

**NEWTON DENIZ PIOVESAN**

**ANÁLISES BIOMÉTRICAS E MOLECULARES VISANDO O  
DESENVOLVIMENTO DE LINHAGENS DE SOJA COM ALTO TEOR  
PROTÉICO E PRODUTIVAS**

**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  
2008**

**NEWTON DENIZ PIOVESAN**

**ANÁLISES BIOMÉTRICAS E MOLECULARES VISANDO O  
DESENVOLVIMENTO DE LINHAGENS DE SOJA COM ALTO TEOR  
PROTÉICO E PRODUTIVAS**

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: 24 de julho de 2008.

---

**Prof. Carlos Siguelyuki Sedyama  
(Co-Orientador)**

---

**Prof. Cosme Damião Cruz  
(Co-Orientador)**

---

**Dra. Cláudia Teixeira Guimarães**

---

**Prof. José Eustáquio de Souza Carneiro**

---

**Prof. Maurílio Alves Moreira  
(Orientador)**

A Deus,

Aos meus pais Adley (in memoriam) e Elsa, pelo amor e incentivo,

À minha esposa Lia, pelo amor, carinho, incentivo e compreensão,

Aos meus irmãos, pelo amor, apoio e amizade,

Aos meus parentes e amigos,

Dedico.

## AGRADECIMENTOS

A Deus, pela constante proteção e por todas as bênçãos.

À Universidade Federal de Viçosa, ao Colegiado do Bioagro e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade da realização do curso.

Ao Professor Maurílio Alves Moreira, pela amizade, pela orientação, pelos ensinamentos e pelo exemplo de dedicação ao progresso científico da UFV e do país.

Ao Professor Everaldo Gonçalves de Barros, pela amizade, pelos ensinamentos, pelo apoio, pela agradável convivência e pelo exemplo de dedicação didática e científica.

Ao Professor Carlos Sigueyuki Sedyama, pela amizade, pelos ensinamentos, pelo apoio, pelo respeito e pelo exemplo de ética, honestidade e competência.

Ao Professor Cosme Damião Cruz, pela amizade, pelos ensinamentos, pelas críticas e sugestões, pela disposição em ajudar e pelo exemplo de competência e simplicidade.

Ao Professor José Eustáquio de Souza Carneiro, pela amizade, pelas críticas e sugestões.

À Doutora Claudia Teixeira Guimarães, pela amizade, pelo respeito, pelas críticas e sugestões.

Ao professor Múcio Silva Reis, pela amizade, pela convivência e pelos ensinamentos.

Aos professores Sebastião, Wagner, Sérgio, Valéria, Elza, Maria Goreti e Elizabeth Fontes pela amizade, pela convivência e pelo incentivo.

A todos os Professores que ministraram as disciplinas do Programa de Genética e Melhoramento, pelos ensinamentos.

Aos amigos José Luis, Derval, Lúcio e Cupertino, pela amizade, pela ajuda e pela agradável convivência.

Aos colegas que aqui passaram e se tornaram grandes amigos: Tais, Telma, Rita, Inês, Márcia Flores, Márcia Costa, Sílvia, Ana Lília, Francismar, Lucimara, Lucinete, Telma, Luciano, Ivan, Abelmon, Adésio, Geraldo, Lucas, Vagner, Júlio, Carlos, Sidney, José Carlos, Francisco.

Aos amigos dos laboratórios: Loretta, Fernanda, Janaína, Suelen, Gabriela, Josie, Josiane, Danielle, Liliane, Gisele, Cassiana, Maria Andréia, Beatriz, Solange, Eleonice, Pedro Ivo, Kléver, Fábio, Demerson, Thiago, Rafael, Julierme, Leandro, Marcelo, Matheus, Renato, Bruno, Anderson e Daniel, pela amizade, pela ajuda e pela agradável convivência.

Aos funcionários do Bioagro e PPG: José Fausto, Edson, Pedro, Carlos, Adão, Expedito, Domingos, Gil, Evandro, Celso, Maurício, Alberto, Lili, Sandra Alessandra, Ludovina, Sueli, Margarida, Solange e Cláudia, pelo respeito, pela amizade e agradável convivência.

Aos funcionários do Programa Soja: Naldo, José Carlos, Glaucia, Cássio e Marcos (*in memoriam*), pela amizade, pelo auxílio e colaboração durante a realização deste trabalho.

Aos antigos e novos amigos de república: Sting, Berilo, Waldir, José Geraldo, Lucas, Marcinho, Marcio Garbeloto, Daniel, Hannuar, Ricardo, Rodrigo Tiussi, Rodrigo Darós, Vagner, Max, Paulo, Luciano e Diego, pela amizade e agradável convivência, que muito contribuíram para minha realização pessoal e profissional.

À minha esposa Eliéria, pelo amor, dedicação, companheirismo, carinho, compreensão e também por tornar a minha vida mais feliz.

A todos aqueles que colaboraram e torceram pelo meu sucesso.

## **BIOGRAFIA**

Newton Deniz Piovesan, filho de Adley Eridano Piovesan e Elsa Bissoli Piovesan, nasceu no dia 20 de abril de 1969, em Vitória, ES.

Em março de 1988, ingressou no curso de Agronomia da Universidade Federal de Viçosa, Viçosa, MG, graduando-se em janeiro de 1993.

No período de abril de 1993 a junho de 1996, atuou como Pesquisador/Bolsista do Programa de Melhoramento da Qualidade da Soja do Bioagro/UFV, desenvolvendo trabalhos na área de Genética Molecular de Plantas aplicados ao melhoramento da soja.

Em julho de 1996, foi contratado como técnico de nível superior pela Universidade Federal de Viçosa, como Engenheiro Agrônomo, para atuar na área de Genética Molecular aplicada ao melhoramento da soja.

Em outubro do ano de 1996, ingressou no Curso de Mestrado em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, MG, submetendo-se à defesa de tese em dezembro de 1999.

Em março do ano de 2004, ingressou no Curso de Doutorado em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, MG, submetendo-se à defesa de tese em julho de 2008.

## CONTEÚDO

RESUMO .....	xiii
ABSTRACT .....	xi
1. INTRODUÇÃO .....	01
2. REVISÃO BIBLIOGRÁFICA .....	04
2.1. Cultura da soja .....	04
2.1.1. Origem e evolução da soja cultivada .....	04
2.1.2. Aspectos taxonômicos, genéticos, botânicos e diversidade .....	05
2.1.3. Importância econômica da soja no Brasil .....	06
2.2. Melhoramento da soja .....	08
2.2.1. Herdabilidade .....	10
2.2.2. Correlação entre caracteres .....	16
2.3. Procedimentos de seleção .....	21
2.3.1. Delineamento com famílias intercaladas com testemunhas .....	22
2.3.2. Estratégias de seleção .....	24
2.3.3. Seleção simultânea de caracteres .....	26
2.3.4. Melhoramento da soja para alto teor protéico .....	28
2.4. Melhoramento genético de plantas auxiliado por marcadores moleculares .....	33
2.4.1. Mapeamento genético .....	35
2.4.2. Identificação de QTLs .....	37
2.4.3. Mapeamento de QTLs em soja .....	41
2.4.4. Seleção assistida por marcadores moleculares .....	45
3. OBJETIVOS E METAS .....	49
3.1. Objetivos .....	49
3.2. Meta .....	49
4. MATERIAL E MÉTODOS .....	50
4.1. Material genético e delineamento experimental .....	50
4.1.1. População para fins de seleção.....	50
4.1.2. População para fins de mapeamento de QTLs.....	52
4.2. Correção ambiental dentro do experimento .....	53
4.3. Procedimentos de seleção .....	54
4.3.1. Procedimentos de seleção no primeiro ciclo .....	54

4.3.2. Procedimentos de seleção no terceiro ciclo .....	55
4.4. Determinação do conteúdo de proteínas totais .....	56
4.5. Determinação do conteúdo de óleo .....	56
4.6. Determinação da composição do grão pela metodologia NIR .....	57
4.7. Análises estatísticas e genéticas da população .....	57
4.7.1. Análise de variância do experimento com linhas segregantes intercaladas pelos progenitores com informação entre e dentro de fileiras .....	57
4.7.2. Análise de variância do experimento com linhas segregantes intercaladas pelos progenitores com informação entre médias de fileiras .....	61
4.7.3. Análise de variância das linhagens selecionadas e plantadas em 2006/2007 .....	64
4.8. Esquema de seleção e predição de ganhos por seleção.....	66
4.8.1. Seleção entre famílias .....	66
4.8.2. Seleção entre e dentro de famílias .....	67
4.8.3. Seleção massal .....	68
4.8.4. Seleção combinada .....	68
4.8.5. Seleção simultânea utilizando índices de seleção .....	69
4.8.6. Seleção simultânea de caracteres entre e dentro de famílias .....	71
4.8.7. Ganhos de seleção em relação ao progenitor recorrente .....	73
4.9. Análises estatísticas e genéticas da população de mapeamento	73
4.9.1. Teste de normalidade .....	73
4.9.1.1. Teste de assimetria .....	73
4.9.1.2. Teste de curtose .....	74
4.9.1.3. Teste de Lilliefors .....	74
4.9.2. Análise de variância da população de mapeamento .....	76
4.10. Extração, amplificação e análise do DNA por marcadores SSR .....	76
4.11. Obtenção do mapa de ligação .....	77
4.12. Mapeamento de QTLs para conteúdo de proteína .....	78
4.13. Recursos computacionais .....	78
5. RESULTADOS E DISCUSSÃO .....	79
5.1. Análise das testemunhas e correção dos dados da população de seleção.....	79



5.2. Estimaco de parmetros genticos e predico de ganhos para produo de gros utilizando anlise de geraes com informao entre e dentro de famlias .....	79
5.2.1 Anlise de varincia .....	80
5.2.2. Predico de ganhos .....	82
5.3. Estimaco de parmetros genticos e predico de ganhos entre famlias para produo de gros por plantas, teor de protena e teor de leo .....	83
5.3.1. Anlise de varincia e estimaco de parmetros genticos .....	83
5.3.2. Predico de ganhos .....	86
5.4. Estimaco de parmetros genticos e predico de ganhos para produo de gros por planta e teor de protena utilizando anlise de geraes com informao entre e dentro de famlias .....	89
5.4.1. Anlise de varincia .....	90
5.4.2. Predico de ganhos .....	95
5.5. Avaliaco do desempenho e predico de ganhos em prognies F <sub>4:6</sub> em experimento realizado no ano agrcola 2006/2007 .....	97
5.5.1. Anlise de varincia .....	97
5.5.2. Predico de ganhos .....	100
5.6. Avaliaco gentica e estatstica da populao de mapeamento .....	105
5.6.1. Testes de normalidade .....	105
5.6.2. Anlise de varincia .....	105
5.6.3. Deteco de polimorfismos, anlise de segregaco e formao dos grupos de ligaco .....	107
5.6.4. Identificaco de QTLs .....	110
6. CONCLUSES .....	114
REFERNCIAS BIBLIOGRFICAS .....	116

## RESUMO

PIOVESAN, Newton Deniz, D.Sc., Universidade Federal de Viçosa, julho de 2008. **Análises biométricas e moleculares visando o desenvolvimento de linhagens de soja com alto teor protéico e produtivas.** Orientador: Maurílio Alves Moreira. Co-orientadores: Carlos Siqueyuki Sedyama, Cosme Damião Cruz e Everaldo Gonçalves de Barros.

A melhoria do potencial produtivo das cultivares de soja é o principal objetivo de todos os programas de melhoramento genético conduzidos no país. No entanto, em alguns programas de melhoramento, o aumento do teor de proteína nos grãos vem também sendo privilegiado. Com o propósito de desenvolver variedades produtivas e com alto conteúdo de proteína, este trabalho teve como objetivos específicos: a) comparar diferentes estratégias de seleção para a predição de ganhos em produção de grãos (PROD), altura da planta na maturação (APM), número de dias para maturação (NDM), teor de proteína (PTN) e teor de óleo (OL); b) comparar diferentes critérios de seleção por meio de índices de seleção; e c) identificar marcadores SSR ligados a QTLs que contribuam para o aumento do teor de proteína em uma população de linhagens quase isogênicas (NILs). Para alcançar estes objetivos, foram utilizadas duas populações neste trabalho, a primeira chamada população de seleção foi derivada de retrocruzamento parcial envolvendo uma linhagem de alto teor protéico e uma possuindo resistência ao herbicida glifosato. Nesta população RC<sub>1</sub>F<sub>4</sub> segregante, foram estimados os parâmetros genéticos e praticada a seleção durante três ciclos em função do ano agrícola. Na primeira etapa, utilizaram-se as seguintes estratégias: seleção entre (SE), seleção entre e dentro (SED), seleção massal (SM) e seleção combinada (SC). A herdabilidade no sentido restrito dentro apresentou reduzido valor para PROD (8,1%), comprometendo a eficiência de seleção em nível de indivíduo. Na segunda etapa, utilizaram-se 205 progênies, pré-selecionadas para PROD, na seleção simultânea de famílias para as características PROD, PTN e OL empregando os critérios de seleção baseados nos índices de Smith e Hazel (SH), Kempthorne e Nordskog (KN), Pesek e Baker (PB) e Pesek e Baker generalizado por Tai (PB-Tai). Correlações genéticas significativas só foram detectadas para PTN × OL com valor de -0,398. Foram selecionadas as 40 famílias que maximizaram os ganhos para PTN e que proporcionaram ganhos

moderados em PROD e redução mínima em OL. O índice PB-Tai apresentou melhores predições, em função do objetivo proposto, com um ganho de 4,2%, 8,9% e -3,68%, enquanto a seleção direta para PTN proporcionou um ganho de 4,45%, 3,91% e -5,85% para PTN, PROD e OL, respectivamente. Na terceira etapa, procedeu-se à estimação dos parâmetros genéticos e à predição de ganhos tanto para PROD e PTN com informação de família e planta, empregando-se diferentes estratégias e critérios de seleção. Para PTN, a herdabilidade restrita entre foi de 60,7% (superior à restrita dentro que foi de 15,6%), o que revelou uma pequena superioridade dos ganhos proporcionados por estratégias que também empregam a informação do indivíduo. Nesta etapa, a combinação da seleção simultânea utilizando o índice clássico de Smith e Hazel com a seleção praticada entre e dentro proporcionou ganhos significativos para PTN e PROD, porém inferiores aos da segunda etapa devido à redução da variabilidade genética. O terceiro ciclo de seleção foi realizado no ano 2006/2007 em um ensaio utilizando o DBC com as 84 famílias RC<sub>1</sub>F<sub>4.6</sub> selecionadas no primeiro e multiplicadas no segundo ciclo com o objetivo de aumentar PTN e PROD e reduzir NDM. Para este fim, utilizou-se a seleção simultânea baseada nos índices de SH, PB, KN e PB-Tai. Novamente o índice PB-Tai foi o critério de seleção mais indicado, pois proporcionou a seleção de um grupo de progênies que em média possuíam maior produtividade e menor ciclo do que a variedade comercial Monarca e outro grupo com maior PTN e menor NDM do que Monarca. Na segunda população, foi realizado um estudo de mapeamento para identificar marcadores SSR associados à QTLs que controlam o teor de proteína nos grãos de soja. Esta população chamada de população de mapeamento foi composta de 168 NILs (isolinhas na geração RC<sub>4</sub>F<sub>6</sub>) e foi obtida a partir de retrocruzamentos entre o genótipo com alto teor protéico BARC-8 com a variedade recorrente Monarca. De um total de 350 *primers* de microssatélites utilizados nos progenitores, apenas 20 foram polimórficos na população de NILs. Utilizando-se a análise de associação pelo método da ANOVA e regressão linear simples, foram encontrados dois marcadores (Satt 239 e Satt 384) associados à QTLs que explicavam 19,04% e 7,42% da variação do fenótipo. Empregando-se regressão linear múltipla com o procedimento *stepwise*, esses mesmos QTLs foram mantidos e explicaram 26,01% da variação fenotípica. Utilizando-se o mapeamento por intervalo

simples no GL E, foi estimado o efeito aditivo do QTL (+ 0,54) na expressão do caráter e determinada a posição do marcador em relação ao QTL. Esta marca foi posicionada muito próxima ou coincidente do QTL, o que resulta em uma marca extremamente confiável para futuros trabalhos de seleção assistida.

