D As herdabilidades no sentido restrito, para a população RC2F2, foram estimadas com o avanço desta para a geração RC2F3, segundo SMITH e KINMAN (1965):

$$h^2 = \frac{\hat{b}}{2\theta_{XY}},$$

em que:

h^2 = estimador da herdabilidade no sentido restrito;

\hat{b} = estimativa do coeficiente de regressão da média das linhas da geração RC2F3 sobre plantas da geração RC2F2; e

θ_{XY} = coeficiente de parentesco de MALÉCOT (1948), entre as geração RC2F2 e RC2F3, em que $\theta_{XY} = \frac{1}{2}(1 + F_X)$, sendo F_X o coeficiente de endogamia da geração RC2F2.

Assim, a herdabilidade no sentido restrito foi estimada pela expressão:

$$h^2 = \frac{8}{15}\hat{b}, \text{ sendo } F_X = \frac{7}{8}.$$

As estimativas de herdabilidade no sentido restrito foram também obtidas com base nos coeficientes de regressão padronizados (b'), para minimizar o efeito das diferenças ambientais decorrentes do fato de cada geração ter sido conduzida em ano diferente, conforme o método proposto por FREY e HORNER (1957), pela seguinte expressão:

$$h^2 = \frac{8}{15} \hat{b}' ,$$

em que:

$$\hat{b}' = \text{estimador do coeficiente de regressão padronizado} = \frac{\hat{\sigma}_{\text{PRC}_2\text{F}_2}}{\hat{\sigma}_{\text{PRC}_2\text{F}_3}} \hat{b} ;$$

$\hat{\sigma}_{\text{PRC}_2\text{F}_2}$ = estimador do desvio-padrão fenotípico das plantas RC2F2;

$\hat{\sigma}_{\text{PRC}_2\text{F}_3}$ = estimador do desvio-padrão fenotípico das médias RC2F3; e

\hat{b} = estimador do coeficiente de regressão das médias das famílias RC2F3 em função dos dados das plantas da geração RC2F2.

3.3.3. Respostas esperadas

Para se comparar as diferentes metodologias de seleção, foram estimados os ganhos esperados e realizados, utilizando-se das seguintes expressões:

$$\text{GP} (\%) = h^2 (\bar{X}_{\text{RC2F2}}^{\text{S}} - \bar{X}_{\text{RC2F2}}) 100 ; \text{ e}$$

$$\text{GR} (\%) = \frac{\bar{X}_{\text{RC2F3}}^{\text{S}} - \bar{X}_{\text{RC2F2}}}{\bar{X}_{\text{RC2F2}}} 100$$

em que:

GP = ganho predito;

GR = ganho realizado;

h^2 = herdabilidade estimada;

\bar{X}_{RC2F2} = média da população RC2F2;

$\bar{X}_{\text{RC2F2}}^{\text{S}}$ = média das linhagens selecionadas na população RC2F2; e

\bar{X}_{RC2F3}^S = média das mesmas linhagens selecionadas na população RC2F3.

4. RESULTADOS E DISCUSSÕES

As estimativas de médias e os desvios-padrão do teor protéico da semente de cada população segregante nas diferentes gerações são apresentadas no Quadro 1. De modo geral, as médias estimadas das populações segregantes foram sempre superiores à do progenitor recorrente correspondentes. Isso demonstra que sempre houve linhagens com maior teor de proteína que o progenitor recorrente, nas diferentes fases de seleção do processo de retrocruzamentos.

De modo geral, as plantas F2 apresentaram maiores valores de desvios-padrão indicando maior variabilidade e as plantas RC2F2, os menores valores, demonstrando uma maior uniformidade para teor de proteína nas linhagens. Era esperado que as plantas RC2F3 apresentassem os menores valores, isto não foi observado, provavelmente devido ao efeito de ambiente, pois foram conduzidas em localidades diferentes daquelas onde se praticou seleção.

O maior e o menor valores da média para teor protéico nas sementes de soja foram observados na população oriunda da variedade OCEPAR 13. Entretanto, a menor média foi observada em F2 e a maior em F3. Essa diferença pode ser devido ao efeito de dominância negativa na geração F2, e na geração F3 efeito ambiental positivo, provavelmente, devido ao plantio fora da época, janeiro, normalmente recomendada.

O resumo das análises de variância da geração RC2F3 de todas as populações são apresentadas no Quadro 2. Observa-se que em todas as populações há interação significativa das famílias RC2F3 com o ambiente,

QUADRO 1 – Médias e desvios-padrão (s) das populações segregantes para teor protéico nas sementes da soja

Progenitor/geração	Época de plantio	Localidade	% proteína	
			Média	s
COODETEC 201	abril e outubro 2000	Viçosa	34,89	-
F2	setembro 1999	Viçosa	36,65	6,45
F3	janeiro 2000	Viçosa	44,95	3,45
RC1F2	abril 2000	Viçosa	43,61	5,26
RC1F3	outubro 2000	Viçosa	43,22	3,87
RC2F2	janeiro 2001	Viçosa	40,51	1,78
RC2F3	maio 2001	Capinópolis e Oratórios	39,11	2,62
COODETEC 202	abril e outubro 2000	Viçosa	36,10	-
F2	setembro 1999	Viçosa	36,50	4,66
F3	janeiro 2000	Viçosa	43,77	3,60
RC1F2	abril 2000	Viçosa	34,26	5,05
RC1F3	outubro 2000	Viçosa	45,23	4,50
RC2F2	janeiro 2001	Viçosa	39,04	1,84
RC2F3	maio 2001	Capinópolis e Oratórios	39,07	2,64
COODETEC 205	abril e outubro 2000	Viçosa	36,01	-
F2	setembro 1999	Viçosa	35,41	6,92
F3	janeiro 2000	Viçosa	44,99	2,54
RC1F2	abril 2000	Viçosa	38,36	5,79
RC1F3	outubro 2000	Viçosa	37,82	2,55
RC2F2	janeiro 2001	Viçosa	44,37	2,72
RC2F3	maio 2001	Capinópolis e Oratórios	39,72	2,44
OCEPAR 13	abril e outubro 2000	Viçosa	34,79	-
F2	setembro 1999	Viçosa	35,05	4,44
F3	janeiro 2000	Viçosa	46,10	1,84
RC1F2	abril 2000	Viçosa	44,01	4,34
RC1F3	outubro 2000	Viçosa	40,90	4,13
RC2F2	janeiro 2001	Viçosa	42,62	1,66
RC2F3	maio 2001	Capinópolis e Oratórios	45,93	2,80

indicando que os efeitos dos fatores famílias RC2F3 e ambiente, isoladamente, não explicam toda a variação observada na característica avaliada. Segundo FALCONER (1987), a interação significa que a melhor linhagem em um ambiente poderia não ser em outro. Assim, uma estratégia seria selecionar linhagens específicas para cada ambiente ou identificar linhagens sub-ótimas para os municípios de Capinópolis e Oratórios.

QUADRO 2 – Resumo da análise de variância da característica teor protéico avaliada na geração RC2F3 e as respectivas estimativas dos coeficientes de variação. Capinópolis e Oratórios, ano 2001

F.V	Progenitor recorrente							
	COODETEC 201		COODETEC 202		COODETEC 205		OCEPAR 13	
	Gl	QM	gl	QM	gl	QM	Gl	QM
Bloco	2	9,46*	2	0,89 ^{ns}	2	2,45 ^{ns}	1	11,50**
Ambiente	1	773,91**	1	683,26**	1	274,55**	1	41,05**
Famílias RC2F3	22	7,84**	23	8,16**	29	7,55**	20	11,50**
Fam x amb	18	9,96**	17	9,47**	15	6,70**	5	5,41**
Resíduo	29	2,81	28	2,79	20	1,18	12	0,89
CV (%)		4,29		2,05		2,73		2,05

(**),(*) Significativos a 1% e 5 % de probabilidade, pelo teste F, respectivamente.
(^{ns}) Não-significativo, a 5% de probabilidade, pelo teste F.