## ABSTRACT

PIOVESAN, Newton Deniz, D.Sc., Universidade Federal de Viçosa, July, 2008.

**Biometrics and molecular analysis aiming at the development of soybean lines with high-protein and productive.** Adviser: Maurílio Alves Moreira. Co-Advisers: Carlos Sigueyuki Sedyama, Cosme Damião Cruz and Everaldo Gonçalves de Barros

The improvement of the productive potential of soybean cultivars is the main objective of all the genetic breeding programs conducted in the country. However, in some breeding programs, the increase of the protein content in grains has been focused. In order to develop productive varieties with high protein content, this work had the following specific objectives: a) use different selection strategies for the prediction of gains in grain production (PROD), plant height in maturation (APM), number of days for maturation (NDM), protein content (PTN) and oil content (OL); b) compare different selection criteria through selection indices; and c) identify SSR markers linked to the QTL's which contribute for the increase of the protein content in a population of almost isogenic lineages (NIL's). To achieve these objectives, two populations were used in this work; the first one, called selection population, derived from the partial backcrossing involving a lineage with high protein content and another presenting resistance to the glyphosate herbicide. In the segregant population RC<sub>1</sub>F<sub>4</sub>, the genetic parameters were estimated and the selection was carried out during three cycles, due to the agricultural year. In the first phase, the following strategies were used: selection between (SE), selection between and inside (SED), massal selection (SM) and combined selection (SC). The heritability, in the restrict sense inside, presented reduced value for PROD (8.1%), affecting the selection efficiency at the individual level. In the second phase, 205 pre-selected progenies were used for PROD, in the simultaneous selection of families for the characteristics PROD, PTN and OL employing the selection criteria based upon the indices of Smith and Hazel (SH), Kempthorne and Nordskog (KN), Pesek and Baker (PB) and Pesek and Baker, generalized by Tai (PB-Tai). Significant genetic correlations were detected only for PTN × OL, with the value of -0.398. It was performed the selection of the 40 families

that maximized gains for PTN and provided moderate gains in PROD and minimum reduction in OL. The PB-Tai index presented better predictions as to the proposed objective, with gains of 4.2%, 8.9% e -3.68%, while the direct selection for PTN provided gains of 4.45%, 3.91% and -5.85% for PTN, PROD and OL, respectively. In the third phase, it was carried out the genetic parameter estimation and the gain prediction for PROD and PTN, with family and plant information, through the use of different strategies and selection criteria. For PTN, the restricted heritability between was of 60.7% (higher than the restrict inside, which was of 15.6%), which revealed a small superiority of gains provided by the strategies that employed the information of the individual. In this phase, the combination of the simultaneous selection using the classic index of Smith and Hazel, with the selection used between and inside, provided significant gains for PTN and PROD, lower, however, than those of the second phase, due to the reduction of the genetic variability. The third selection cycle was carried out in the year of 2006/2007, in an essay that used the DBC, with the 84 families  $RC_1F_{4:6}$  selected in the first cycle and multiplied in the second cycle, with the objective of increasing PTN and PROD and reducing NDM. For such, the simultaneous selection was used, based upon the indices of SH, PB, KN and PB-Tai. Again, the PB-Tai index was the most suitable selection criterion, because it provided the selection of a group of progenies which presented, in average, higher productivity and smaller cycle than the commercial variety Monarca and another group, with higher PTN and lower NDM, in comparison to Monarca. A mapping study was carried out, in the second population, in order to identify SSR markers associated to the QTL's, which control the protein content in soybean grains. This population, called mapping population, was composed of 168 NIL's (isolines in the generation  $RC_4F_6$ ) and was achieved from the backcrossing between the BARC-8 genotype, with high protein content, and the recurrent Monarca variety. Out of the 350 *primers* of microsatellites used in the parents, only 20 were polymorphic in the population of NIL's. Using the association analysis by the method of the ANOVA and the simple linear regression, two markers were found (Satt 239 and Satt 384), associated to QTL's, which explained 19.04% and 7.42% of the phenotype variation. With the use of the multiple linear regression, with the *stepwise* procedure, these very QTL's were maintained and explained 26.01%

of the phenotypical variation. With the use of the mapping by simple interval in the GL E, it was estimated the additive effect of the QTL (+ 0.54) in character expression and it was determined the position of the marker in relation to the QTL. This mark was placed very near to the QTL or on it, resulting in an extremely trustable mark for future works of assisted selection.

## 1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é a mais importante leguminosa cultivada no mundo, devido a seus altos conteúdos de proteína (40%) e óleo (20%), além da alta produtividade (2.823 kg ha<sup>-1</sup>) e baixo custo de produção (R\$ 1.100,00 a 1.400,00 ha<sup>-1</sup>) (CONAB, 2008). Essas excelentes características, aliadas às suas múltiplas utilidades, geram uma demanda mundial elevada e crescente, superior a 237,4 milhões de toneladas no ano de 2007 (SAFRAS e MERCADO, 2008).

O crescimento da produção e o aumento da capacidade competitiva da soja brasileira sempre estiveram associados aos avanços científicos e à disponibilização de tecnologias para o setor produtivo. Cultivares melhoradas, portadores de genes capazes de expressar alta produtividade, ampla adaptação e boa resistência ou tolerância a fatores bióticos ou abióticos adversos, representam usualmente uma das mais significativas contribuições à eficiência do setor produtivo.

O melhoramento da soja no Brasil tem sido um processo contínuo e dinâmico empenhado em resolver problemas reais ou potenciais objetivando o desenvolvimento de variedades mais produtivas, resistentes a pragas e doenças, de melhor qualidade nutricional, mais estáveis e mais bem adaptadas às diferentes regiões e sistemas de cultivo. Diversos programas de melhoramento genético foram essenciais para atender a esta demanda.

Contudo, nos últimos anos, têm sido frequentes os questionamentos, em alguns estados brasileiros, sobre a redução do teor de proteína nas novas cultivares de soja. Isso é reflexo do passado, pois os principais objetivos dos programas de melhoramento de soja eram produtividade e teor de óleo, o que reduziu a variabilidade e estreitou a base genética para teor de proteína nas atuais populações melhoradas, agravado pelo fato de existir uma correlação negativa elevada entre teor de óleo e proteína.

O melhoramento da soja está baseado, principalmente, nas hibridações para a formação de populações segregantes, a partir das quais são selecionados os genótipos mais promissores, segundo os objetivos do programa. Dessa forma, a possibilidade de prever ganhos por intermédio de estratégias de seleção constitui uma das mais importantes contribuições da Genética Quantitativa para o



melhoramento genético, pois as informações obtidas permitem decidir, com base científica e sem custo adicional, as técnicas de seleção, tornando mais eficiente o processo de obtenção de novas variedades. Entretanto, para a obtenção de genótipos superiores, torna-se necessária a reunião de uma série de atributos favoráveis que confirmem rendimento elevado e níveis adequados de óleo e proteína, satisfazendo assim as exigências do mercado.

Para tal, a teoria de índice de seleção permite combinar as múltiplas informações contidas na unidade experimental, de modo que seja possível a seleção com base em um complexo de variáveis que reúnam vários atributos de interesse econômico. Desse modo, o índice de seleção constitui-se em um caráter adicional, estabelecido pela combinação linear ótima de vários caracteres que permitem efetuar com eficiência a seleção simultânea.

Vários são os procedimentos utilizados pelo melhorista para identificar os genótipos superiores numa população. Alguns levam em conta o comportamento do indivíduo, enquanto outros se fundamentam, primeiramente, no desempenho da família e, secundariamente, na superioridade relativa dos indivíduos dentro da família. Há também, a estratégia que utiliza, simultaneamente, as informações do indivíduo e de seus parentes, chamada de seleção combinada, sendo esta sempre igual ou superior aos métodos de seleção individual, de famílias e entre e dentro de famílias (LUSH, 1964; FALCONER, 1987). A resposta à seleção combinada pode ser ainda maximizada se toda informação obtida nos parentes for também utilizada no processo seletivo (WEBER, 1982).

Muitos caracteres de importância agronômica apresentam baixas herdabilidades ou são de difícil medição e identificação, podendo a eficiência de sua seleção ser aumentada pelo uso de uma característica auxiliar que esteja correlacionada com o caráter principal, podendo ser esta característica auxiliar representada pelos marcadores moleculares.

Os marcadores moleculares têm sido sugeridos como ferramenta útil em vários aspectos do melhoramento de plantas, como a descrição de variedades, a construção de mapas genéticos, a medição das distâncias genéticas entre as linhagens, mapeamento e identificação de QTLs e a própria seleção assistida por marcadores (SAM), principalmente para características controladas por poucos genes, como a seleção de locos que conferem resistência a doenças.

A seleção de características agronômicas monitorada por marcadores moleculares baseia-se no princípio de que se um gene, ou um bloco de genes, encontra-se ligado a um marcador genético de fácil identificação, então, esse marcador pode ser usado para selecionar a característica de interesse em um programa de melhoramento. Segundo Hayward *et al.* (1994), o método mais eficaz para este tipo de associação envolve a produção de mapas genéticos detalhados nos quais são localizadas regiões contendo os locos para as características quantitativas (QTLs).

Para determinar se existe ligação genética entre marcadores e QTLs, é necessário haver desequilíbrio de ligação entre os alelos do loco marcador e alelos do QTL, o que é interpretado como evidência de ligação física entre os dois locos. Para detectar essas associações, foram desenvolvidos programas computacionais, tais como: o Mapmaker-QTL (LANDER *et al.*, 1987), QTLSTAT (KNAPP, 1995), PGRI (LIU, 1998), QTL-Cartographer (BASTEN *et al.*, 1999), e GQMol (<http://www.ufv.br/dbg/gqmol/gqmol.htm>).

Um ponto importante no uso da seleção assistida por marcadores é a identificação de associações específicas marcador – QTL significativas, tanto para características agronômicas em soja, como para resistência a doenças e para qualidade do grão. Para que a seleção assistida por marcadores (SAM) seja eficiente, é necessário inicialmente que se disponha de QTLs que no conjunto expliquem grande parte da variação fenotípica da característica. Para atingir tal objetivo, deve-se priorizar a saturação do genoma pelos marcadores e considerar ainda o tamanho e o tipo da população utilizada, o marcador empregado, o efeito do ambiente, os genótipos parentais, além da metodologia para detecção dos QTLs.

Este trabalho é parte integrante do programa de melhoramento da qualidade da soja desenvolvido pelo Bioagro/UFV que procura reunir conhecimentos tanto da biometria como da genética molecular e aplicá-los no desenvolvimento de variedades produtivas e com alto teor protéico.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. Cultura da soja

#### 2.1.1. Origem e evolução da soja cultivada

A soja é uma espécie exótica no continente americano. Segundo Probst e Judd (1973), essa cultura tem como centro de origem a região leste da China, onde é cultivada há mais de cinco mil anos, mas só foi domesticada por volta do século XI a.C. A partir daí a soja foi introduzida em outras regiões e países do Oriente, como Manchúria, Coréia, Japão, Rússia e países do sudeste da Ásia, tornando-se a base alimentar do seu povo. No Ocidente, sua introdução se deu a partir do século XVIII, em 1739, quando foram plantadas experimentalmente, na Europa, sementes recebidas pelo Jardim Botânico de Paris. Já no continente americano, o primeiro relato data de 1765, com registro de plantio na Geórgia (EUA). Sua disseminação pela Europa e América do Norte foi facilitada por ela ter sido domesticada em regiões de clima temperado com latitudes compreendidas entre 30 e 45°N e pelo fato de ser uma planta de dias curtos (noites longas). No Brasil, a introdução ocorreu na Bahia, em 1882, por Gustavo Dutra (BONETTI, 1981). Mas, durante muitas décadas, a soja foi plantada somente em caráter experimental por algumas instituições de pesquisa. As primeiras estatísticas revelando a utilização da soja para a produção de grãos no Brasil datam de 1941, no estado Rio Grande do Sul, onde foram produzidas 450 toneladas em 640 ha de área cultivada, com rendimento médio de 700 kg ha<sup>-1</sup> (VERNETTI *et al.*, 1983).

A partir da década de 1960, a cultura da soja passou a adquirir importância no país, inicialmente na região Sul (latitudes 30 a 22°S), chamada de região tradicional, onde apresentou melhor adaptação devido à semelhança com as regiões de cultivo do sul dos EUA. Justamente por essa semelhança ambiental, diversas introduções dos EUA foram utilizadas nos anos iniciais de plantio. Posteriormente, tais introduções serviram de base genética para o desenvolvimento de novas variedades obtidas pela hibridação e seleção (VERNETTI *et al.*, 1983).

Na década de 1960, a produção no Paraná cresceu rapidamente até atingir 60.000 toneladas. A partir da década de 70, a cultura da soja evoluiu

significativamente nos estados produtores, não só no Sul, mas também nos estados do Centro-Oeste do Brasil com o desenvolvimento de novas cultivares adaptadas a diferentes regiões agroclimáticas do País, tornando-se o Brasil o segundo maior produtor de soja do mundo.

O desenvolvimento de variedades adaptadas às baixas latitudes do centro-oeste brasileiro foi o principal desafio dos programas de melhoramento para a expansão das fronteiras da soja no Brasil. Inicialmente, por meio de hibridações e seleção de materiais de ciclo tardio, foram desenvolvidas as primeiras variedades adaptadas. Vale destacar o papel do IAC e da UFV que desenvolveram as primeiras variedades bem sucedidas para o cerrado, no caso a IAC-2 e a UFV-1, respectivamente. Posteriormente, a incorporação da característica período juvenil longo (P JL) nas novas variedades – procedimento desenvolvido pelo Dr. Romeu A. S. Kiihl, pesquisador do IAC e posteriormente da Embrapa Soja – permitiu a consolidação definitiva da cultura no cerrado brasileiro.

#### 2.1.2. Aspectos taxonômicos, genéticos, botânicos e diversidade

A soja cultivada (*Glycine max* (L.) Merril) pertence ao gênero *Glycine*, família Leguminosae e subfamília Papilionoideae. Dentro do gênero *Glycine*, podem-se distinguir três subgêneros: *Bracteata*, *Glycine* e *Soja*. No gênero *Glycine* Willd, incluem-se espécies perenes no subgênero *Glycine* e espécies anuais no subgênero *Soja*. A soja cultivada (*Glycine max*) e seu ancestral anual selvagem (*Glycine soja*) pertencem ao subgênero *Soja*. São espécies alotetraplóides ( $2n = 40$ ), com comportamento meiótico de um diplóide normal, carregando genomas similares, sendo compatíveis sexualmente, podendo produzir híbridos férteis e apresentam alto grau de auto-polinização (SINGH e HYMOWITZ, 1988; PALMER *et al.*, 1996). Existem ainda 22 outras espécies perenes reconhecidas dentro do gênero *Glycine*, das quais *Glycine tabacina* e *Glycine tomentella* são neopoliplóides ( $2n = 78, 80$ ; HYMOWITZ, 2004). Segundo Shoemaker *et al.* (2003), o genoma da soja é constituído por 1.115 milhões de pares de base por genoma haplóide, sendo mais de 35% constituído de heterocromatina, com braço curto em seis dos 20 cromossomos bivalentes completamente heterocromáticos (SINGH e HYMOWITZ, 1988). Apresenta ainda, de 40 a 60% de seqüências de DNA repetitivas, sendo que das sondas de RFLP testadas, cerca de 90% detectaram locos duplicados, dentre essas sondas

aproximadamente 60% detectaram três ou mais locos (SHOEMAKER *et al.*, 1996), podendo ter como explicação o seu passado de poliploidia.

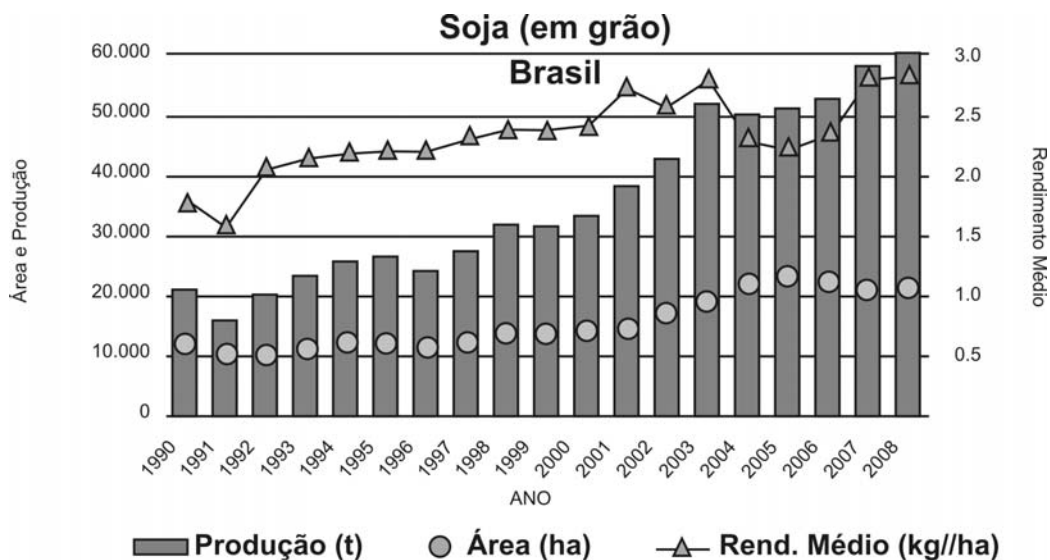
Por ser o país de origem da soja, a China é o berço da diversidade de germoplasma dessa espécie, possuindo uma coleção superior a 30.000 acessos cultiváveis, dos quais cerca de 5.000 são silvestres. Além da coleção chinesa, grande parte da variabilidade genética da soja é também mantida e conservada em Bancos de Germoplasma existentes em vários países orientais e ocidentais. O Departamento de Agricultura dos Estados Unidos da América (USDA), por exemplo, mantém uma coleção de aproximadamente 15.000 acessos de soja. No Brasil, existe uma coleção de germoplasma com aproximadamente 4.000 acessos (genótipos), que está sendo conservada em câmaras climatizadas no Banco Ativo de Germoplasma da Embrapa Soja, em Londrina, PR. Esta mesma coleção é mantida, em condições de conservação a longo prazo, na Coleção Base da Embrapa Recursos Genéticos, em Brasília (DF). A maioria desses acessos foi introduzida da coleção de germoplasma norte-americana e é composta principalmente por genótipos procedentes da China, do Japão e de outros países onde ocorreu a diversificação da espécie (ALMEIDA *et al.*, 1997).

Nas coleções de germoplasma de soja, a variabilidade genética para caracteres fisiológicos, morfológicos e agrônômicos é considerada bastante ampla. Entretanto, essa variabilidade existente nas coleções de germoplasma ainda é pouco utilizada nos programas de melhoramento da soja. Isso se deve, em geral, à opção dos melhoristas de utilizar variedades de alto desempenho (cultivares e linhagens mais adaptadas) em seus programas de cruzamentos, o que pode conduzir a um estreitamento da variabilidade genética. Muitos pesquisadores consideram que a base genética das cultivares comerciais de soja é restrita (HIROMOTO e VELLO, 1986), o que representa um fator de risco para a estabilidade da cultura, haja vista a não existência de nenhuma variedade comercial resistente ao fungo causador da ferrugem asiática da soja.

### **2.1.3. Importância econômica da soja no Brasil**

A soja é a mais importante leguminosa cultivada no mundo, sendo amplamente cultivada em vários países. Atualmente, os principais países produtores são os Estados Unidos, o Brasil, a Argentina e a China. O Brasil destaca-se como o segundo maior produtor mundial e o primeiro maior exportador

de soja. A produção brasileira de soja no ano agrícola 2007/2008 foi de 58,726 milhões de toneladas em uma área de 21,2 milhões de hectares, com uma produtividade média próxima a 3.000 kg/ha (Figura 1).



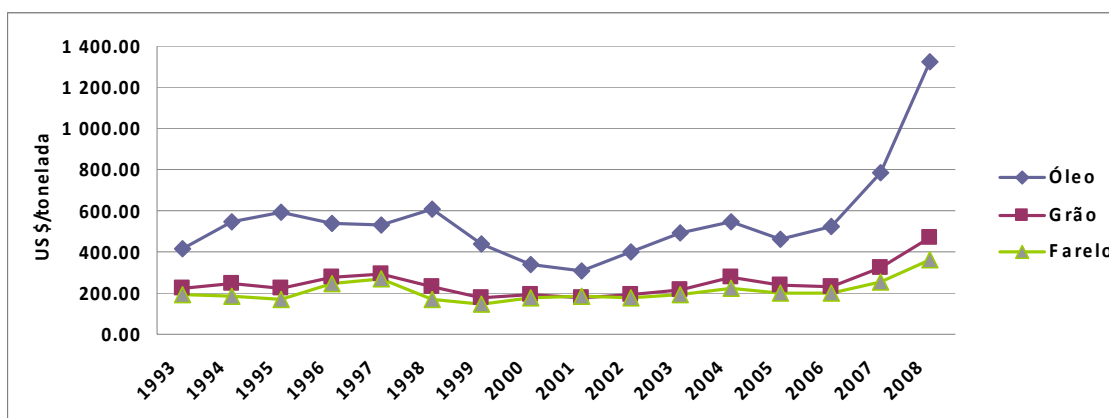
**Figura 1** - Evolução da soja no Brasil (Fonte: CONAB 2008).

A soja é uma das principais fontes de proteína e óleo, sendo responsável por cerca de 60% do suprimento de proteína de origem vegetal e 25% do óleo do mundo. No Brasil, a soja é responsável por fornecer aproximadamente 94% do farelo protéico e 90% dos óleos vegetais aqui consumidos. Destacam-se como os principais estados produtores Mato Grosso, Paraná, Rio Grande do Sul e Goiás (EMBRAPA, 2007).

A soja é o principal *commodity* agrícola do Brasil. O complexo soja, em 2007, exportou aproximadamente 23,8 milhões de toneladas de grãos, 12,9 milhões de toneladas de farelo e 2,5 milhões de toneladas de óleo gerando em torno de 10,18 bilhões de dólares (CONAB, 2008). Com isso, a cadeia da soja destaca-se, tanto pela sua expressiva participação nas exportações sob a forma de farelo, óleo e grãos, como também para o suprimento do mercado interno de óleos comestíveis e derivados protéicos.

O preço do grão, do óleo e do farelo de soja no mercado mundial vem mostrando muita volatilidade nos últimos anos, principalmente o óleo, com variações de mais de 300% entre os vários anos. Entre 1997 e 2008, a cotação do óleo de soja exportado pelo Brasil variou entre o máximo de US\$ 1350/t em maio de 2008 e o mínimo de US\$ 300/t, em 2001 (Figura 2). Essas cotações

determinaram a maior ou menor participação do Brasil no comércio mundial do produto e também sua viabilidade de uso na produção do biodiesel.



**Figura 2** - Evolução do preço do complexo soja (Fob Porto Paranaguá) (Fonte: Abiove, 2008).

## 2.2. Melhoramento da soja

A criação de novas cultivares tem sido uma das tecnologias que mais têm contribuído para os aumentos de produtividade e estabilidade de produção, sem custos adicionais ao agricultor. Uma cultivar de soja deve ter alta produtividade, estabilidade de produção e ampla adaptabilidade aos mais variados ambientes existentes na região onde é recomendada. É importante destacar que as cultivares de soja lançadas atualmente têm capacidade de produtividade superior a 4.000 kg/ha, muito maior que a média nacional, porém, fatores bióticos e abióticos interagem para que essa capacidade de produção não seja alcançada.

De acordo com Sedyama *et al.* (1999), programas de melhoramento de soja devem focar os seguintes caracteres: adaptação quanto ao ciclo, hábito de crescimento, altura da planta e da inserção da primeira vagem, acamamento, deiscência das vagens, qualidade da semente, adaptação quanto ao nível de fertilidade do solo, resistência a doenças e pragas, aspectos qualitativos, tolerância a herbicida, produtividade e período juvenil longo.

O ganho genético proporcionado pelas novas cultivares ao setor produtivo tem sido muito significativo, maior que 1,0% ao ano, por isso é que o desenvolvimento de variedades é dinâmico, necessitando sempre de novos cruzamentos, seleção de progênes e ensaios de avaliação, a fim de serem

lançadas cultivares cada vez mais produtivas. A utilização da tecnologia RR (Roundup Ready) continua crescendo entre os produtores e vem também sendo o caráter de maior oferta pelas instituições de pesquisa, por meio das novas cultivares. Em 2005, estavam registradas no Ministério da Agricultura, Pecuária e Abastecimento apenas 22 cultivares tolerantes ao glifosato. Já em abril de 2008 já estavam inscritas 144 cultivares RR no Registro Nacional de Cultivares. Um novo desafio para os melhoristas de soja é o desenvolvimento de variedades mais tolerantes à ferrugem asiática. No desenvolvimento de novas variedades, devem-se combinar os diferentes alelos dos genes de resistência Rpp1, Rpp2, Rpp3 e Rpp4 com genes que conferem resistência horizontal. O correto posicionamento e a colheita antecipada são novas exigências dos produtores, por isso deve-se dar prioridade às variedades mais precoces.

Nos últimos anos, tem sido freqüente no Brasil o questionamento, em alguns estados, sobre a redução do teor de proteína nas novas cultivares de soja. Este fato pode ser explicado pela não priorização deste caráter no processo de seleção e pela correlação negativa entre produção de grãos e teor de proteínas. Portanto, torna-se necessário que os programas de melhoramento da soja se empenhem em aumentar essa característica nas novas cultivares, pois empresas exportadoras de farelo de soja têm reduzido o preço pago pelos grãos oriundos das variedades que possuem baixos teores de proteína.

Esta tendência foi evidenciada por Bonato *et al.* (2000), que avaliaram o desempenho de cultivares do Rio Grande do Sul para a produção de óleo e proteína comparando cultivares desenvolvidas antes e após 1990. De acordo com os autores, as cultivares lançadas entre 1991 e 1996 apresentaram menor teor de proteína e maior teor de óleo que as cultivares desenvolvidas em anos anteriores, devido principalmente ao emprego das cultivares União e Industrial como genitores dos programas de melhoramento.

Já nos Estados Unidos, de acordo com Wilcox e Guodong (1997), a composição média das sementes das cultivares comerciais, com poucas exceções, tem permanecido inalterada, nos últimos 70 anos, em aproximadamente 400 g de proteína e 210 g de óleo por kg de grãos. No entanto, Voldeng *et al.* (1997) afirmaram que de 1976 a 1992, o aumento do rendimento dos grãos das cultivares cultivadas naquele país foi de 0,7% ao ano, enquanto o nível de proteína foi reduzido em 4 g/kg/ano e o teor de óleo aumentou em 4 g/kg/ano.