Neste experimento, o fator ambiente foi significativo, isto indica que nos ambientes testados, Capinópolis e Oratórios, exerceram efeitos diferenciados sob as populações testadas. Isto era esperado, pois há, entre os dois ambientes diferenças edafo-climáticas.

A observação da significância de todas as famílias RC2F3 indica a existência de variabilidade para teor protéico nas sementes nas populações estudadas, evidenciando-se a necessidade da continuação do programa de retrocruzamento e possibilidade de progresso genético.

Os coeficientes de variação (C.V.) obtidos entre 2,05 a 4,29 % foram considerados baixos na experimentação agrônômica e demonstra a boa precisão experimental e estão próximos dos valores observados por PIOVESAN (2000), MELO FILHO (2002).

Os resultados apresentados a seguir não consideraram o delineamento experimental nas análises. Ainda violam algumas pressuposições do modelo animal, por se tratar de uma espécie autógama, no entanto, várias considerações foram feitas no material e métodos para superar as dificuldades advindas destas pressuposições. É importante ressaltar que os parâmetros genéticos estimados são válidos especificamente para a população estudada; e se admite que o modelo utilizado representa corretamente a natureza da amostragem dos dados e reflete a biologia do problema.

Diferentes plantas de um progenitor foram consideradas idênticas geneticamente. As sementes F1, RC1F1 e RC2F1 foram consideradas gerações de passagem, onde não ocorre oportunidade de avaliar a variabilidade dentro das famílias, portanto não foram analisadas para teor de proteína.

No Quadro 3, são apresentadas as estimativas das herdabilidades no sentido restrito e as variâncias fenotípica e aditiva para teor de proteína avaliadas nas quatro populações.

Observa-se que variância das linhagens que utilizaram como progenitor recorrente a variedade OCEPAR 13, apresentaram menor valor de variância fenotípica, sendo indicativo que o progenitor doador PTN I12 seria de menor divergência genética, expressando seu potencial genético de maneira mais uniforme, quando comparada com as outras populações. MIRANDA (1997) e NAOE (2001), estudando variabilidade e divergência reportaram que cruzamentos entre progenitores pouco divergentes tendem a apresentar menor variação fenotípica em suas progênes.

A população oriunda de COODETEC 205 apresentaram sempre com os maiores valores de variância fenotípica entre as populações estudadas. Isto evidencia maior variabilidade entre plantas para teor de proteína, possibilitando uma seleção fenotípica com maior rigor.

A variância genética aditiva foi reduzida quando foram consideradas apenas as gerações RC2F2 e RC2F3, situação C. Isto era esperado pois o processo de seleção e do retrocruzamento nas gerações anteriores diminuiu a variância genética. No método de melhoramento por retrocruzamento poucos

QUADRO 3 – Estimativas das herdabilidades no sentido restrito (h^2) e das variâncias fenotípica (σ_f^2) e genética aditiva (σ_a^2) para teor de proteína avaliadas nas quatro populações de soja

População	Situação	σ_f^2	σ_a^2	h^2
COODETEC 201				
	A	30,24	7,79	0,26
	B	33,17	8,96	0,27
	C	4,87	0,44	0,09
	D	-	-	0,22
COODETEC 202				
	A	20,71	7,83	0,38
	B	22,99	9,12	0,40
	C	5,00	0,25	0,05
	D	-	-	0,02
COODETEC 205				
	A	34,34	5,55	0,16
	B	39,00	8,38	0,21
	C	6,33	1,02	0,16
	D	-	-	0,19
OCEPAR 13				
	A	18,21	4,97	0,27
	B	18,89	4,85	0,26
	C	3,90	1,55	0,40
	D	-	-	0,18

^{1/} A - baseada na REML com matriz de parentesco completa.

B - baseada na REML com matriz de parentesco completa, excetuando a geração RC2F3.

C - baseada na REML considerando somente as gerações RC2F2 e RC2F3.

D - baseada na metodologia de SMITH e KINMAN (1965) e FREY e HORNER (1957).

genitores são selecionados em cada ciclo (ALLARD, 1960).

Observou-se variação de na estimativas de herdabilidade nas diferentes populações de retrocruzamentos nas diferentes situações, isto era esperado, pois cada situação considera uma determinado número de indivíduos e população. O valor da herdabilidade depende da magnitude de todos os componentes de variância, uma alteração em qualquer um deles afeta o seu valor (FALCONER, 1987).

As estimativas de herdabilidade, no sentido restrito, cujos os componentes de variância foram estimadas pelo método da Máxima Verossimilhança Restrita, apresentaram-se baixas para teor de proteína nas sementes, nas quatro populações de retrocruzamentos estudadas. Quando a característica avaliada apresenta baixa herdabilidade, uma das vantagens do procedimento REML para estimação de componentes de variância é a de não se obterem componentes negativos, uma vez que existe restrição no espaço paramétrico das soluções (LOPES, et al., 1998).

Valores estimados baixos ocorrem quando há baixa variabilidade genética, interação genótipo x ambiente ou elevados erros experimentais (FALCONER, 1987). JOHNSON e BERNARD (1963), ao reportarem baixa herdabilidade para teor protéico, salientaram a dificuldade de se identificar e selecionar linhagens superiores. Estimativas de herdabilidade no sentido restrito na geração de retrocruzamentos por REML não foram encontradas na literatura, entretanto estão de acordo às observadas em gerações F2, F3 e F4 por SHORTER et al. (1976) e BRIM e BURTON (1979) que estimaram a herdabilidade com base na regressão. Os baixos valores de herdabilidade obtidos sugerem alta influência de fatores ambientais sobre a característica teor de proteína e base genética estreita da população base.

Na estimativa de herdabilidade para a população oriunda de OCEPAR 13, na situação C (0,40), aparentemente o valor está superestimado, quando comparados as demais estimativas. Isto foi devido ao menor valor de estimação da variância fenotípica e do maior valor da variância genética aditiva para a situação C. Erros de estimação componentes de variância estão associados ao

pequeno número de repetições, manejo inadequado da parcela no ensaio e erros de mensuração, no caso provavelmente nas linhagens RC2F2, as famílias seriam mais semelhantes entre si em relação as outras, como foi discutido, e por conseguinte a população estaria com maior homozigose. A variância genética aditiva é a variância dos valores genéticos aditivos que é a principal causa de semelhança entre as gerações, se uma população de uma planta autógama possui maior variância genética aditiva significa que maior número de genes estão em homozigose em relação a outra população da mesma espécie.

Os resultados de herdabilidade indicam que a seleção para teor de proteína, nessa condições ambientais obtidas, pode ser ineficiente baseando-se apenas nos valores fenotípicos do indivíduo. Todavia, o conhecimento de associações entre caracteres pode auxiliar na seleção de características de baixa herdabilidade (FALCONER, 1987; CRUZ e REGAZZI, 1997; VENCOVSKY e BARRIGA, 1992).

Para propiciar ganhos genéticos contínuos com a seleção ao longo de várias gerações devem ser selecionadas linhagens com média elevada e de ampla variabilidade genética.(ALLARD, 1960; JENSEN, 1988). Os futuros genitores são selecionados segundo suas informações individuais, e, também, de acordo com as informações de seus parentes.

Na seleção direta das seis melhores linhagens avaliando-se diretamente os valores observados na geração RC2F2 ou nos valores preditos pelo BLUP, nota-se que eleva o grau de parentesco em todas as populações de retrocruzamentos. Entretanto, no método de melhoramento utilizado, prevê-se que ao final dos recruzamentos haverá o cruzamento entre as linhagens recuperadas para que alelos desejáveis diferentes em cada linhagem recuperada se unam, assim devemos evitar selecionar plantas oriundas da mesma planta F2, para que genes que estejam em blocos gênicos diferentes tenham possibilidade de se unir. Entretanto, vale ressaltar que HELMS e ORF (1998), relataram que a seleção para incremento de proteína nas sementes podem resultar em decréscimo de produtividade.

Nos Quadros 4, 5, 6 e 7 são apresentadas as linhagens selecionadas de todas as populações para a continuidade do programa de retrocruzamentos nas diferentes situações. No Quadro 8 são apresentadas as correlações de Spearman, considerando todos os indivíduos avaliados, em diferentes situações. Observou-se que as linhagens mais promissoras foram oriundas de apenas três plantas F2 para todas as populações estudadas. Isto implicou na decisão de esquematizar os quadros com seis linhagens por população, duas plantas promissoras por planta F2, considerando os méritos observados ou preditos e sua ancestralidade.

As populações oriundas da variedade COODETEC 201 e OCEPAR 13 apresentaram o menor número de linhagens selecionadas, com 10 linhagens promissoras, e com menores números de classificação, isto pode ser indicativo de maior concordância entre os métodos. Isto indica que qualquer um das metodologias poderia ser utilizada para descartar algumas poucas linhagens menos interessantes sem perder as linhagens promissoras. Por exemplo, algumas linhagens RC2F2 poderiam ser descartadas antes de cultivar as sementes RC2F3, sem perda de linhagens promissoras. Corroborando com essa observação, verifica-se no Quadro 8 que apenas as linhagens mais promissoras se assemelham, pois não houve alta correlação de Spearman entre as situações propostas quando se utilizam todos os indivíduos.

Para avaliar as diferentes seleções propostas sob a classificação das linhagens foi utilizado o coeficiente de correlação de Spearman entre as diferentes situações nas quatro populações. A correlação de Spearman estima a relação de ordem, valores entre -0,50 a 0,50 foram considerados baixos e indicando que a ordem de classificação dos indivíduos são diferentes.

As correlações entre as situações tiveram ampla variação nas quatro populações estudadas, apresentando valores que variaram de 0,07 a 0,90, o que já seria suficiente para prever que as classificações nas diferentes situações não se contradiziam pois não houve correlação negativa significativa. Vale ressaltar que a correlação de valor 0,07 demonstra que na população oriunda da variedade OCEPAR 13, a seleção baseada diretamente na geração RC2F2 e utilizando a metodologia de modelos mistos com toda a genealogia tem-se classificação

QUADRO 4 – Resumo da classificação das linhagens selecionadas da população RC2F2 em cada metodologia na população oriunda da variedade de soja COODETEC 201

Linhagem	Planta F2 ^{1/}	Seleção I ^{2/}	Seleção II	Seleção III	Seleção IV
PTC 10566	PTC 323	41,325	0,269	0,305*	0,191
PTC 11568	PTC 323	41,744	0,369*	0,353*	0,302*
PTC 21574	PTC 323	42,208*	0,565*	0,205	0,533*
PTC 10585	PTC 323	42,934*	-0,002	0,295	0,109
PTC 22542	PTC 434	44,259*	-0,062*	-0,101	0,400*
PTC 33544	PTC 434	44,793*	0,981*	0,288*	1,071*
PTC 14546	PTC 434	42,840	-0,408	0,017*	-0,024
PTC 35538	PTC 675	40,932	-0,074*	-0,564*	0,407*
PTC 43577	PTC 675	42,824*	-0,812	-1,116	0,157
PTC 44580	PTC 675	43,231*	-0,346*	-1,059*	0,522*

^{1/} Código numérico da planta F2, ancestral da linhagem.

^{2/} Seleção I – considera apenas os valores fenotípicos (teor de proteína) da geração RC2F2.

Seleção II – baseados nos valores preditos (BLUP) com matriz de parentesco completa.

Seleção III - baseados nos valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

Seleção IV – baseado nos valores preditos (BLUP) de RC2F2 e RC2F3.

* linhagem selecionada.

QUADRO 5 – Resumo da classificação das seis linhagens selecionadas da população RC2F2 em cada metodologia na população oriunda da variedade de soja COODETEC 202

Linhagem	Planta F2 ^{1/}	Seleção I ^{1/}	Seleção II	Seleção III	Seleção IV
PTC2 98370	PTC2 234	39,929*	-0,012*	-1,161*	-0,079*
PTC2 73374	PTC2 234	38,580*	-0,010*	-1,463*	-0,060*
PTC2 45376	PTC2 334	41,341	0,422*	-0,468	0,100
PTC2 46381	PTC2 334	44,259*	0,015	0,184	0,085
PTC 23382	PTC2 334	38,715	0,321*	-1,056	-0,032
PTC2 87384	PTC2 334	42,186	0,142	-0,141	0,539*
PTC2 13385	PTC2 334	42,840*	-0,149	0,005	-0,212
PTC 45395	PTC2 334	42,132	-0,111	0,417*	-0,109
PTC2 78399	PTC2 334	41,144	-0,122	0,196*	0,096*
PTC2 37414	PTC2 678	39,875	0,215*	-0,058	0,206*
PTC2 65417	PTC2 678	40,994*	-0,004	0,190*	0,022
PTC2 18428	PTC2 678	42,248*	0,111*	0,388*	0,140*

^{1/} Código numérico da planta F2, ancestral da linhagem.

^{2/} Seleção I – considera apenas os valores fenotípicos (teor de proteína) da geração RC2F2.

Seleção II – baseados nos valores preditos (BLUP) com matriz de parentesco completa.

Seleção III - baseados nos valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

Seleção IV – baseado nos valores preditos (BLUP) de RC2F2 e RC2F3.

* Linhagem selecionada.

QUADRO 6 – Resumo da classificação das seis linhagens selecionadas da população RC2F2 em cada metodologia na população oriunda da variedade de soja COODETEC 205

Linhagem	Planta F2 ^{1/}	Seleção I ^{1/}	Seleção II	Seleção III	Seleção IV
PTC5 23457	PTC5 4	46,072	1,043*	0,342	1,491*
PTC5 86458	PTC5 4	49,346*	0,953	0,692*	0,255
PTC5 14462	PTC5 4	49,003*	0,104	0,656	0,620
PTC5 82465	PTC5 4	47,895	-0,238	0,147	0,236
PTC5 24466	PTC5 4	46,949	-0,244	0,046	0,387
PTC5 25478	PTC5 4	48,266	0,961*	0,975*	0,680*
PTC5 23482	PTC5 23	48,931*	0,997*	1,046*	1,245*
PTC5 56483	PTC5 23	44,363	-0,322	-0,565	0,044
PTC5 89485	PTC5 23	45,596	-0,457	-0,433	-0,396
PTC5 13490	PTC5 23	46,100*	0,877*	-0,379*	0,651*
PTC5 23500	PTC5 43	49,117*	0,019	-0,014*	0,411
PTC5 14503	PTC5 43	47,764*	1,064*	-0,159*	1,104*
PTC5 46504	PTC5 43	45,881	1,009*	-0,360	1,029*

^{1/} Código numérico da planta F2, ancestral da linhagem.

^{2/} Seleção I - considera apenas os valores fenotípicos (teor de proteína) da geração RC2F2.

Seleção II - baseados nos valores preditos (BLUP) com matriz de parentesco completa.

Seleção III - baseados nos valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

Seleção IV – baseado nos valores preditos (BLUP) de RC2F2 e RC2F3.

* Linhagem selecionada.

QUADRO 7 – Resumo da classificação das seis linhagens selecionadas da população RC2F2 em cada metodologia na população oriunda da variedade de soja OCEPAR 13

Linhagem	Planta F2 ^{1/}	Seleção I ^{1/}	Seleção II	Seleção III	Seleção IV
PTO3 36441	PTO3 42	44,821	1,887*	-0,064	2,935*
PTO3 64442	PTO3 42	46,278*	1,521*	0,128	2,818*
PTO3 25448	PTO3 42	43,697	-0,399	0,280*	0,335
PTO3 45483	PTO3 42	45,186*	-0,388	0,058	0,551
PTO3 12487	PTO3 42	45,137	-0,422	0,303*	0,090
PTO3 65493	PTO3 64	45,784*	0,953*	-0,312*	1,552*
PTO3 60495	PTO3 64	44,710*	-0,678*	-0,452*	0,261*
PTO3 65460	PTO3 80	43,333*	0,932*	-0,038	0,457*
PTO3 25464	PTO3 80	43,518*	0,941*	-0,013*	1,202*
PTO3 97475	PTO3 80	42,726	0,926	-0,001*	0,306

^{1/} Código numérico da planta F2, ancestral da linhagem.