Então, para tentar solucionar este problema é necessário que os programas de melhoramento também considerem como critério de seleção o teor de proteína. Para que este objetivo seja alcançado, é necessário um profundo conhecimento da magnitude da variância genética das características agronômicas e de qualidade a serem melhoradas, principalmente produtividade, teor de óleo e proteína, bem como a de seus componentes, como estimativas de herdabilidades e de correlações. Este conhecimento possibilita a avaliação da potencialidade da população para o melhoramento e facilita as decisões de escolha do método de seleção mais eficiente a ser empregado, disso dependendo a escolha acertada dos genitores e do método de melhoramento mais eficiente (VERNETTI, 1983). Portanto, enfocando esses conhecimentos é que serão feitas algumas considerações a seguir.

### **2.2.1. Herdabilidade**

O conceito de herdabilidade, introduzido para separar as diferenças genéticas e não genéticas entre indivíduos, é de fundamental importância no melhoramento genético. A estimativa do coeficiente de herdabilidade atende a duas finalidades básicas: a primeira, a de demonstrar a relativa facilidade com que diferentes caracteres são selecionados em um determinado esquema de melhoramento; e a segunda, a de permitir que estimativas dos progressos esperados com a seleção sejam obtidas.

Hanson (1963) define herdabilidade como a fração da variância fenotípica de uma unidade a ser transmitida para a progênie, ou em termos de unidade de seleção, a fração do diferencial de seleção esperado pelo ganho, quando a seleção é praticada em uma referida unidade. Para Falconer (1987), a herdabilidade de um caráter métrico é uma das mais importantes propriedades de uma população, porque expressa a proporção da variância total atribuída ao efeito médio dos genes, o que determina o grau de semelhança entre parentes, sendo a seleção efetuada, fundamentalmente, nessa semelhança. Dois tipos de herdabilidades podem ser estimados em uma população: a herdabilidade no sentido amplo e a herdabilidade no sentido restrito. Quando a herdabilidade é alta, a seleção nas gerações iniciais de autofecundação é eficaz. Por outro lado, sendo o seu valor baixo, a seleção deve ser praticada apenas nas gerações mais avançadas, uma vez que o aumento da homozigose se reflete em um aumento da

herdabilidade no sentido restrito, conseqüência do aumento da variância genética aditiva. Entretanto, deve-se atentar para a avaliação desses coeficientes, pois maiores respostas à seleção não estão necessariamente associadas a características de maior herdabilidades, uma vez que altos valores de herdabilidade podem ocorrer em características de pequena variância genética aditiva.

Dudley e Moll (1969) discutem que a estimativa de herdabilidade refere-se a uma característica e a uma população, sendo específica para as condições experimentais nas quais os genótipos foram estudados. Assim, é difícil generalizar estimativas de uma população para outra, ou para diferentes condições experimentais. Além disso, é normal a ocorrência de erros associados às estimativas de herdabilidades e de outros componentes da variância genética, principalmente quando são estimadas pela análise de variância, existindo uma grande faixa de variação nas estimativas da herdabilidade de uma mesma característica. Portanto, é necessária precaução na comparação entre herdabilidades obtidas em diferentes populações e condições experimentais.

Na literatura, são descritas várias metodologias para obtenção de estimativas do coeficiente de herdabilidade. A utilização da variabilidade entre as plantas na geração  $F_2$  em relação à variabilidade entre as plantas dos progenitores foi utilizada por Mahmud e Kramer (1951), para a obtenção do coeficiente de herdabilidade. Warner (1952) propôs uma metodologia para estimar a herdabilidade com base nos progenitores, gerações  $F_1$  e  $F_2$  e dos dois retrocruzamentos. Hanson *et al.* (1956) determinaram estimativas do coeficiente de herdabilidade com base nos componentes de variância. Frey e Horner (1957) propuseram estimar a herdabilidade em unidades-padrão, obtidas calculando-se a regressão sobre os dados codificados em termos de desvio padrão. Smith e Kinman (1965) utilizaram fatores de correção do coeficiente de regressão pai-filho para estimar a herdabilidade, utilizando progenitores endogâmicos para algumas gerações de auto-fecundação.

De maneira geral, a produção de grãos é um caráter que apresenta de média a baixa herdabilidade, que pode ser atribuída ao comportamento quantitativo desse caráter, em função do grande número de genes que o controlam, permitindo uma maior influência ambiental e a existência de genes que apresentam ação devido aos desvios de dominância. Já as características teor de

proteína e teor de óleo apresentam herdabilidades de média a elevada, como poderá ser observado nos diversos trabalhos descritos a seguir.

Hanson e Weber (1961) estimaram a herdabilidade de diversas características de soja, estudando as gerações  $F_2$ ,  $F_3$  e  $F_4$  e obtiveram estimativas por meio dos componentes de variância, sendo estimados os valores de 32, 79 e 86% para maturidade nas gerações  $F_2$ ,  $F_3$  e  $F_4$ , respectivamente. Para altura de planta na maturidade, foram estimados valores de 35, 73 e 82% para as mesmas gerações; para peso de sementes os valores foram de 35, 46 e 58%; para teor de óleo os valores foram de 34, 63 e 74%; e para produção foram estimados valores de 12, 40 e 53% para as mesmas gerações, respectivamente.

Johnson e Bernard (1963), estudando herdabilidade para várias características em soja, dentre elas o teor de proteína, mostram que em gerações precoces ( $F_2$ ) esta apresenta baixos valores, quando comparadas com gerações mais avançadas ( $F_4$ ). Isso implica dificuldade de identificar e selecionar genótipos superiores em gerações precoces no caso  $F_2$ , resultando em pouco progresso esperado por seleção.

Kwon e Torrie (1964), procurando estimar a herdabilidade para caracteres de interesse em soja, estudaram duas populações derivadas da seleção de 64 plantas  $F_2$  para cada cruzamento. Essas populações foram avaliadas na geração  $F_3$  em um ano e um local e as gerações  $F_4$  e  $F_5$  em dois anos e dois locais. Os parâmetros genéticos extraídos da análise de variância proporcionaram estimativas de herdabilidade livre dos efeitos da interação genótipo x ambiente. Os autores verificaram que a interação genótipo x ano foi maior que a interação genótipo x local, para muitas características, principalmente produção de sementes. A herdabilidade para produção de sementes foi mais baixa devido a uma maior interação genótipo x ano, ficando na ordem de 3 e 10%, respectivamente, para cada população. Para os caracteres teor de óleo e proteína, a herdabilidade só foi estimada para um dos cruzamentos, apresentando valores de 51 e 57%, respectivamente.

Para estudar a potencialidade do uso de germoplasma exótico de alto teor de proteína como fonte de variabilidade para os teores de proteína e óleo em soja, Thorne e Fehr (1970) sintetizaram seis populações oriundas de cruzamentos biparentais que foram avançadas até a geração  $F_6$ , e seis populações de cruzamentos tri-parentais que foram avançadas até a geração  $F_5$ . Foram estimados os coeficientes de herdabilidade por meio dos componentes de

variância para as 12 populações derivadas dos cruzamentos bi e tri-parentais, sendo que estes valores variaram de 81 a 96% para teor de proteína e de 77 a 95% para teor de óleo nas sementes.

Shannon *et al.* (1972), trabalhando com seis populações com interesse para teor de proteína e produção de grãos, obtiveram estimativas de herdabilidade com base nos parâmetros genéticos obtidos da análise de variância. Essas populações foram sintetizadas a partir de um dialelo entre dois progenitores de alto teor de proteína e dois progenitores de alta produtividade e foram avançadas até a geração  $F_4$  e testadas em dois locais. As estimativas de herdabilidade nas seis populações variaram de 0 a 73% para produção de grãos e de 81 a 96% para teor de proteína.

Shorter *et al.* (1976), estudando o comportamento de três populações obtidas por três pares de cruzamentos biparentais envolvendo progenitores com baixo, médio e alto teor protéico, avaliadas em duas localidades nas gerações  $F_3$  e  $F_4$ , obtiveram estimativas de herdabilidade com base nos componentes de variância da geração  $F_4$ , variando de 73 a 86% para teor de proteína e de 75 a 84% para teor de óleo. Quando considerados os dois ambientes, a herdabilidade diminui tanto para teor de proteína, variando de 54 a 60%, como para teor de óleo variando de 47 a 84%. A herdabilidade também foi calculada no sentido restrito com base na regressão de  $F_4$  sobre  $F_3$ , tendo encontrado estimativas para teor de proteína variando de 21 a 57% e para teor de óleo variando de 1 a 66%.

Brim e Burton (1979) conduziram um programa de seleção recorrente a fim de aumentar o teor de proteína nas sementes em duas populações, sendo uma delas obtida do cruzamento entre duas linhagens produtivas e com altos teores de proteína e a outra proveniente do primeiro ciclo de retrocruzamento de uma linhagem produtiva com nove introduções de plantas com alto teor de proteína. Ao final de vários ciclos de seleção recorrente, foram estimadas as herdabilidades realizadas baseando-se na regressão cumulativa em resposta ao diferencial de seleção, estas estimativas referem-se à herdabilidade no sentido restrito e assumiram valores que variavam de 20 a 29% para a primeira população e de 34 a 37% para a segunda população.

Erickson *et al.* (1981) estimaram a herdabilidade para teor de proteína em quatro populações obtidas a partir de quatro pares de cruzamentos biparentais interespecíficos entre *Glycine max* e *Glycine soja*. Para o cálculo da herdabilidade no sentido restrito, foi utilizada a regressão das médias de famílias

$F_3$  sobre as famílias parentais  $F_2$  e para a herdabilidade no sentido amplo, foi calculada pelos componentes de variância para as médias de famílias  $F_3$ . A estimativa de herdabilidade para teor de proteína, no sentido restrito, foi de 27% e no sentido amplo de 78%, com base nas 100 famílias  $F_3$  avaliadas em dois locais.

Cianzio e Fehr (1982) avaliaram a herdabilidade no sentido amplo para teor de proteína em populações de retrocruzamento sintetizadas a partir de cruzamentos biparentais, contrastando para teor de proteína e produção de grãos. O coeficiente de herdabilidade foi calculado até a segunda geração de retrocruzamento, e os valores variaram de 76 a 89%, tendo sido também constatada uma diminuição na variância genética para teor de proteína à medida que se aumentava o grau de recuperação do progenitor recorrente por meio dos retrocruzamentos.

Mckendry *et al.* (1985), estudando a herança dos teores de óleo e proteína das sementes por meio de cruzamentos biparentais entre três variedades de ciclo precoce, avaliaram as gerações  $P_1$ ,  $P_2$ ,  $F_1$ ,  $F_2$ ,  $F_3$  e o retrocruzamento. A herdabilidade no sentido amplo foi calculada a partir dos componentes de variância de geração  $F_2$  extraídos da análise efetuada segundo modelo de Mather e Jinks, sendo estimada para teor de óleo variando de 57 a 68% e para teor de proteína variando de 68 a 83%. A herdabilidade no sentido restrito foi obtida de duas maneiras: a primeira utilizando a variância genética aditiva em  $F_2$  dividida pela variância fenotípica, que variou de 57 a 62% para teor de óleo e para teor de proteína variou de 34 a 72%; e a outra maneira utilizando a regressão das médias de famílias  $F_3$  sobre as famílias parentais  $F_2$ , apresentando valores para teor de óleo variando de 43 a 70% e para teor de proteína variando de 76 a 81%.

Buscando estabelecer estratégias para o melhoramento simultâneo para produção de grãos e teor de proteína, Prado (1994) avaliou quatro cruzamentos biparentais de soja na geração  $F_8$ , obtendo estimativas de herdabilidades no sentido restrito com base em médias para o caráter produção de grãos e teor de óleo variando de 50 a 70%, para teor de proteína a herdabilidade foi mensurada em apenas um cruzamento e apresentou valor de 61%.

Wilcox (1998) conduziu um programa de seleção recorrente para aumentar o teor de proteína nas sementes em duas populações de soja, utilizando macho-esterilidade. Estimativas de herdabilidade foram calculadas a cada ciclo de seleção recorrente, que variaram de 55 a 89% entre o primeiro e o

sétimo ciclo. Houve também um aumento da variabilidade para teor de proteína, à medida que se aumentava o número de ciclos de seleção recorrente.

Helms e Orf (1998), trabalhando com dez diferentes populações na geração  $F_5$ , testadas em sete ambientes diferentes, calcularam a herdabilidade considerando-se todos os ambientes para teor de proteína, encontrando valores elevados para a maioria das populações.

Azevedo Filho *et al.* (1998) avaliaram linhagens de soja na geração  $F_{6:7}$ , obtidas de 45 cruzamentos dialélicos em dois locais contrastantes para saturação de alumínio, tendo observado ampla variação entre os coeficientes de herdabilidade para um mesmo caráter em diferentes cruzamentos, e atribuíram isso à diversidade genética entre os genitores e ao tipo de resposta às variações ambientais. Obtiveram coeficientes de herdabilidade de 58,0 e 84,0% para número de dias para a maturação, 64,0 e 62,0% para altura da planta, e de 38,0 e 38,0% para produção de grãos, para alta e baixa saturação de alumínio, respectivamente. Valores mais elevados de herdabilidades para altura de planta na maturação e produção de grãos foram encontrados por Costa *et al.* (2004), quando avaliaram 1.200 plantas derivadas de 26 populações  $F_2$  derivadas de cruzamentos biparentais em delineamento em blocos aumentados de Federer (1955).

Mauro *et al.* (2000), avaliando as características altura da planta na maturação, altura de inserção da primeira vagem e a produção de grãos em 30 linhagens de soja por três anos, encontraram estimativas de herdabilidade de 46,0, 29,0 e 20,0% para altura da planta, altura de inserção da primeira vagem e produção de grãos, respectivamente. Observaram ainda que o caráter altura da planta foi menos sensível às alterações ambientais do que altura de inserção da primeira vagem e produção de grãos.

Backes (2000), utilizando um esquema experimental de famílias intercaladas de progenitores, avaliou cinco populações de soja, três na geração  $F_6$  e duas na  $F_5$ , tendo obtido estimativas de herdabilidade no sentido restrito entre famílias para a característica altura de planta na maturação variando de 59,4 a 92,2% para este caráter, valores próximos aos encontrado por Santos *et al.* (1995).

Teixeira (2003), estudando uma população nas gerações  $F_2$  e  $F_3$ , encontrou estimativas de herdabilidades de 56,2% para o teor de proteínas totais,

62,1% para o teor da subunidade 7S e 43,4% para o teor da subunidade 11S, mostrando a possibilidade de melhoramento para a qualidade protéica da soja.