^{2/} Seleção I - considera apenas os valores fenotípicos (teor de proteína) da geração RC2F2.

Seleção II - baseados nos valores preditos (BLUP) com matriz de parentesco completa.

Seleção III - baseados nos valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

Seleção IV – baseado nos valores preditos (BLUP) de RC2F2 e RC2F3.

* Linhagem selecionada.

QUADRO 8 – Estimativas dos coeficientes de correlação de Spearman nas quatro populações de retrocruzamento de soja entre as diferentes classificações

	Populações			
	CD 201 ^{1/}	CD 202	CD 205	OCEPAR 13
Classificação I ^{2/}				
Classificação II	0,09 ^{ns}	0,26	0,46	0,07 ^{ns}
Classificação III	0,29	0,71	0,66	0,16
Classificação IV	0,17	0,19	0,50	0,46
Classificação II				
Classificação I	0,09 ^{ns}	0,26	0,46	0,07 ^{ns}
Classificação III	0,69	0,48	0,83	0,38
Classificação IV	0,75	0,90	0,79	0,82
Classificação III				
Classificação I	0,29	0,71	0,66	0,16
Classificação II	0,69	0,48	0,83	0,38
Classificação IV	0,17	0,21	0,46	0,22

^{1/} CD = COODETEC.

^{2/} Classificação I – considera apenas os valores fenotípicos da geração RC2F2.

Classificação II – baseados nos valores preditos (BLUP) com matriz de parentesco completa.

Classificação III - baseados nos valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

Classificação IV – baseado nos valores preditos (BLUP) de RC2F2 e RC2F3.

(ns) Não significativo a 5% de probabilidade pelo teste t.

diferente, mas não contraditória.

No Quadro 9 observa-se que a classificação II tem correlação alta com a classificação IV, isto é interessante, pois com a perda de genealogia anteriores a RC2F2 poderemos obter uma classificação semelhante, entretanto a seleção entre famílias ficaria prejudicada. Quando ocorre perda de genealogia pode-se prever seus parentescos utilizando-se marcadores moléculas ou técnicas de análise multivariada.

Um dos indicativos da viabilidade de uma população para fins de melhoramento foi a possibilidade de prever ganhos em virtude da seleção (SILVA, 1982; FALCONER, 1987). No Quadro 9 são apresentados as médias gerais, as médias das linhagens selecionadas e os ganhos preditos e realizado nas diferentes metodologias de seleção. Esta análise tem por finalidade verificar se poderia prever o comportamento da geração RC2F3.

Na população oriunda da COODETEC 201 observou-se que nas situações que utilizaram o BLUP foram observadas maiores ganhos realizados, entretanto houve ganho realizado negativo no teor de proteína quando se selecionou com base nos valores fenotípicos de RC2F2. Estes resultados contradizem o esperado (ganho predito), mas demonstra a utilização promissora de BLUP. As populações oriundas de COODETEC 202 e OCEPAR 13 também apresentaram maiores ganhos realizados quando utilizado o BLUP.

Observou-se que na população oriunda de COODETEC 205, apesar da seleção positiva, não houve ganho positivo, obtendo-se valores fenotípicos, em RC2F3, abaixo do esperado. Isto se deve, provavelmente, ao fato de as linhagens terem sido selecionadas anteriormente para o microambiente de Viçosa - MG e as linhagens RC2F3 terem sido cultivadas em outras localidades. VENCOVSKY e BARRIGA (1992), já havia reportado que se a comparação entre ganho predito e ganho realizado quando ocorre em diferentes condições ambientais, além da diferença de magnitude, o sentido do ganho poderia mudar. Apesar de apresentar ganho negativo, quando utilizando BLUP houve menor perda de ganho.

Como foi observado, apesar do ganhos diferenciados entre as diferentes metodologias, diferentes alternativas podem ser adotadas, sendo o que vai

QUADRO 9 – Estimativas das médias e ganhos das populações de retrocruzamento de soja nas diferentes situações

População	II	II	III	IV
COODETEC 201				
\bar{X}_{RC2F2}	40,51	40,51	40,51	40,51
\bar{X}_{RC2F2} selecionadas	43,37	43,73	42,86	42,88
\bar{X}_{RC2F3} selecionadas	40,71	40,20	41,11	40,92
Ganho predito	62,49	70,13	61,25	61,67
Ganho realizado	0,50	-0,76	1,49	1,03
COODETEC 202				
\bar{X}_{RC2F2}	39,04	39,04	39,04	39,04
\bar{X}_{RC2F2} selecionadas	41,48	42,50	40,66	40,69
\bar{X}_{RC2F3} selecionadas	40,17	40,15	40,86	40,24
Ganho predito	2,21	3,14	61,58	62,84
Ganho realizado	2,90	2,83	4,67	3,08
COODETEC 205				
\bar{X}_{RC2F2}	44,37	44,37	44,37	44,37
\bar{X}_{RC2F2} selecionadas	47,84	48,79	46,52	47,79
\bar{X}_{RC2F3} selecionadas	38,39	39,93	41,60	41,04
Ganho predito	67,60	85,96	34,36	54,75
Ganho realizado	-13,48	-10,02	-6,25	-7,52
OCEPAR 13				
\bar{X}_{RC2F2}	42,62	42,62	42,62	42,62
\bar{X}_{RC2F2} selecionadas	44,80	45,32	44,64	44,24
\bar{X}_{RC2F3} selecionadas	46,71	46,86	47,43	47,99
Ganho predito	40,01	49,52	54,44	43,68
Ganho realizado	9,58	9,93	11,27	12,59

^{I/} I - considera apenas os valores fenotípicos da geração RC2F2.

II - idem situação I, evitando-se selecionar linhagens com alta covariância genética.

III - baseados nos maiores valores preditos (BLUP) com matriz de parentesco completa, excetuando a geração RC2F3.

IV - idem situação III, evitando-se selecionar linhagens com alta covariância genética.

determinar a melhor estratégia será a quantidade de informação disponível, pois os resultados mostram que o BLUP tem boa precisão, pois os resultados foram os melhores nas diferentes populações de retrocruzamentos, e também apresenta boa acurácia pois absorveu bem os erros sistemáticos (seleção sem considerar a covariância).

Pode-se concluir, com base nos resultados do Quadro 9, que a seleção baseado no BLUP e evitando-se ou não o aumento de parentesco mostrou-se mais promissora. O cultivo da geração RC2F3 em ambiente diferente onde havia ocorrido seleção, provavelmente favoreceu a seleção que utiliza informação de parentesco. Para utilização do BLUP o conhecimento do parentesco é fundamental, existem outras formas de obtenção como por exemplo: o uso de marcadores moleculares, podendo ser utilizado em programas iniciais de melhoramento; quando se desconhece o parentesco. O BLUP pode auxiliar em metodologias de seleção que necessitam de dispor do conhecimento do parentesco como no “método genealógico”.

A utilização do BLUP para seleção de linhagens apresentou-se promissor no programa de retrocruzamento da soja para elevar o teor protéico na soja, lembrando que condução das populações deve ser planejada para tal procedimento, assim empregar o BLUP para auxiliar na tomada de decisão passa a ser uma alternativa viável e promissora.

5. CONCLUSÕES

- a) As estimativas de herdabilidade, no sentido restrito, cujos os componentes de variância foram estimadas pelo método da Máxima Verossimilhança Restrita, apresentaram-se baixas para teor de proteína nas sementes, nas quatro populações de retrocruzamentos de soja estudadas;
- b) Selecionando-se somente os maiores valores preditos pelo BLUP, haverá aumento de parentesco na população;
- c) Houve diferenças entre ganhos preditos e realizados, entretanto observou-se que com a utilização do BLUP houve melhores ganhos realizados;

CAPITULO 2

DISTRIBUIÇÃO DE FREQUÊNCIA DOS TEORES DE PROTEÍNA NAS SEMENTES DE POPULAÇÕES DE RETROCRUZAMENTO EM SOJA

1. RESUMO

No presente estudo, trabalhou-se com a característica quantitativa teor protéico nas sementes de soja. Foram estudadas sete populações oriundas de retrocruzamentos cujo objetivo foi de traspassar genes para alto teor de proteínas nas sementes, para variedades comerciais. As distribuições das observações e, em conseqüência, a natureza da variação genética, foi avaliada por meio da análise da curtose e da assimetria das distribuições. Observou-se que a média do teor protéico foi deslocada positivamente, indicando potencial para o desenvolvimento de linhagens promissoras, com maior teor de proteína nas sementes; a variabilidade do teor de proteína dentro das populações depende somente da presença de poucos alelos nas diferentes populações; no segundo retrocruzamento, as populações tendem a apresentar distribuição mesocúrtica ou leptocúrtica e, por meio da estimativa de assimetria, foi possível constatar efeito de dominância em algumas das populações avaliadas.