Reis *et al.* (2004), avaliando a característica produção de grãos por planta em três populações de soja na geração  $F_5$  em um delineamento experimental de famílias intercaladas de progenitores, encontraram valores variando de 25,8 a 77% e 5,53 a 37,7% para as herdabilidades no sentido restrito, entre e dentro de famílias, respectivamente.

Miranda (2006), avaliando 207 famílias nas gerações  $F_2$  e  $F_3$  derivadas de cruzamentos contratantes para teor de proteína, encontrou estimativa de herdabilidade no sentido restrito, calculada pela regressão pai/filho segundo Smith e Kinman (1965) no valor de 43,40%. Pela análise de variância entre famílias  $F_3$ , a estimativa de herdabilidade no sentido amplo foi de 62,36% para teor de proteína, 71,5% para teor de óleo e de 56,5% para produção de grãos.

## 2.2.2. Correlação entre caracteres

Um aspecto genético de grande valor para o melhoramento de plantas, que deve receber atenção dos melhoristas, é a estimativa das correlações entre caracteres, pois reflete o grau de associação entre esses caracteres. Seu conhecimento é importante porque mostra como a seleção para um caráter influencia a expressão de outros caracteres. Nos programas de melhoramento, geralmente, além de visar ao aprimoramento de um caráter principal, busca-se também manter ou melhorar a expressão de outros caracteres.

Segundo Falconer (1987), a correlação entre caracteres tem duas causas principais: uma genética e outra de ambiente. As causas genéticas são principalmente em virtude da pleiotropia e do desequilíbrio de ligação gênica, sendo o desequilíbrio de ligação gênica uma causa transitória, particularmente quando as populações são derivadas de progenitores divergentes e de fecundação cruzada. A correlação de ambiente entre caracteres ocorre quando eles são influenciados pelas mesmas diferenças do ambiente. A correlação pode ser estimada em três níveis: fenotípico, genotípico e de ambiente. No nível fenotípico, estão envolvidas tanto as causas genéticas como as de ambiente. Se ambos os caracteres correlacionados possuem baixas herdabilidade, então a correlação fenotípica é determinada, principalmente, pela correlação de ambiente.

Se eles têm altas herdabilidades, então a correlação genética é a mais importante.

Os estudos sobre correlações genóticas, fenotípicas e de ambiente em soja têm envolvido os caracteres avaliados desde o florescimento até a maturação, destacando-se a produtividade e seus componentes e, mais recentemente, envolvendo análises quantitativas e qualitativas de óleo e proteína (SHARMA, 1979; CECON *et al.*, 1993; AKHTER e SNELLER, 1996a,b; TAWARE *et al.*, 1997; WILCOX, 1998; XINHAI *et al.*, 1999; MORRISON *et al.*, 2000; WILCOX e SHIBLES, 2001; ALT *et al.*, 2002).

Nos trabalhos em que foram estimadas as correlações genóticas e fenotípicas constata-se que geralmente elas têm o mesmo sinal e que, com poucas exceções, as correlações genóticas são maiores que as fenotípicas (JOHNSON *et al.*, 1955; ANAND e TORRIE, 1963; KWON e TORRIE, 1964; FEHR e WEBER, 1968; SMITH e WEBER, 1968).

Wakawkar *et al.* (1976) sugeriram levar em conta, em programas de seleção visando ao aumento de produção em soja, os caracteres número de ramos, número de dias para florescimento e maturação e número de vagens por planta, pois foram os caracteres que tiveram altos efeitos positivos diretos e indiretos, via todos os componentes de produção.

Estimativas da magnitude das correlações genóticas e fenotípicas foram obtidas em soja por Johnson *et al.* (1955) para todas as possíveis combinações entre 24 caracteres avaliados. Por meio de duas populações segregantes provenientes de cruzamentos avaliados na geração F<sub>4</sub> em dois e três locais, respectivamente, os autores estimaram as correlações genóticas e fenotípicas entre teor de óleo e de proteína em -0,69 e -0,70 para uma das populações e -0,48 e -0,48 para a outra. As correlações entre teor de proteína e produção de grãos ficaram em -0,12 e -0,80 para a primeira e -0,64 e -0,33 para a segunda população.

Kwon e Torrie (1964) estimaram correlações genéticas e fenotípicas em duas populações segregantes obtidas de cruzamentos biparentais e avaliadas na geração F<sub>5</sub> em cinco locais. As correlações genéticas e fenotípicas entre teor de óleo e de proteína foram estimadas em -0,66 e -0,54, respectivamente, em apenas um dos cruzamentos. Produção de grãos e teor de proteína estavam negativamente correlacionadas em -0,58 e -0,42, respectivamente.



Thorne e Fehr (1970) estudaram cruzamentos entre cultivares de soja adaptadas e introduções com alto teor de proteína. As populações foram avançadas até a geração  $F_6$  para cruzamentos biparentais e  $F_5$  para cruzamentos triparentais. A correlação fenotípica estimada para as diversas populações variou entre -0,84 e -0,54 para teor de proteína e teor de óleo, e entre -0,38 e -0,28 para teor de proteína e produção de grãos, o que seria suficiente para permitir uma seleção de linhagens com alta produção e alto teor de proteína.

Shannon *et al.* (1972), trabalhando com seis populações obtidas por cruzamentos biparentais entre quatro progenitores, sendo dois progenitores produtivos e dois com alto teor protéico e avaliado na geração  $F_4$  em dois locais, estimaram correlações genéticas e fenotípicas para os caracteres teor de proteína, teor de óleo e produção de grãos. A correlação genética entre teor de proteína e teor de óleo, e entre teor de proteína e produção de grãos variou entre -0,92 e -0,46 e entre -0,61 e -0,24, respectivamente. Uma correlação fenotípica foi estimada variando entre -0,68 e -0,30 para teor de proteína e teor de óleo, entre -0,21 e 0,15 para teor de proteína e produção de grãos.

Shorter *et al.* (1976), estudando o comportamento de três populações obtidas por três pares de cruzamentos biparentais envolvendo progenitores com baixo, médio e alto teor protéico, avaliados em duas localidades nas gerações  $F_3$  e  $F_4$ , obtiveram correlações genotípicas entre teor de proteína e produção de grãos variando entre -0,50 a 0,34.

Miller e Fehr (1979), após praticar um ciclo de seleção recorrente para teor de proteína em soja, estimaram a correlação genética entre teor de óleo e teor de proteína em -0,69 e a correlação fenotípica em -0,72. Obtiveram aumento de 1,5% na média para teor de proteína, mas com uma redução de 0,70% no teor de óleo. Em condições semelhantes, praticaram um ciclo de seleção para baixo teor de óleo, obtendo um aumento indireto no teor de proteína de 0,80% e uma redução no teor de óleo de 0,80%.

Openshaw e Hadley (1984) estudaram, em duas populações que incluíam progenitores de alto teor de proteína e que foram avaliados em  $F_3$  e  $F_4$ , a adequação de índices de seleção para modificar a concentração do teor de proteína, óleo e açúcares nas sementes de soja. As correlações genéticas entre teor de proteína e teor de óleo e entre teor de proteína e conteúdo de açúcares encontrados foram de -0,83 e -0,68 para uma população e -0,68 e -0,60 para a outra população.

Holbrook *et al.* (1989) avaliaram a utilização de índices de seleção para aumento de produção com a manutenção do teor de proteína na semente e, apesar da freqüente correlação negativa entre estes dois caracteres, obtiveram aumentos significativos na produção sem alteração no teor de proteína por meio da utilização de índices de seleção restrito.

Xu e Wilcox (1992) estimaram correlações fenotípicas em uma população submetida a quatro ciclos de seleção recorrente, utilizando macho-esterilidade para promover os intercruzamentos, tendo sido as populações selecionadas a cada ciclo para teor de proteína. Coeficientes de correlações foram estimados variando de 0,18; 0,15; -0,21; -0,13; -0,17, respectivamente, da população inicial até o quarto ciclo de seleção para produção de grãos e teor de proteína.

Wilcox e Cavins (1995), estudando correlações entre teor de proteína e produção de grãos, por meio de uma série crescente de retrocruzamentos entre dois progenitores contrastantes para estas duas características, verificaram uma correlação negativa decrescente entre teor de proteína e produção de grãos à medida que se avançaram os retrocruzamentos, obtendo ao final do 3º retrocruzamento linhagens produtivas e com alto teor de proteína.

Scott e Kephart (1997) estudaram o comportamento de linhagens F<sub>5</sub> derivadas de oito cruzamentos entre materiais adaptado × adaptado e adaptado × não-adaptado com níveis diferentes de proteína, tendo encontrado correlações positivas entre produção de grãos e teor de proteína em todos os cruzamentos, com uma variação de 0,10 a 0,58. Quando praticada uma seleção com 10% de selecionados, obtiveram ganhos diretos para teor de proteína que atingiram até 4,0% e reduziram a produção e teor de óleo em até -1,2% e quando se praticou a seleção direta para produção obtiveram ganhos de até 16,5%, mas com uma redução de 1,1% para teor de proteína.

Xinhai *et al.* (1999) determinaram correlações genéticas entre teor de proteína e produção de grãos e entre teor de proteína e componentes de produção em uma população derivada do cruzamento entre *Glycine max* e *Glycine soja*, utilizando cinco diferentes métodos de seleção, e encontraram correlações entre teor de proteína e produção de grãos variando de -0,117 a -0,271.

Wilcox e Shibles (2001) estudaram a associação entre teores de proteína, óleo, carboidratos totais, sacarose e enxofre em uma população que

continha uma ampla variação para teor de proteína. Neste trabalho, 43 linhagens com concentração de proteína, variando de 413 a 468 g kg<sup>-1</sup>, foram cultivadas em três ambientes. As linhagens apresentaram variações significativas para cada uma das características avaliadas. Os teores de carboidratos totais não tiveram associação com produtividade. O aumento no teor de proteína ocorreu à custa do teor de óleo, carboidratos totais e sacarose com coeficientes de correlação de -0,88, -0,71 e -0,66, respectivamente. No entanto, as proporções S/N foram constantes em todos os níveis de proteína. Estes autores concluíram que a diminuição no teor de carboidratos com o aumento do teor proteína deverá contribuir para um aumento do valor nutricional do farelo obtido dessas linhagens melhoradas. Além disto, a consistente proporção S/N em todas as linhagens com diferentes teores de proteína indica que os teores de aminoácidos sulfurados não foram reduzidos com o aumento de proteína.

Farias Neto *et al.* (2001) estimaram a correlação fenotípica e genotípica entre porcentagem de óleo na semente (%OL), produtividade de óleo (PO) e produtividade de grãos (PG). Neste experimento, foram utilizadas progênies F<sub>3:4</sub> provenientes de cruzamentos em cadeia envolvendo 40 genitores. As correlações fenotípicas estimadas foram de:  $r_{PG \times PO} = 0,98$ ;  $r_{PG \times \%OL} = 0,04$  e  $r_{PO \times \%OL} = 0,17$ . Para o produtor, um maior rendimento ou produtividade de óleo por área pode ser mais desejável que um maior teor de óleo na semente. Isso se dá porque o rendimento de óleo por área está também ligado à produtividade da lavoura. A baixa correlação entre PO e %OL pode ser explicada por este balanço. Em contrapartida, para a indústria de esmagamento, sementes de soja com maior teor de óleo (%OL) são mais desejáveis, pois levam a um maior rendimento na extração de óleo.

Miranda (2006) estimou correlações genotípicas entre os caracteres produção de grãos (PRO), número de dias para maturação (NDM), teor de óleo (OLEO) e teor de proteína (PTN) em 207 famílias F3 de soja derivadas de cruzamentos contratantes para teor de proteína e óleo. Encontrou valor de correlação entre PRO e PTN igual a -0,08; entre PRO e OLEO igual a 0,35; entre PRO e NDM igual a 0,71; entre NDM e PRO igual a -0,44; entre NDM e OLEO igual a 0,43; e entre PTN e OLEO um alto valor negativo igual a -0,89. Estes valores de correlações demonstram apenas dificuldades para obter ganhos simultâneos para óleo x proteína e produção x precocidade.

### 2.3. Procedimentos de seleção

Os programas de melhoramento genético da cultura são essenciais para atender à crescente demanda por maiores produções, possibilitando, pela criação de variabilidade e ampliação da base genética, a seleção dos melhores genótipos de uma população, capazes de superar os patamares atuais de produtividade.

Contudo, selecionar progênies superiores não é tarefa fácil, uma vez que os caracteres de importância, em sua maioria, são quantitativos e apresentam comportamento complexo, por serem influenciados pelo ambiente e estarem inter-relacionados, de tal forma que a seleção de um provoca uma série de mudanças em outros (CRUZ, 2001).

Partindo-se de uma população base de soja, várias estratégias de seleção podem ser usadas. No entanto, o melhorista sempre se preocupa em usar uma estratégia que promova o maior ganho por unidade de tempo e custo. Para tal fim, dispõe-se de técnicas de genética quantitativa que permitem prever o quanto conseguirá de ganho na geração seguinte, utilizando determinada estratégia de seleção. Assim, pode-se decidir pela melhor estratégia, com bases científicas, tendo em vista o quanto de progresso será conseguido na geração seguinte (VENCOVSKY, 1987).

O tamanho e o número de populações que podem ser avaliadas para determinados caracteres, assim como os objetivos de um programa de melhoramento, são os determinantes quando da escolha do método de melhoramento de plantas. Métodos tradicionais de melhoramento de plantas autógamas são utilizados para desenvolver novas variedades de soja, tais como: introdução de germoplasma melhorado, genealógico, população ('bulk'), genealógico modificado (SSD), retrocruzamentos e seleção recorrente. Combinações e ou variações desses métodos são também bastante utilizadas pelos melhoristas durante o processo de avanço de gerações segregantes. Recentemente, técnicas da biotecnologia como clonagem, transformação genética e marcadores moleculares têm sido empregados com sucesso no desenvolvimento de novas variedades, como no caso da soja Roundup Ready.

As atividades da maioria dos programas de melhoramento envolvem quatro etapas principais: escolha dos parentais, cruzamentos entre parentais e obtenção de genótipos segregantes, avanço das gerações até a homozigose e avaliação das linhagens. A etapa intermediária correspondente ao avanço das

gerações de endogamia tem sido feita de forma relativamente rotineira, com a finalidade principal de desenvolver genótipos homozigóticos, que, estando livres das combinações alélicas heterozigóticas e tendo fixado as combinações epistáticas favoráveis, aumentam a eficiência dos testes de desempenho agrônomo. Além disso, as linhagens homozigóticas podem ser avaliadas com precisão experimental superior, pois dispõem de um maior número de sementes para locais, épocas de cultivo e anos agrícolas. Por outro lado, o avanço das gerações de endogamia tem como desvantagens o aumento do número de anos de cada ciclo do programa de melhoramento e a demanda adicional de recursos humanos e financeiros. Essas limitações poderiam ser contornadas pela eficiente e eficaz escolha dos parentais e pela avaliação e seleção dos genótipos promissores logo nas gerações iniciais, de maneira que somente estes genótipos selecionados sejam avançados até originarem linhagens superiores.

### **2.3.1. Delineamento com famílias intercaladas com testemunhas**

Os delineamentos experimentais obedecem a três princípios básicos: repetição, casualização e controle local. Então, será da observação destes princípios que depende a maior ou menor validade das conclusões obtidas. O ideal é utilizar grande número de repetições, buscando maior precisão experimental, pela diminuição dos efeitos ambientais. A casualização tem por finalidade evitar que as repetições de um determinado tratamento venham a ser continuamente favorecidas ou desfavorecidas pelo ambiente, quando distribuídas nas diversas unidades experimentais, enquanto o controle local deve ser exercido pela utilização de áreas mais homogêneas ou pela divisão da área em blocos. A observação destes princípios garante validade às inferências estatísticas, dando confiabilidade à comparação de médias no processo de seleção (GOMES, 1990; RAMALHO *et al.*, 2000).

Porém, na avaliação de gerações segregantes em programas de melhoramento de plantas, a exemplo da soja, pelo menos três fatores limitam a utilização de repetições: a pequena quantidade de sementes disponíveis; o grande número de materiais a serem avaliados; e a falta de homogeneidade genética nas primeiras gerações de autofecundação, mesmo entre sementes oriundas de uma única planta. Assim, não se pode considerar como repetições parcelas entre as quais, além das diferenças ambientais, exista heterogeneidade

genética. Mesmo na avaliação de linhagens e clones, muitas vezes o que dificulta a utilização de delineamentos experimentais é o grande número de tratamentos a serem avaliados.

Uma alternativa para contornar estas dificuldades é a utilização de blocos aumentados (FEDERER, 1955). Neste delineamento, são considerados dois tipos de tratamentos: os tratamentos comuns a todos os blocos, ou seja, as testemunhas, e o segundo tipo, os tratamentos regulares, que aparecem em apenas um dos blocos, ou seja, sem repetições. Neste delineamento, as observações dos tratamentos regulares de cada bloco são corrigidas em função da diferença entre a média geral das testemunhas e a média das testemunhas naquele bloco. Desta forma, todos os tratamentos regulares de um mesmo bloco são ajustados por um único índice obtido com base nas testemunhas.

Em soja, Spehar (1994) mostrou que estes delineamentos foram eficientes na identificação de genótipos tolerantes à toxidez de alumínio. As vantagens em relação aos ensaios com repetições recaíram sobre o custo efetivo. O autor recomenda testes para a seleção de variedades contrastantes a serem usadas como testemunhas, esperando que elas representem melhor o germoplasma e que a correção para os efeitos de blocos seja mais precisa e útil às comparações. Entretanto, Gomes (1995) e Rios (1997) afirmam que este tipo de delineamento não se mostrou eficiente para a estimação de correlações entre caracteres, alertando para a necessidade de estudos sobre sua utilização na estimação de parâmetros genéticos. Contudo, os autores compartilham da opinião de que o emprego deste tipo de delineamento seja factível de ser implementado em programas de melhoramento de soja, sobretudo para os testes de genótipos nas etapas iniciais, quando muitas progênies precisam ser avaliadas e a disponibilidade de sementes é baixa.

Backes *et al.* (2003) implementaram modificações no delineamento de blocos aumentados em que as famílias sem repetições são intercaladas regularmente com testemunhas com repetições, possibilitando a estimação do componente ambiental associado à variância fenotípica da população em estudo e, conseqüentemente, os parâmetros genéticos e com base na repetições das testemunhas procede-se à correção do efeito ambiental nas famílias.

### 2.3.2. Estratégias de seleção

A possibilidade de o melhorista prever o ganho esperado pela seleção antes mesmo que ela seja realizada é uma das maiores contribuições da genética quantitativa (RAMALHO e VENCOVSKY, 1978; HALLAUER e MIRANDA FILHO, 1981). Para CRUZ e REGAZZI (1997). De acordo com a estratégia de seleção e o ganho que ela proporcionará, pode-se orientar, de maneira mais efetiva, um programa de melhoramento, bem como prever o sucesso do esquema seletivo adotado, decidindo, com bases científicas, quais esquemas podem resultar em maior ganho genético.

Os progressos genéticos referem-se às alterações observadas nas características de interesse, durante um ciclo de seleção, com a recombinação e/ou a multiplicação das unidades selecionadas. Tais modificações ocorrerão em magnitude e sentido variados, dependendo da estratégia e dos critérios de seleção adotados. Assim, uma das atribuições mais importantes do melhorista de plantas é identificar critérios ou estratégias de seleção capazes de promover alterações, no sentido desejado, nas características de interesse dentro de um programa de melhoramento.

O progresso genético direcionado em qualquer espécie está associado à existência de variabilidade genética, à seleção natural e/ou artificial e ao ajuste dos genótipos aos ambientes existentes. Comprovada a presença da variabilidade genética e sobretudo seu valor em relação à variação não-genética, a seleção assume grande importância no progresso genético. Por fim, a seleção objetiva acumular alelos favoráveis à característica de interesse em determinada população.

Vários são os procedimentos ou estratégias utilizados pelo melhorista para identificar os genótipos superiores numa população. Alguns levam em conta o comportamento do indivíduo, enquanto outros se fundamentam, primeiramente, no desempenho da família e, secundariamente, na superioridade relativa dos indivíduos dentro da família.

Há, também, a estratégia que utiliza, simultaneamente, as informações do indivíduo e de seus parentes, chamada de seleção combinada, sendo sempre superior aos métodos de seleção individual, de famílias e entre e dentro de famílias (FALCONER, 1987). A seleção combinada proporciona resposta superior ou, no mínimo, equivalente à seleção de famílias ou à seleção massal (LUSH,

1964), e a resposta à seleção pode ser maximizada se toda informação obtida nos parentes for também utilizada no processo seletivo (WEBER, 1982).

Os testes de progênies, instrumentos dos mais importantes para o trabalho do melhorista, têm sido usados na estimação dos parâmetros genéticos e seleção de indivíduos entre e dentro de progênies, quando se procura avaliar a magnitude e a natureza da variância genética disponível com vistas a quantificar os ganhos com a seleção e predizer o melhor método de seleção a ser utilizado. Nos referidos testes, podem ser discriminados os indivíduos superiores a serem utilizados em novos ciclos de melhoramento.

A seleção combinada, por outro lado, baseia-se em um índice que considera, simultaneamente, o comportamento do indivíduo e sua família. Dessa forma, em função dos pesos de ponderação do indivíduo e da família, é possível selecionar indivíduos superiores, porém de famílias de desempenho intermediário, ou ainda, indivíduos de desempenho intermediário pertencentes a famílias superiores (FALCONER, 1987; RESENDE e HIGA, 1994). Essa idéia não é nova no melhoramento genético. Lush (1945, 1947) realizou a primeira análise dessa questão e propôs a seleção combinada com base em valores individuais com recuperação da informação de famílias. Nesse método, a seleção é baseada em medidas genéticas (valores genéticos líquidos) e não fenotípicas, dos candidatos à seleção. Nesse contexto, o uso da seleção combinada poderá proporcionar ganhos mais elevados que os obtidos pela seleção massal ou de família.

Erickson *et al.* (1981) estudaram seleção para proteína em gerações precoces do cruzamento entre *Glycine soja* e *Glycine max*. A herdabilidade do teor de proteína foi medida em F<sub>2</sub> e F<sub>3</sub> e o ganho por seleção massal e seleção de famílias foi comparado na geração F<sub>3</sub> destes cruzamentos. *G. soja* foi usado como progenitor masculino em todos os cruzamentos e a porcentagem de seleção foi de 10%. A herdabilidade para proteína na geração F<sub>2</sub> foi de 27%, obtida pela regressão pai-filho. A herdabilidade no sentido amplo em famílias F<sub>3</sub> cultivadas em dois locais em uma estação foi de 78%. A média do teor de proteína em cada população selecionada foi maior do que a média da população não selecionada. As médias dos teores de proteína dos métodos de seleção foram: seleção massal em F<sub>2</sub> e F<sub>3</sub> 48,0%; seleção massal em F<sub>3</sub> seguida de uma única semente por descendência, 48,8%; seleção entre famílias F<sub>3</sub>, 47,6%; e seleção entre e dentro de famílias F<sub>3</sub>, 47,5%. O método de seleção massal foi superior ao método de seleção de famílias em nível de 1,0%. Estes autores comentam que, apesar de a



seleção precoce ter elevado o teor de proteína da população, a segregação continuou a originar genótipos de baixo teor de proteína. Conseqüentemente, passos adicionais de seleção deverão ser necessários.

Reis *et al.* (2004), em trabalho de melhoramento de soja, procuraram predizer o ganho genético para produção de grãos, advindo de diferentes procedimentos de seleção e indicar, entre as populações estudadas, a mais promissora quanto à expectativa de ganhos, adotando para isso quatro estratégias de seleção para comparar os ganhos: seleção entre e dentro de famílias; seleção combinada e seleção individual para produção de grãos; e, também, para a seleção livre de pesos e de parâmetros, o número de dias e altura da planta por ocasião da maturação. Os indivíduos selecionados na geração F<sub>5</sub> foram conduzidos e observados em campo no esquema cultivares-padrão intercalados às linhas segregantes. Os resultados indicaram que, considerando a família no processo seletivo, a estratégia de seleção combinada foi a mais promissora. Entretanto, a estratégia de seleção individual deteve maior expectativa de progresso genético entre as quatro estratégias. Quanto aos cruzamentos, o cruzamento entre os progenitores CEPS 89-26 e FT-Cristalina foi o que apresentou maior expectativa de ganho genético.

### 2.3.3. Seleção simultânea de caracteres

Para a obtenção de genótipos superiores, é necessária a reunião de uma série de atributos favoráveis que confirmem rendimento comparativamente mais elevado e satisfaçam as exigências do mercado. Com isso, a seleção baseada em uma ou poucas características mostra-se inadequada, conduzindo a um produto final superior apenas em relação aos caracteres selecionados (CRUZ e REGAZZI, 1997).

Assim, a seleção simultânea de um conjunto de caracteres de expressividade econômica aumenta a chance de êxito de um programa de melhoramento. Para tal, a teoria de índice de seleção permite combinar as múltiplas informações contidas na unidade experimental, possibilitando a seleção com base em um complexo de variáveis que reúna vários atributos de interesse econômico. Desse modo, o índice de seleção constitui-se num caráter adicional, estabelecido pela combinação linear ótima de vários caracteres, que permite efetuar, com eficiência, a seleção simultânea, proporcionando maiores ganhos

totais (CRUZ e REGAZZI, 1997). Dois índices de seleção não restritos muito empregados no melhoramento genético são o índice clássico proposto por Smith (1936) e Hazel (1943) e o índice com base nos ganhos desejados proposto por Pesek e Baker (1969). Mas, segundo Resende (2002), muitas vezes torna-se importante impor restrições a estes índices, visando a atender às necessidades do melhorista. Uma dessas restrições impostas ao índice clássico apresentada por Kempthorne e Nordskog (1959) é a manutenção de  $r$  características inalteradas, ao mesmo tempo em que se procura maximizar o ganho genético em  $n$  características, dentre um total  $q = (n + r)$ . Outro tipo de restrição apresentada por Tai (1977) ao índice proposto por Pesek e Baker (1969) é a inclusão de caracteres tidos como secundários no índice para auxiliar no melhoramento dos caracteres principais, necessitando apenas dos ganhos desejados dos caracteres e impondo que o valor econômico dos caracteres secundários sejam nulos.

Segundo Cruz (1990), o índice de seleção tem sido mais freqüentemente utilizado nos programas de melhoramentos animal e vegetal de espécies perenes e semi-perenes, com uso muito restrito em espécies anuais e de ciclo curto.

Trabalhos similares já foram realizados com milho (GARCIA e SOUZA JÚNIOR, 1999; GRANATE *et al.*, 2002), eucalipto (PIRES, 1996; PAULA *et al.*, 2002), arroz (MORAIS, 1992; RODRIGUEZ *et al.*, 1998) e soja (OLIVEIRA *et al.*, 1999; REIS *et al.*, 2004; COSTA *et al.*, 2004; 2008).

Shorter *et al.* (1976) estudaram a eficiência relativa da seleção direta e indireta por meio do índice na obtenção de ganhos para a produtividade de proteína e de óleo e para os teores de proteína e óleo em três populações obtidas por três pares de cruzamentos biparentais envolvendo progenitores com baixo, médio e alto teores protéicos, avaliados em duas localidades nas gerações  $F_3$  e  $F_4$ . Nesse estudo, obtiveram-se ganhos preditos pela seleção direta variando de 5,9 a 26,1% para as produtividades de proteína e óleo, enquanto para os teores destas duas características obtiveram-se ganhos variando de 1,5 a 6,3%. Utilizando o índice de seleção, foi obtida uma eficiência de até 14% a mais do que na seleção direta para promover o aumento da produtividade dessas duas características.

Miller e Fehr (1979) avaliaram a eficiência do melhoramento de uma população para conteúdo de proteína em semente de soja por seleção recorrente direta para alto teor de proteína e indireta baseada em baixo teor de óleo, tendo o teor de proteína aumentado de 43,1% para 44,6% no ciclo 0 da população de alto

teor de proteína e para 43,9% na população de baixo teor de óleo. A seleção direta causou diminuição significativa no teor de óleo e carboidratos, e a seleção indireta diminuiu significativamente o teor de óleo, mas não alterou o teor de carboidratos. Estes autores concluíram que a seleção indireta (baseada em baixo teor de óleo) não pode ser considerada um método eficiente para melhoramento do teor de proteína quando o custo e o tempo para análise de proteína não forem fatores limitantes.