Termos para indexação: *Glycine max*, proteína, distribuição, frequência.

2. INTRODUÇÃO

Os programas de melhoramento de soja do Brasil, embora o país seja o segundo maior produtor mundial de leguminosa, apresentam grande dependência de materiais exóticos pelo fato da base genética do germoplasma brasileiro ser estreita. Programas de melhoramento visam o desenvolvimento de variedades que sejam produtivas e de melhor qualidade. Um aspecto referente à qualidade do grão de soja é o teor e a qualidade da proteína. A qualidade da proteína refere-se ao teor de aminoácidos sulfurados e também às propriedades funcionais da proteína (MOREIRA et al. 1979, 1990).

Quando processados, 100 quilos de soja produzem, em média, 79 quilos de farelo (que possui cerca de 50% de proteína) e 18,4 quilos de óleo. Por meio do melhoramento genético é possível o desenvolvimento de variedades com características especiais de qualidade, que podem permitir à indústria de alimentos o desenvolvimento de produtos ou de processos tecnológicos mais econômicos e eficientes, viabilizando, assim, aceitabilidade mais efetiva, em razão da oferta de produtos de soja de melhor qualidade.

O preço do farelo para exportação é dependente da classificação quanto ao padrão de qualidade, em que “Normal” deve conter no mínimo 46 % de proteína, servindo como base de preço, “Low Pro” deve conter um mínimo de 43,5 % de proteína, sofrendo um deságio no preço e “Hy Pro” com um mínimo de 48%, ganhando um ágio em relação ao preço da classificação “Normal” (MOREIRA, 1999).

Os estudos de ação gênica relacionados ao teor protéico nas sementes de soja indicam que o efeitos gênicos aditivos são mais importantes que da dominância e da epistasia (BRIM e COCKERHAM, 1961, THORNE e FEHR, 1970, McKENDRY et al., 1985, PIOVESAN, 2000). Dessa forma, há indicação de que nas populações trabalhadas por esses autores, o teor de proteína das sementes pode ser eficientemente melhorado ainda em gerações precoces.

O conhecimento da natureza genética de uma característica possibilita tornar o processo seletivo mais eficiente. Entretanto, no caso de retrocruzamentos em que se trabalha com populações pequenas, as estimativas são viesadas, pois a mudança nas frequências gênicas em função da amostragem ocorrem em todas as gerações.

No presente estudo, trabalhou-se com a característica quantitativa teor protéico nas sementes. Foram estudadas sete populações oriundas de retrocruzamentos que visavam a introdução, em variedades comerciais, de genes para alto teor de proteína nas sementes, sendo avaliadas quanto à natureza da sua variação genética.

3. MATERIAL E MÉTODOS

3.1. Material genético e município

Foram utilizadas neste trabalho plantas de soja de ciclo precoce/semiprecoce oriundas de retrocruzamentos das variedades de soja OCEPAR 13, OCEPAR 14, COODETEC 201, COODETEC 202, COODETEC 203, COODETEC 205 e COODETEC 206. Essas variedades apresentam teores de proteína em níveis comumente observados nas variedades comerciais, sendo indicadas para plantio comercial no sul do Brasil e tendo sido utilizadas como progenitores recorrentes. Comercialmente são conhecidas pelas siglas OC 13, OC 14, CD 201, CD 202, CD 203, CD 205 e CD 206, respectivamente. Como progenitores doadores, foram utilizadas as linhagens de alto teor protéico BARC 8 (LEFFEL, 1992) e BR 80 14887, neste trabalho foram consideradas como variedades para facilitar a discussão. Os progenitores recorrentes são pertencentes à Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda (COODETEC) e os progenitores doadores foram desenvolvidos pelo USDA-ARS nos Estados Unidos e EMBRAPA/Soja, respectivamente.

Os cruzamentos iniciais foram realizados em casa de vegetação do BIOAGRO, no mês de fevereiro de 1999, localizado no Campus da Universidade Federal de Viçosa, em Viçosa, na Zona da Mata do Estado de Minas Gerais,

cujas coordenadas geográficas são 20°45'20''S e 42°52'40''W e, aproximadamente, 650 m de altitude.

As sementes F1's têm a seguinte genealogia:

- a) PTN I12 x OCEPAR 13;
- b) PTN I12 x OCEPAR 14;
- c) BR 80 14887 x CD 201;
- d) BR 80 14887 x CD 202;
- e) BR 80 14887 x CD 203;
- f) PTN I12 x CD 205;
- g) PTN C1 x CD 206.

O progenitor à esquerda do esquema de cruzamentos sempre foi o genitor materno. A linhagem PTN I12 foi oriunda do cruzamento de (BARC 8 x IAC-12) x IAC 12 e a linhagem PTN C1 de (BARC 8 x CAC 1) x CAC 1. As variedades IAC 12 e CAC 1 não apresentam alto teor protéico nas sementes e foram desenvolvidas pelo Instituto Agrônômico de Campinas e pela Cooperativa Agrícola de Cotia, respectivamente.

As sementes F1 foram cultivadas em maio de 1999, sendo o cultivo realizado em vaso, com uma planta. Os cruzamentos oriundos de OC 14, CD 201, CD 202 e CD 205 foram confirmados por meio de análise de marcadores RAPD, pois não se dispunha de marcador morfológico. A colheita e a debulha foram efetuadas, manualmente, para cada planta, quando 100% das vagens estavam secas para colheita.

As populações segregantes foram conduzidas como descrito no capítulo 1.

3.2. Análise de dados

Na análise dos dados, foram calculadas as variâncias, indicativas da variabilidade fenotípica. Foram, também, calculadas a assimetria e a curtose. A

assimetria é a medida do grau de desvio em relação à distribuição normal. Se a curva de frequência da distribuição tem uma cauda mais longa à direita da ordenada máxima que à esquerda, diz-se que a distribuição é desviada para a direita, ou que ela tem assimetria positiva. Se o inverso ocorre, diz-se que ela é desviada para a esquerda ou que tem assimetria negativa. Na avaliação do grau de assimetria, utiliza-se do terceiro momento centrado na média, expresso sob forma não-dimensional. O coeficiente do momento de assimetria de uma amostra populacional é definido por:

$$a = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x - \bar{x}}{s} \right)^3 ,$$

em que:

a = coeficiente do momento de assimetria;

s = desvio padrão da amostra;

n = número total de indivíduos da amostra;

x = dados coletados;

\bar{x} = média aritmética.

A curtose é o grau de achatamento de uma distribuição, considerado usualmente em relação à distribuição normal. A distribuição que tem um pico relativamente alto é denominada leptocúrtica, enquanto aquela que tem o topo achatado é denominada platicúrtica. A distribuição normal, que não é nem muito pontiaguda nem muito achatada, é denominada mesocúrtica. Baseada no quarto momento centrado na média, expressa sob forma não-dimensional, o coeficiente do momento de curtose de uma amostra é definido por:

$$k = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} ,$$

em que:

k = coeficiente do momento de curtose;

s = desvio padrão da amostra;

n = número total de indivíduos da amostra;

x = dados coletados;

\bar{x} = média aritmética.

Assumindo-se que os progenitores recorrentes e doadores utilizados no presente trabalho têm os alelos fixados (frequência igual a 1 ou zero) em todos os locos, ou seja, a distribuição das observações obtida nas variáveis quantitativa dentro desses progenitores tem distribuição próxima da normal. Portanto qualquer ocorrência do desvio da distribuição normal na população oriunda de cruzamentos será resultante da introgressão de novos alelos, os quais estariam segregando e seriam detectados pelos testes de curtose e, ou de assimetria.