Estudando a adequação da seleção direta e a do índice proposto por Smith-Hazel para modificar a concentração do teor de proteína, óleo e açúcares nas sementes de soja em duas populações oriundas de progenitores de alto teor protéico e avaliados em F<sub>3</sub> e F<sub>4</sub>. Openshaw e Hadley (1984) observaram que a seleção direta para proteína proporcionou um ganho de 1,42% na proteína e uma redução de 1,14% no teor de óleo para as progênies selecionadas. Quando se utilizou o índice de Smith-Hazel com pesos econômicos iguais para proteína e óleo e ganhos desejados de 1,0% e 0,25% para as respectivas características, foram obtidos ganhos de 0,28% e 0,14% para proteína e óleo, respectivamente.

#### **2.3.4. Melhoramento da soja para alto teor protéico**

A incorporação de novas características aos grãos abre perspectivas de maior eficiência de produção, processamento ou utilização, transformando grãos comuns, comercializados como simples *commodities*, em produtos especializados com alto valor agregado e comercial. Neste contexto, os programas que visam ao desenvolvimento de cultivares de melhor qualidade nutricional e tecnológica têm se concentrado no aumento do teor de óleo e proteínas, na alteração da composição da fração óleo e na diminuição de fatores antinutricionais das sementes, além, é claro, da máxima produtividade, objetivo comum de todo e qualquer programa de melhoramento.

A semente de soja apresenta composição média em base seca de 40% de proteína, 21% de lipídios, 5% de cinzas e 34% de carboidratos com os teores de proteína e óleo variando de 32 a 44% e 17 a 24%, respectivamente, nas variedades comerciais (SEDIYAMA *et al.*, 1999).

É importante destacar que o farelo protéico resultante da extração do óleo é o principal produto de uma unidade de esmagamento de soja (Tabela 1), e o teor de proteína é a característica principal para determinação de sua qualidade

e valor. Deve-se considerar no melhoramento de soja visando à alteração na composição, a existência de uma correlação genética negativa de alta magnitude entre teor de óleo e teor de proteína. Assim, dificilmente uma variedade de soja irá apresentar teores elevados de óleo e proteína simultaneamente. Portanto, uma das duas características precisa ser priorizada no processo de seleção e será o mercado que irá estabelecer a importância relativa destas duas frações, ficando a cargo do melhorista estabelecer qual das duas frações será priorizada em algum dos seus subprogramas. O grande desafio dos melhoristas de soja é o de obter ganhos para proteína ou óleo sem haver redução ou até mesmo conseguir ganhos moderados para a característica secundária (teor de óleo ou proteína). Para isto, o emprego de índices de seleção parece ser a metodologia mais adequada. Não se pode pensar apenas no momento atual para definir quais dos dois componentes deverão ser priorizados em um programa de melhoramento de soja, pois o óleo de soja está muito valorizado, sendo o valor da tonelada de óleo cerca de 3,5 vezes maior do que o preço da tonelada de farelo, mas essa relação já chegou a menos que o dobro em 2001 (Figura 2).

Tabela 1 – Composição do grão de soja, de seus subprodutos, resíduos e estimativa de rendimento e perda em uma unidade esmagadora

Composição relativa (%)	Soja	Óleo	Farelo	Casca	Resíduo	Perda	Lecitina
Umidade	12,5	0,15	12,5	7,25	11	60	0
Óleo	19,5	99,7	2	0,77	0,9	10	24,2
Proteína	36,6	0	47	8,88	20	10	0
Fibras	5,5	0	3,92	34,88	0	10	0
Impurezas	0,2	0	0	0	30	10	0
Fosfato	1,0	0,15	0,8	0,72	0	0	75,8
Carboidratos	24,7	0	33,78	51,78	20	0	0
Produção final	100%	18,13%	70,74%	2,81%	0	7,84%	0,48%

Fonte: Demborguski (2003).

Nos EUA, até a década de 1950, o óleo de soja era o principal produto extraído dos grãos de soja. A torta residual, altamente protéica, era considerada um subproduto de uso restrito. Naquela conjuntura, os programas de melhoramento tinham como objetivo, além do aumento da produtividade, alcançar maiores conteúdos de óleo. Entretanto, este panorama mudou rapidamente durante as décadas de 60 e 70 quando a importância do farelo de soja cresceu de forma significativa, vindo a suplantando a importância econômica do óleo. Este panorama influenciou de forma marcante o desenvolvimento de variedades,

principalmente pelo fato da existência de correlação negativa entre conteúdo de óleo e proteína.

Segundo vários relatos, essa composição média tem pouco variado ao longo dos anos, considerando a maioria das variedades comerciais mais utilizadas. Por exemplo, Wilcox e Guodong (1997) mostram que a composição média das sementes de cultivares plantadas nos EUA, com poucas exceções, permaneceu inalterada, nos últimos 70 anos. De fato, Yaklich *et al.* (2002) demonstraram, pela análise de resultados de testes no período de 1948 a 1998, que as concentrações de óleo e proteína das principais variedades americanas variaram erraticamente ao longo dos anos. Contudo, não atribuíram precisamente um fator de variação, embora a influência de condições climáticas e de cultivo possa ser indicada como uma causa maior que a seleção realizada por meio de programas de melhoramento.

As primeiras variedades de soja cultivadas no Brasil nas décadas de 1960 e início de 1970 foram introduções de variedades americanas ou derivadas de intercruzamentos destas variedades. Sabe-se também que estas variedades foram selecionadas principalmente para produção de grãos e teor de óleo. Pelo fato de estes dois caracteres serem negativamente correlacionados com teor de proteína, este teor se tem mantido em níveis mais baixos (em torno de 40%) do que potencialmente pode ser conseguido (em torno de 50% em materiais não adaptados). As informações disponíveis indicam que a correlação entre o teor de proteína e a produção de grãos não tem sido muito consistente, mas se verifica uma tendência de diminuição ao longo dos anos. Já a correlação entre teor de proteína e teor de óleo se mantém sempre abaixo de -0,50.

Vários estudos, com diferentes métodos de melhoramento e diferentes fontes protéicas, têm sido realizados tentando reduzir a correlação negativa entre produtividade e teor de proteína, a fim de desenvolver cultivares de soja com altos teores de proteína e produtivas (WEHRMANN *et al.* 1987; WILCOX e CAVINS 1995).

Fehr e Weber (1968) avaliaram a eficiência de três ciclos de seleção massal pelo tamanho da semente e densidade específica sobre a porcentagem de proteína e óleo em duas populações heterogêneas de soja. Observaram que o progresso máximo foi obtido para alto teor de proteína e baixo teor de óleo quando se selecionou para sementes maiores e de maior densidade específica. O

contrário, alto teor de óleo e baixo teor de proteína foram observados quando se selecionou para sementes pequenas e de baixa densidade específica.

Brim e Burton (1979) praticaram quatro e seis ciclos de seleção recorrente para teor de proteína em quatro populações de soja. Obtiveram aumentos de 0,30 a 0,70% por ciclo na média do caráter. Mas como a seleção não levou em consideração a produção de grãos, ocorreu sua diminuição em duas populações.

Wehrman *et al.* (1987) utilizaram o método de retrocruzamentos na tentativa de transferir genes que aumentam o teor de protéico de linhagens pouco adaptados para três cultivares de alta produção e de ciclos diferentes. Os autores trabalharam com três populações formadas por cruzamentos biparentais, avançadas até  $F_3$  e retrocruzadas duas vezes com os materiais recorrentes. Correlações genéticas foram calculadas após uma geração de autofecundação no segundo retrocruzamento, e a correlação entre produção de grãos e teor de proteína foi estimada em -0,86, -0,64 e -0,54 para as populações cujos progenitores possuíam ciclo precoce, médio e tardio, respectivamente. Os autores chegaram à conclusão de que dois retrocruzamentos possibilitaram selecionar linhagens com alto teor protéico e produtivas, principalmente naquela população de ciclo mais tardio.

Hartwig (1994) demonstrou que progressos podem ser obtidos na seleção de linhagens de soja de alto teor de proteína e alta produtividade quando se utiliza o progenitor de alto teor de proteína com tipo agrônomico melhorado e com produtividade média. Em seu trabalho, uma população  $F_2$ , com 438 plantas, foi obtida do cruzamento entre a cultivar Forrest (alta produtividade) com uma linhagem D76-8070 (alto teor de proteína e média produtividade). O teor de proteína das plantas  $F_2$  variou de 40,3 a 51,1%. Linhagens com alto teor de proteína e com alta produtividade foram desenvolvidas selecionando 50 plantas com maior teor de proteína em  $F_2$ . Sete linhagens avançadas em  $F_5$  apresentaram teores de proteína similares ao progenitor de alto teor de proteína e produtividade, comparável ao progenitor de alta produtividade. Este autor concluiu neste trabalho que a seleção, em gerações precoces para alto teor de proteína, provavelmente, não acarretará nenhum problema no desenvolvimento de cultivares altamente produtivas e com altos teores de proteína.

Wilcox e Guodong (1997) estudaram a relação entre produtividade e teor de proteína em populações de soja de crescimento determinado e indeterminado.

As progênies de crescimento determinado e indeterminado foram oriundas de cruzamentos entre linhagens de alto teor de proteína de crescimento indeterminado com linhagens de crescimento determinado e médio teor de proteína. As análises demonstraram que as progênies de crescimento determinado constituíram melhores fontes de seleções, as quais combinaram mais alta produtividade e mais alto teor de proteína do que as progênies de crescimento indeterminado.

Wilcox (1998) estudou as mudanças nos teores de proteína e óleo e a variabilidade dessas duas características durante oito ciclos de seleção recorrente em soja. A seleção recorrente aumentou a média do teor de proteína nas sementes em 0,58% e diminuiu a média do teor de óleo em 2,3% por ciclo. Nos últimos ciclos, foi observada uma correlação negativa entre proteína e óleo e diminuição da variabilidade dessas duas características. A maioria dos alelos para alto teor de proteína foi acumulada até o quinto ciclo de seleção recorrente. Uma média de 53% das plantas do sexto até o oitavo ciclo apresentou 48,0% ou mais de proteína. Progênies destas plantas poderiam ser excelentes fontes de fenótipos que combinassem alto teor de proteína com boas características agronômicas.

Helms e Orf (1998) avaliaram o ganho para aumento do teor de proteína por seleção direta e a resposta correlacionada à produtividade e teor de óleo. Seus resultados indicaram que, em média, a seleção para alto teor de proteína resultou em diminuição do teor de óleo e da produtividade.

Cober e Voldeng (2000) avaliaram melhoramento para alto teor de proteína e produtividade em soja por meio de cruzamento simples em um único ciclo de retrocruzamento. Estes autores concluíram que linhagens avançadas oriundas de um único ciclo de retrocruzamento não foram significativamente diferentes das linhagens avançadas obtidas do cruzamento simples quanto à produtividade e teor de proteína. Ambas as estratégias de melhoramento produziram linhagens com teor de proteína significativamente maior que o progenitor recorrente. Contudo, nenhuma dessas linhagens teve produtividade significativamente superior ao progenitor recorrente ou a um cultivar de alto teor de proteína usado como padrão. As populações obtidas exibiram baixas ou nenhuma associação entre produtividade e teor de proteína. Apesar de os progenitores serem fontes úteis de alelos que não exibiram os efeitos pleiotrópicos de baixa produtividade e alto teor de proteína, concluiu-se que o uso

de um ciclo de retrocruzamento para progenitor adaptado não foi eficiente para o desenvolvimento de linhagens com alto teor de proteína e de alta produtividade.

Alt *et al.* (2002) sugerem uma estratégia de seleção baseada na seleção para alto teor de proteína e semente de tamanho grande em populações obtidas de dois e três cruzamentos. Dois progenitores com sementes grandes (LS) e com alto teor de proteína (HP) e um progenitor normal para teor de proteína (N) e tamanho de semente foram usados para obter três tipos de população: LSHP × LSHP; LSHP × N e LSHP × (LSHP × N). A porcentagem das linhagens F<sub>2:3</sub>, com tamanho de semente e proteína maior ou igual ao menor progenitor LSHP, variou de 4% para o cruzamento LSHP × N, de 88% para LSHP × LSHP e de 31,0% para o cruzamento triplo LSHP × (LSHP × N). Estes autores concluem que as populações dos cruzamentos LSHP × LSHP e do cruzamento triplo LSHP × (LSHP × N) seriam mais úteis para desenvolver cultivares com alto teor de proteína e semente de tamanho grande.

Jamago e Nelson (2006) sugeriram, como estratégia para aumentar o teor protéico de progênies derivadas de cruzamentos entre progenitores de elevada produtividade e progenitores de alto teor protéico, a seleção precoce (geração F<sub>2</sub>) para teor de protéico e posteriormente na geração F<sub>4</sub> a seleção para produtividade. Apesar de a linhagem mais produtiva não ter estatisticamente a mesma produtividade de grãos que a variedade padrão Dwight, esta teve a mesma produção de proteína por hectare quando comparada com a Dwight.

#### **2.4. Melhoramento genético de plantas auxiliado por marcadores moleculares**

Com o advento das técnicas de biologia molecular, tornou-se possível a manipulação do DNA, que culminou no surgimento dos vários tipos de marcadores moleculares disponíveis atualmente, destacando-se RFLP (*Restriction Fragment Length Polymorphisms*), RAPD (*Randon Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*), microssatélites ou SSR (*Simple Sequence Repeats*) e mais recentemente os SNPs (*Single Nucleotide Polymorphisms*). Os marcadores moleculares apresentam várias vantagens sobre os marcadores morfológicos por fornecer um número ilimitado de polimorfismos distribuídos aleatoriamente por todo o genoma



e por serem independentes dos efeitos ambientais e do estado fisiológico da planta.

Os tipos de marcadores moleculares mais usados no melhoramento de plantas são o RAPD (*Randomly Amplified Polymorphic DNA*) e microssatélites (*Simple Sequence Repeats – SSR*). Ambos os marcadores, baseados na técnica de reação em cadeia da polimerase (*Polimerase Chain Reactions - PCR*), constituem ferramentas muito úteis e práticas por serem tecnicamente simples e de fácil execução, requererem poucas quantidades de DNA e permitem que dezenas de plantas possam ser analisadas de uma só vez (WAUGH e POWELL, 1992). Recentemente, têm sido utilizados também os marcadores baseados em polimorfismo de base única (SNPs - *Single Nucleotide Polymorphism*) (KWOK, 1996).