Para a característica teor protéico, cujo interesse foi o de aumentar o seu nível nas sementes, a estimativa de assimetria nula ($-0,5 < a < 0,5$) ou positiva ($a \geq 0,5$) indica expectativa de seleção eficiente, enquanto a assimetria negativa ($a \leq -0,5$) indica seleção não eficiente, ou seja, há efeito da dominância para alto teor protéico. Quanto à curtose, se a distribuição for mesocúrtica ($-1 < k < 1$) ou leptocúrtica ($k \geq 1$), os genótipos possuem divergência baixa ou nula, visto que os alelos entre os genótipos devem ser iguais ou de efeitos semelhantes; se for platicúrtica ($k \leq -1$) são divergentes, pois a combinação de diferentes alelos produziriam indivíduos muito contrastantes. A regra para discriminar as classes dos coeficientes foi baseada em FOSTER (1986) e a interpretação genética em LYNCH e WALSH (1998).

4. RESULTADOS E DISCUSSÃO

A distribuição das frequências dos diferentes retrocruzamentos de soja, na geração RC2F2, as estimativas de assimetria, curtose e médias estão apresentadas no Quadro 1. Observa-se que todas as populações de retrocruzamentos apresentaram, em relação aos seus progenitores recorrentes, média do teor protéico deslocada positivamente, indicando potencial para o desenvolvimento de linhagens promissoras, com maior teor de proteína nas sementes.

A ordem de classificação decrescente, quanto ao teor protéico das sementes dos progenitores recorrentes, foi OC 13 > CD 206 > CD 205 > CD 203 > CD 201 > CD 202 > OC 14. Avaliando-se somente as melhores plantas oriundas dos retrocruzamentos, a classificação foi CD 205 > CD 203 > OC 13 > CD 201 = OC 14 > CD 202 > CD 206. Comparando-se as médias populacionais de cada retrocruzamento observa-se CD 205 > CD 203 > OC 13 > CD 201 > CD 202 > CD 206 > OC 14. Observando-se essas classificações nota-se que a variedade CD 205 foi a que apresentou a melhor classificação após os ciclos de retrocruzamento, provavelmente em razão da introgressão de alelos de efeito maior que conferem alto teor protéico, pois apresentou a assimetria nula, mostrando que nesta população a seleção é eficiente, enquanto que na CD 206, por exemplo, existe maiores dificuldades de seleção.

Quanto à estimativa da variância dentro da amostra, o cruzamento com o progenitor recorrente CD 205 apresentou maior valor entre os retrocruzamentos estudados, além do alto valor estimado da média. Esse resultado é interessante,

QUADRO 1 – Distribuição de frequências, média da população e do progenitor recorrente, assimetria e curtose do teor protéico nas sementes da soja, na geração RC2F2 de sete cruzamentos. Viçosa, MG

Classe (% proteína)	Progenitor recorrente						
	Ocepar 13	Ocepar 14	Coodetec 201	Coodetec 202	Coodetec 203	Coodetec 205	Coodetec 206
49						4	
48						5	
47						5	
46	2				5	8	
45	8	1	1		8	4	
44	10		2	4	12	11	
43	13		7	2	13	7	1
42	13	1	6	5	13	6	4
41	14	8	13	3	9	6	10
40	2	8	13	12	3	1	9
39	1	15	13	12	1		10
38		14	4	8		2	6
37	1	8	1	5			2
36		8	1	10			1
35		5		2			2
34		1					
33							1
32							1
31		1					2
Média da população	42,62	38,24	40,51	39,04	42,98	44,40	38,91
Média do progenitor recorrente	39,77	35,92	37,71	36,23	38,02	38,06	38,87
Variância da amostra	2,77	4,59	3,18	5,63	3,40	8,23	7,55
Assimetria	-0,25	0,24	0,21	0,36	0,03	-0,05	-1,48*
Curtose	0,48	0,60	-0,04	-0,50	-0,58	-0,57	2,08*

* a distribuição afasta-se da distribuição normal.

pois evidência maior variabilidade entre plantas, aumentando a possibilidade de se conseguir linhagens com a média de teor protéico superiores às dos progenitores. A menor variância foi observada em OC 13, entretanto a existência de pouca variabilidade neste caso não prejudica o processo seletivo das plantas, pois algumas plantas apresentam valores superiores a 15 % em relação à média do progenitor recorrente.

A assimetria está relacionada com o grau de dominância. Na análise do coeficiente de assimetria, populações oriundas de OC 13, OC 14, CD 201, CD 202, CD 203 e CD 205 apresentaram valores próximos de zero, indicando ausência de efeito de dominância. De acordo com THORNE e FEHR (1970), sabe-se que existe dominância para o caráter estudado, mas em pequena magnitude. Quanto à população oriunda de CD 206, verificou-se estimativa negativa indicando haver efeito da dominância para alto teor protéico. Assim, a segregação da progênie RC2F3 deve ser posteriormente avaliada para se assegurar maiores ganhos na seleção.

A curtose está relacionada com a presença de alelos diferentes na população e de maneira indireta, nos progenitores. Todas as populações, exceto a oriunda de CD 206, apresentaram estimativas nulas. Portanto, a curva de distribuição dos dados da amostra é mesocúrtica, indicando baixa divergência entre os indivíduos. Esse fato era esperado, pois, a princípio, a divergência foi reduzida, quando foram selecionadas plantas F2 e RC1F2 similares ao progenitor recorrente para continuar o programa de retrocruzamento; entretanto, existem populações com maior valor de variância do que outras, indicando que a variância fenotípica observada foi dependente somente da presença de poucos alelos dentro de cada população. A população oriunda de CD 206 apresentou-se leptocúrtica, isto provavelmente em razão da pré-existência de alelos de efeito maior que conferem alto teor protéico (O'DONALD, 1971), com acréscimo do efeito da dominância.

5. CONCLUSÕES

- a) A média do teor protéico foi deslocada positivamente, indicando potencial para o desenvolvimento de linhagens promissoras, com maior teor de proteína nas sementes;
- b) A variabilidade do teor de proteína observada dentro das populações depende somente da presença de poucos alelos nas diferentes populações;
- c) No segundo retrocruzamento, as populações tendem a apresentar a curva de distribuição mesocúrtica ou leptocúrtica;
- d) Por meio da estimativa de assimetria foi possível constatar ausência e presença de efeito de dominância nas populações avaliadas; sendo que a população oriunda de COODETEC 206 apresentou efeito de dominância para alto teor protéico.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELNOOR, R.V.; BARROS, E.G.; MOREIRA, M.A. Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Brazilian Journal of Genetics**, v. 18, n. 2, p. 265-273, 1995.
- ALLARD, R. W. **Principles of plant breeding**. New York: John Wiley, 1960, 485p.
- ANDERSON, R.D. **Variance components**. In: QUAAS, R.L.; ANDERSON, R.D.; GILMOUR, A.R. Use of mixed models for prediction and for estimation of (co)variance components. University of New England, AGBU, p. 77-145, 1984.
- ANDRÉ, C.M.G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica**. Lavras, UFLA, 101p, 1999. (tese de mestrado).
- BERNARDO R. Best linear unbiased prediction of the performance of crosses between untested maize inbreds. **Crop Science**, v. 36, n. 4, p. 872-876, 1996.
- BOLDMAN, K.G.; VAN VLECK, L.D. Derivative-free restricted maximum likelihood estimation in animal models with a sparse matrix solver. **Journal of Dairy Science**, v. 74, n. 12, p. 4337-4343, 1991.
- BOLDMAN, K.G.; KRIESE, L.A.; VAN VLECK, L.D.; VAN TASSELL, C.P. and KACHMAN, S.D. A manual for use of MTDFREML. A set of programs to obtain estimatives of variances and covariances [DRAFT]. University of Nebraska, Lincoln: U.S. Department of Agriculture, Agricultural Research Service, April 1995, 114p.

- BORÉM, A.; SANTOS, F.R. **Biotecnologia simplificada**. Viçosa: Folha da Mata, 2003, 250p.
- BORÉM, A. **Melhoramento de plantas**. Viçosa: UFV, 1997, 547p.
- BRIM, C.A.; COCKERHAM, C.C. Inheritance of quantitative characters in soybeans. **Crop Science**, v. 1, n. 3, p. 187-190, 1961.
- BRIM, C.A.; BURTON, J.W. Recurrent selection in soybeans II. Selection for increased percent protein in seeds. **Crop Science**, v. 19, n. 4, p. 494-498, 1979.
- BUENO, L.C.S.; MENDES, A.N.G.; CARVALHO, S.P. **Melhoramento genético de plantas princípios e procedimentos**. Lavras: UFLA, 2001, 282p.
- CÂMARA, G.M.S. **Soja: tecnologia da produção**. Piracicaba: ESALQ, 2000, 450p.
- CARNEIRO, P.L.S. **Oscilação genética e comparação de métodos de seleção tradicionais e associados a marcadores moleculares**. Viçosa: UFV, 94p, 2002. (Tese de doutorado).
- CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 2 edição. Viçosa: UFV, 1997, 390p.
- DESTRO, D.; MONTALVÁN, R. **Melhoramento genético de plantas**. Londrina: UEL, 1999, 818p.
- ERICKSON, L.R.; VOLDENG, H.D.; BEVERSDORF, W.D. Early generations selection for protein in *Glycine Max* x *Glycine soja* crosses. **Canadian Journal of Plant Science**, v. 6, n. 4, p. 901-908, 1981.
- FALCONER, D.S. **Introdução à genética quantitativa**. Viçosa: UFV, 1987, 279p.
- FEHR, W.R. **Principles of cultivar development**. Theory and technique. New York: Macmillan, 1987, 536p.
- FEHR, W.R.; CAVINESS C.E.; BURMOOD D.T.; PENNINGTON J.S. Stage of development descriptions for soybeans, *Glycine max*. (L) Merrill. **Crop Science**, v. 11, n. 6, p. 929-931, 1971.
- FISHER, R.A. The correlation between relatives on the supposition of mendelian inheritance. **Transactions of the Royal Society of Edinburgh**, v. 52, p. 399-433, 1918.

- FOSTER, G. **Financial Statement Analysis**. 2 ed. New Jersey: Prentice-Hall, 1986. 625p.
- FREY, K.J.; HORNER, T. Heritability in standard units. **Agronomy Journal**, v. 49, n. 2, p. 59-62, 1957.
- GUIMARÃES, C.T.; MOREIRA, M.A. Genética molecular aplicada ao melhoramento de plantas. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: Editora UFV, p. 715-740, 1999.
- HANSON, W.D. The theoretical distribution on lengths of parental gene blocks in the gametes of an F1 individual. **Genetics**, v. 44, n. 2, 197-209, 1959.
- HARTLEY, H.O.; RAO, J.N.K. Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, v. 54, n. 1, p. 93-108, 1967.
- HARTWIG, E.E. Varietal development. In: B.E. Caldwell (ed.). **Soybeans: Improvement, production and uses**. Madison: American Society of Agronomy, p. 187-210, 1973.
- HELMS, T.C.; ORF, J.H. Protein, oil and yield of soybean lines selected for increased protein. **Crop Science**, v.38, n. 3, p. 707-711, 1998.
- HENDERSON, C.R. Selection index and expected genetic advance. In: **Statistical genetics and plant breeding**. Washington DC: NAS-NRC Publication 982, p. 141-163, 1963.
- HENDERSON, C.R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, v. 32, n. 1, p. 69-83, 1976.
- HENDERSON, C.R. Best linear unbiased prediction of breeding values not in the model for records. **Journal of Dairy Science**, v. 60, n. 6, p.783-787, 1977.
- HENDERSON, C.R. Estimation of variance and covariance components. **Biometrics**, v. 17, n. 1, p. 226-252, 1953.
- HENDERSON, C.R. General flexibility of linear model for sire evaluation. **Journal of Dairy Science**, v. 57, n. 8, p. 963 – 972, 1974.
- HENDERSON, C.R.; KEMPTHORNE, O.; SEARLE, S.R.; von KROSIGK, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v. 15, p. 192-218, 1959.

- HERBACH, L.H. Properties of model II type of variance tests. A: optimum natura of the F-test for model II, in the balanced case. **Annals of Mathematical Statistics**, v. 30, n. 4, p. 939-959, 1959.
- HIGGINS, T.J.V. Synthesis and regulation of major proteins in seeds. **Annual Review of Plant Physiology**, v. 35, n. 1, p. 191-221, 1984.
- IAL - Instituto Adolfo Lutz. **Normas analíticas do Instituto Adolfo Lutz**. 3 ed. São Paulo, v. 1, 1985. 533p.
- IEMMA, M. **Uso do melhor preditor linear não viesado (BLUP) em análise dialélicas e predição de híbridos**. Piracicaba: ESALQ, 81p., 2003. (tese de mestrado).
- JENSEN, N.F. **Plant breeding methodology**. Ed. John Wiley & Sons, New York, 1988. 676p.
- JOHNSON, H.W.; BERNARD, R.L. Soybean genetics and breeding In: NORMAN, A.G. ed. **The soybean**. New York: Academic Press, p. 1-73, 1963.
- JOHNSON, H.W.; ROBINSON, H.F.; CONSTOCK, R.E. Genotypic and phenotypic correlations in soybeans and their implications in selection. **Agronomy Journal**, v. 47, n. 3, p. 477-483, 1955.
- KIIHL, R.A.S.; ALMEIDA, L.A.; ARANTES, N.E.; HIROMOTO, D.M.; SOUZA, P.I.M.; ASSUNÇÃO, M.S.; ZUFFO, N. Cultivar de soja EMBRAPA 20 (DOKO RC). In: Congresso Brasileiro de soja, 1. **Anais ...** Londrina: Embrapa – CNPSo, 1999, p. 487.
- KNOW, S.H.; TORRIE, J.H. Heritability of and interrelationships among traits of two soybean populations. **Crop Science**, v. 4, n.1, p. 196-198, 1964.
- LEFFEL, R.C. Registration of high-protein soybean germplasm lines BARC-6, BARC-7, BARC-8 and BARC-9. **Crop Science**, v. 32, n. 2, p. 502, 1992.
- LOPES, P.S.; MARTINS, E.N.; SILVA, M.A.; REGAZZI, A.J. **Estimação de componentes de variância**. Viçosa:UFV: cadernos didáticos 39, 1998. 61p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer, 1998. 980p.
- MALÉCOT, G. Les Mathematiques de l'Heredité. **Masson et Cie**, Paris, 1948.

- MARTINS, E.N.; LOPES, P.S.; SILVA, M.A.; TORRES JUNIOR, R.A.A. **Uso de modelos mistos na avaliação genética animal**. Viçosa: UFV:cadernos didáticos 18, 1997, 121p.
- McKENDRY, A.L.; McVETTY, P.B.E.; VOLDENG, H.D. Inheritance of seed protein and seed oil content in early maturing soybean. **Canadian Journal of Genetics and Cytology**, v. 27, n. 5, p. 603-607, 1985.
- MELLO FILHO, O. L. **Caracterização de populações de soja de alto teor de proteína quanto à produtividade e à qualidade de fisiologia das sementes**. Viçosa:UFV, 2002, 68p. (Tese de mestrado).
- MIRANDA, G.V. **Diversidade genética e desempenho de cultivares elites de soja como progenitores**. Viçosa: UFV, 117p., 1998. (Tese de doutorado).
- MOREIRA, M.A.; HERMODSON, M.; LARKINS, B.A.; NIELSEN, N.C. Partial characterization of the acidic and basic polypeptides of glycinin. **Journal of Biological Chemistry**, v. 254, n. 19, p. 9921 - 9926, 1979.
- MOREIRA, M.A.; REZENDE, S.J.; SEDIYAMA, C.S.; BARROS, E.G. Obtenção de cultivares de soja de sabor agradável e com sementes de alta qualidade fisiológica. In: TORRES, A. C.; CALDAS, L. S. (Eds.). **Técnicas e aplicações da cultura de tecidos de plantas**. Brasília: ABCPT, p. 417 – 426, 1990.
- MOREIRA, M.A. Programa de melhoramento genético da qualidade de óleo e proteína da soja desenvolvido na UFV. In: **Congresso brasileiro de soja**, 1. Anais ... p.99. Londrina, 1999.
- NAOE, L.K.; SEDIYAMA, C.S.; CRUZ, C.D.; MOREIRA, M.A.; MIRANDA, G.V. Divergência entre progenitores e variabilidade das populações segregantes de soja. **Revista CERES**, v. 48, n. 276, p. 223 - 237, 2001.
- O'DONALD, P. The distribution of genotypes produced by alleles segregating at a number of loci. **Heredity**, v. 26, n. 2, p. 233 - 241, 1971.
- PANTER, D.M.; ALLEN, F.L. Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing Parents. **Crop Science**, v. 35, n. 2, p. 397-405. 1995a.
- PANTER, D.M.; ALLEN, F.L. Using best linear unbiased predictions to enhance breeding for yield in soybean: II. Selections of superior crosses from a limited number of yield trials. **Crop Science**, v. 35, n. 2, p. 405-410, 1995b.
- PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, n. 2, p. 545-554, 1971.

- PIOVESAN, N.D. **Aplicação de cruzamentos dialélicos no melhoramento genético do teor protéico em soja**. Viçosa, UFV, 91p., 2000. (tese de mestrado).
- RAO, C.R. estimation of heteroscedastic variances in linear models. **Journal of American Statistical Association**, v. 65, n. 1, p. 161-172, 1970.
- RAO, C.R. Minimum variance quadratic unbiased estimation of variance components – MINQUE theory. **Journal of Multivariate Analysis**, v. 1, n. 2, p. 445-456, 1971.
- SAS/STAT. **User's Guide**. <http://jeff-lab.queensu.ca/stat/sas/sasman/sashtml/stat/index.htm> (último acesso dez/2003).
- SCHAEFFER, L.R. **Models and Computing Strategies in Animal Breeding**. University of Guelph, 98p, 1993.
- SEARLE, S.R. **Linear model**. Ed. John Wiley & Sons, New York, 533p, 1971.
- SEDIYAMA, T.; PEREIRA, M.G.; SEDIYAMA, C.S.; GOMES, J.L.L. **Cultura da soja I parte**. UFV, 1989, 96p.
- SEDIYAMA, T.; TEIXEIRA, R.C.; REIS, M.S. Melhoramento da soja. In: **Melhoramento de espécies cultivadas**, UFV, 1999.
- SHIBLES, R.M.; ANDERSON, I.C.; GIBSON, R.H. Soybeans. In: EVANS, L.T. **Crop physiology**. London: Cambridge University Press, p. 151-189, 1975.
- SHORTER, R.; BYTH, D.E.; MUNGOMERY, V.E. Estimates of selection parameters associated with protein and oil content of soybean seeds (*Glycine max* (L.) Merril). **Australian Journal of Agricultural Research**, v. 28, n. 2, p. 211-222, 1976.
- SMITH, J.D.; KINMAN, M.L. The use of parent-offspring regression as an estimate of heritability. **Crop Science**, v. 5, n. 3, p. 595–596, 1965.
- SMITH, P.K.; KROHN, R.I.; HERMANSON, G.T.; MALLIA, A.K.; GARTNER, F.H.; PROVENZANO, M.D.; FUJIMOTO, E.K.; GOEKE, N.M.; OLSON, B.J.; KLENK, O.C. Measurement of protein using bicinchoninic acid. **Analytical Biochemistry**, v. 150, p. 76-85, 1985.
- SILVA, M.A. **Melhoramento animal (métodos de seleção)**. Viçosa: UFV; Imprensa Universitária, 1982. 51p.

- SOARES, T.C.B. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa: UFV, 58p., 2000. (tese de mestrado).
- THORNE, J.C.; FEHR, W.R. Incorporation of high-protein exotic germplasm into soybean populations by 2- and 3-way crosses. **Crop Science**, v. 10, n. 6, p. 652-655, 1970.
- Van der WERF, J. Consequences of BLUP selection and genetic properties of the animal model. In: KINGHORN, B.; Van der WERF, J.; DEKKERS, J. **Quantitative genetics for new technologies in animal breeding**. (Course notes) Perth, Australia, 1999.
- VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.
- WEHRMANN, V.K.; FEHR, W.R.; CIANZIO, S.R.; CAVINS, J.F. Transfer of high seed protein to high-yielding soybean cultivars. **Crop Science**, v. 27, n. 5, p. 927-931, 1987.
- WILCOX, J.R.; CAVINS, J.F. Backcrossing high seed protein to a soybean cultivar. **Crop Science**, v. 35, n. 4, p. 1036-1041, 1995.
- WRIGHT, S. Coefficients of inbreeding and relationship. **American Naturalist** v. 56, p. 330-338, 1922.

APÊNDICE

QUADRO 1 – Descrição dos progenitores recorrentes de soja. Viçosa, Médias do verão 2001/inverno 2001/verão 2002/inverno 2002

	Variedades						
	OC 13	OC 14	CD 201	CD 202	CD 203	CD 205	CD 206
Cor da Flor	Roxa	Branca	Branca	Branca	Branca	Branca	Roxa
Cor da pubescência	Marrom	Marrom	Cinza	Cinza	Marrom	Marrom	Marrom
Habito de crescimento	Det	Det	Det	Det	Det	Det	Det
Cor do hilo	Preto	Marrom	Marrom	Marrom	Preto	Marrom	Preto
Cor do tegumento	Amarelo	Amarelo	Amarelo	Amarelo	Amarelo	Amarelo	Amarelo
Dias para maturação	112	112	115	117	120	132	123
Altura da planta (cm)	68	65	84	84	81	92	84
100 sementes (g)	13	12,8	16,1	15,6	15,4	15	14
Deiscência da vagem	Tolerante	Tolerante	Tolerante	Tolerante	Tolerante	Tolerante	Tolerante
Acamamento	Resist	Resist	Resist	Resist	Resist	Resist	Resist
Teor de proteína (%)	36,2	34,7	35,7	36,8	37,0	38,1	39,2
Identificação original*	OC 86-102	OC 85-08	OC 95(4)- 2422	OC 88-127	OC 88-161	CD 91- 671	CD 92-128
Pedigree*	FT-2 x União	Davis x União	OC4 (5) x Willian 20	CEPS 77-16 x Invicta	CEPS 77-16 x OC 73-397	BR 83-147 x OC 87-216	OC 87-5085 x FT- Abyara
Método de seleção*	SSD	SSD	Retr	SSD	SSD	SSD	SSD
Lançamento*	1991	1991	1996	1996	1996	1998	1999
Ano de registro*	-	-	1998	1998	1998	1998	2000
Validade da proteção*	-	-	2013	2013	2013	2013	2015
Área de adaptação*	PR	PR	RS, SC, PR, SP, MS	RS, SC, PR, SP, MS	PR, SC	RS, SC, PR, MG	PR, SP, MG

* Dados fornecidos pela COODETEC.

QUADRO 2 – Descrição dos progenitores doadores. Viçosa, Médias do verão 2001/inverno 2001/verão 2002/inverno 2002

	Linhagens	
	BARC 8	BR 80 14887
Cor da Flor	Branca	Roxa
Cor da pubescência	Marrom	Marrom
Habito de crescimento	Determinado	Determinado
Cor do hilo	Preto	Preto
Cor do tegumento	Amarela	Amarela
Dias para maturação	115	128
Altura da planta (cm)	68	72
100 sementes (g)	16,30	15,70
Acamamento	Resistente	Resistente
Teor de proteína (%)	51,30	49,82
Identificação*	PI 555398	o próprio
Pedigree*	CX797-21 x NC-2-62	?
Área de adaptação	Sem Indicação	Sem Indicação
Obtentor	USDA-ARS	Embrapa