Microssatélites consistem de agrupamentos de seqüências curtas repetidas de DNA (2 a 4 pb) em série presentes no genoma de diversos organismos, sendo que o número de seqüências repetidas varia para cada organismo e para cada região do genoma. Os microssatélites são flanqueados por seqüências de DNA extremamente conservadas. Assim, podem-se usar primers que sejam complementares às regiões flanqueando os microssatélites, para se amplificar um determinado loco, via PCR. As variações existentes nas regiões internas flanqueadas pelo *primer* resultam de um variado número de repetições das seqüências microssatélites, podendo ser identificadas por análise eletroforética, por meio de variações nos tamanhos dos fragmentos gerados. Os marcadores microssatélite apresentam como características principais serem abundantes e bem distribuídos no genoma dos organismos; poderem apresentar polimorfismos; serem co-dominantes; identificados via PCR; e serem bastante acessíveis visto as recentes publicações de *primers* de diversos organismos (WEBER e MAY, 1989; SAGHAI MAROOF *et al.*, 1994). Por esse motivo, essencialmente toda e qualquer população segregante pode ser utilizada como população referência para estudos de ligação e mapeamento genético (MORGANTE *et al.*, 1994), além disso, apresentam alto grau de polimorfismo em soja (AKKAYA *et al.*, 1992).

A maioria dos caracteres de importância agrônômica está sob controle genético complexo, envolvendo vários genes. Esses caracteres são denominados poligênicos ou quantitativos, apresentando uma variação contínua do fenótipo, ao invés de classes fenotípicas discretas, o que dificulta a manipulação e

compreensão destes genes, refletindo em dificuldades para se proceder à seleção do fenótipo desejado. Muitos dos problemas num programa de melhoramento baseado na estimação fenotípica de uma característica agrônômica tais como efeitos ambientais ou herança quantitativa podem ser eliminados por diagnósticos baseados em DNA (PATERSON *et al.*, 1988).

Para a maioria dessas características quantitativas, poucas informações existem sobre o número, posição cromossômica, magnitude do efeito e interações dos locos que controlam sua expressão. Estes locos são denominados QTL (“Quantitative Trait Loci”). O mapeamento de QTLs possibilita mensurar o número de locos quantitativos envolvidos na herança complexa, bem como suas localizações cromossômicas, modo de ação gênica (aditividade, dominância, heterose e epistasia), além de possibilitar a decomposição da interação genótipos por ambientes ao nível de cada QTL (FERREIRA e GRATTAPAGLIA, 1998). A capacidade de detecção de um QTL é função da magnitude do seu efeito sobre a característica, do tamanho da população segregante avaliada, da frequência de recombinação entre o marcador e o QTL, bem como da herdabilidade da característica (PATERSON *et al.*, 1991; TANKSLEY, 1993).

O reconhecimento de que marcadores baseados em DNA podem ser extremamente úteis no melhoramento de plantas foi obtido inicialmente com o desenvolvimento de mapeamento genético de plantas baseado em RFLP (BECKMANN e SOLLER, 1983; TANKSLEY *et al.*, 1989), considerando que marcadores que flanqueiam um gene que determina uma característica podem ser usados para seguir a característica em cruzamentos genéticos e em populações segregantes.

#### 2.4.1. Mapeamento genético

Uma das grandes aplicações dos marcadores de DNA é o mapeamento genético, pois ele possibilita a cobertura completa e a análise de genomas, a decomposição de caracteres complexos em seus componentes mendelianos simples e a localização de regiões do genoma responsáveis pelo controle da expressão de caracteres importantes, sejam eles qualitativos ou quantitativos como os QTLs (*Quantitative Trait Loci*) (FERREIRA e GRATTAPAGLIA, 1998.)

Os primeiros mapas genéticos eram fundamentados em marcadores morfológicos e citológicos. Eles se originaram de *estoques cromossômicos* com

aberrações (como aneuploidias, translocações, deleções e inversões), principalmente nas culturas do milho, tomate e ervilha (COE *et al.*, 1988; RICK e YODER, 1988).

Os estudos realizados com marcadores morfológicos em muito contribuíram para a elucidação do fundamento teórico da análise de ligação gênica e para a construção das primeiras versões de mapas genéticos (KNAPP, 1991). Atualmente, têm sido úteis nos casos em que os grupos de ligação já estão associados a determinados cromossomos, podendo ser também utilizados para ancorar grupos de ligação construídos com base em marcadores moleculares a cromossomos específicos.

Para o mapeamento genético, diferentes tipos de populações segregantes podem ser empregados. Tradicionalmente, são utilizadas populações derivadas do cruzamento de linhas puras, o que origina uma geração  $F_1$ , a qual é autofecundada ou retrocruzada com um dos pais para a produção de uma geração  $F_2$  ou  $RC_1$ , respectivamente (FERREIRA e GRATTAPAGLIA, 1998). Alternativamente, são utilizadas populações  $F_n$  ( $n = 3, 4, \dots, \infty$ ), duplo-haplóides (DH), linhagens endogâmicas recombinantes (RILs) (BURR *et al.*, 1988) e linhagens quase isogênicas (NILs) (LYNCH e WALSH, 1998). A escolha da população deve levar em conta os objetivos do pesquisador, além do tempo disponível para a execução do trabalho e dos recursos disponíveis. Nas plantas  $F_1$ , o desequilíbrio de ligação é máximo, e os estudos das populações derivadas a partir destas plantas  $F_1$  procuram explorar este desequilíbrio (SCHUSTER e CRUZ, 2004).

Segundo Guimarães e Moreira (1999), o número de marcadores moleculares necessários para construção de um mapa genético depende do tamanho do genoma, do número de cromossomos e da frequência de recombinação genética. Pode-se dizer que um mapa é considerado completo quando o número de grupos de ligação obtidos pela análise dos marcadores for igual ao número de cromossomos gaméticos do organismo e quando todos os marcadores genéticos mapeados estiverem ligados, indicando que todas as regiões do genoma estão representadas.

A partir da década de 80, com o advento dos marcadores de DNA, o mapeamento genético tornou-se, efetivamente, ilimitado a todas as espécies (BOSTEIN *et al.*, 1980). Mapas genéticos que levaram décadas para serem

desenvolvidos em virtude da restrita disponibilidade de marcadores morfológicos e isoenzimáticos foram amplamente saturados com o uso de marcas moleculares.

A construção do primeiro mapa genético para a soja usando marcadores moleculares foi realizada por Apuya *et al.* (1988). Esses autores analisaram 27 marcadores RFLP em uma progênie de 60 plantas  $F_2$  derivadas de um cruzamento entre as variedades 'Minsoy' e 'Noir', identificando 11 marcadores contidos em quatro grupos de ligação.

Keim *et al.* (1990) mapearam 150 marcadores RFLP numa população de 60 indivíduos  $F_2$  de soja, identificando 25 grupos de ligação, num total de 1200 cM. Esse mapa foi expandido por Diers *et al.* (1992) para 252 marcadores RFLP distribuídos em 31 grupos de ligação, perfazendo um total de 2147 cM. Posteriormente este mapa foi aumentado por Shoemaker *et al.* (1994) para 359 marcadores RFLP, que identificaram 38 QTLs e 10 locos para características monogênicas. Keim *et al.* (1997) obtiveram um mapa com 840 marcadores, sendo 165 RFLP, 25 RAPD e 650 AFLP, distribuídos em 28 grupos de ligação, num total de 3441 cM, partindo-se de uma população inicial de 300 linhagens recombinantes endogâmicas (RILs).

CREGAN *et al.* (1999) construíram um mapa integrado para a soja utilizando dados de três mapas diferentes. Esse mapa integrado foi definido utilizando-se um total de 606 *primers* SSR e outros marcadores (RFLP, RAPD, AFLP, isoenzimas e marcadores morfológicos), totalizando 1.423 marcadores e estabelecendo 20 grupos de ligação consenso. Um mapa mais completo foi produzido por Song *et al.* (2004) combinando cinco populações de mapeamento. Neste mapa integrado ao longo dos 20 grupos de ligação, foram posicionados 1849 marcadores, incluindo 1.015 SSRs, 709 RFLPs, 73 RAPDs, 24 contendo marcadores morfológicos, 6 AFLPs, 10 isoenzimáticos, e 12 outros, abrangendo um total de 2523,6 cM.

#### 2.4.2. Identificação de QTLs

Diversos métodos têm sido desenvolvidos para a localização de QTLs: a) métodos mais simples que não se baseiam em mapas genéticos, tais como análise por marcas simples (STUBER *et al.*, 1987) que tanto podem utilizar populações segregantes como BSA (Bulked Segregant Analysis) e b) métodos mais complexos (do ponto de vista estatístico) que se baseiam em intervalos

obtidos de mapas genéticos, tais como mapeamento por intervalo simples (“interval mapping”) (LANDER e BOTSTEIN, 1989), mapeamento por intervalo composto (“composite interval mapping”) (JANSEN, 1992, 1993) e (ZENG, 1993), mapeamento para múltiplas características ou para diferentes ambientes (JIANG e ZENG, 1995) e o mapeamento por intervalo múltiplo (KAO *et al.*, 1999).

A primeira metodologia desenvolvida foi o mapeamento por marca simples (STUBER *et al.*, 1987) que consiste na associação da expressão do QTL à presença de um marcador, sendo realizadas análises para cada marcador separadamente. Este método representa a idéia básica do mapeamento de QTLs lançada por Sax (1923) e em última análise foi o procedimento adotado desde as primeiras tentativas de associar um caráter quantitativo a um marcador genético de qualquer natureza. A análise é feita verificando-se a diferença entre as médias fenotípicas do caráter para cada uma das classes genotípicas de um dado marcador, que devem possuir distribuição de frequência correspondente ao tipo de população utilizada. Se as diferenças entre as médias fenotípicas das classes forem estatisticamente significativas, pode-se inferir que existe um QTL (ou mais de um) ligado à marca sob análise. Vários procedimentos estatísticos podem ser utilizados, como o teste ‘t’, ANOVA, regressão linear simples ou método da máxima verossimilhança (BEARZOTI, 2000; SHUSTER e CRUZ, 2004). O método é simples, porém limitado. Como cada marcador é analisado individualmente, não é possível identificar se uma marca está ligada a um ou mais QTLs simultaneamente, e assim determinar o número de QTLs envolvidos na característica. Além disso, o método confunde efeito e posição de QTLs, não havendo distinção entre a ocorrência de um QTL de pequeno efeito situado próximo ao marcador e de um QTL de grande efeito só que distante do marcador. Também, os efeitos genéticos do QTL são subestimados, pois seus estimadores são viesados pela fração de recombinação entre o marcador e o QTL. Esse viés, por sua vez, gera a necessidade de utilizar amostras de grandes tamanhos na análise. Entretanto, se a abordagem for realizada pelo método da máxima verossimilhança, é possível a estimação dos parâmetros relativos aos efeitos e à distância entre o QTL e o marcador, uma vez que os estimadores de máxima verossimilhança são não viesados (BEARZOTI, 2000). Esses dois problemas de posição e de efeito acima citados podem ser minimizados, segundo Tanksley (1993), quando um grande número de marcadores segregantes são utilizados cobrindo todo o genoma (usualmente em intervalos menores do que 15 cM). Sob

estas condições, qualquer QTL potencial estaria ligado fortemente a pelo menos um marcador genético.

Em virtude dessas deficiências, um novo método foi proposto por Lander e Botstein (1989) que foi o mapeamento por intervalo, e se baseia na estimação da máxima verossimilhança dos parâmetros: frequência de recombinação, ordem dos marcadores e magnitude do efeito. Neste método, ao invés de marcas individuais, são consideradas pares de marcas adjacentes. A detecção da presença e a estimação dos efeitos dos QTLs são realizados dentro de cada intervalo entre marcas, analisando-se cada intervalo separadamente, utilizando-se o método da máxima verossimilhança.

Seja, por exemplo, um mapa de ligação construído utilizando-se um conjunto de  $n$  marcadores,  $m_1, m_2, \dots, m_n$ . Como princípio, o método assume a existência de um QTL  $Q$  no intervalo entre as marcas adjacentes  $[m_t, m_{t+1}]$ , sendo realizado o Teste da Razão de Verossimilhança (LR) para a presença do QTL em cada posição ao longo do intervalo  $[m_t, m_{t+1}]$ , obtendo-se as estimativas de verossimilhança tanto para os efeitos do possível QTL quanto para a sua posição. Em seguida, o procedimento é repetido no intervalo seguinte  $[m_{t+1}, m_{t+2}]$  e, assim, sucessivamente, até que todo o genoma coberto pelos marcadores seja mapeado. Uma grande vantagem do mapeamento por intervalo é a representação mais clara da presença de QTLs ao longo do genoma, mostrando, para cada posição, o valor do LR correspondente, sendo apresentado mais facilmente o indício da presença de um ou mais QTLs controlando o caráter. As estimativas dos efeitos genéticos são não-viesadas pela utilização do método da máxima verossimilhança. Outra vantagem é o requerimento de um menor número de progênies para a análise. Ainda é possível o estabelecimento de intervalos de confiança para as posições estimadas dos QTLs, utilizando-se o valor do LR expresso na forma de “LOD score”, de acordo com o critério denominado “one LOD support interval” (LANDER e BOTSTEIN, 1989; LIU, 1998; LYNCH e WALSH, 1998). Uma desvantagem atribuída ao mapeamento fundamentado em intervalo é a de que os outros QTLs fora do intervalo são ignorados no mesmo cromossomo, o que pode resultar em duas consequências principais. Toda a variação genética devida a esses outros QTLs é residual, o que pode diminuir a precisão das estimativas e o poder dos testes, não possibilitando com isso a estimação precisa do número e da posição dos QTLs. Além disso, eventuais QTLs que estejam ligados ao intervalo em questão interferem no processo de

estimação, levando, em casos extremos, a declarar a existência de um QTL no intervalo quando na realidade não há nenhum, denominados QTLs fantasmas (BEARZOTI, 2000).

Com o objetivo de simplificar o mapeamento por intervalo e diminuir a demanda computacional, Harley e Knott (1992) propuseram um enfoque baseado em modelos de regressão (simples e múltiplas). A regressão do valor observado da característica (o fenótipo) a um conjunto de possíveis genótipos para QTLs também possibilita estimar a ocorrência do QTL em cada posição do intervalo entre dois marcadores. Nesta regressão, cada possível genótipo para o QTL possui uma probabilidade de ocorrência em função de sua localização no intervalo e dos genótipos dos marcadores franqueadores. A diferença é que este utiliza apenas a média dos marcadores, e o método da máxima verossimilhança utiliza a informação completa da distribuição da associação marcador característica, sendo por isso considerado mais robusto.

A fim de resolver algumas limitações impostas pelo mapeamento por intervalo simples, Jansen (1993) e Zeng (1993, 1994) propuseram o mapeamento por intervalo composto, o princípio do método é o controle de QTLs fora dos intervalos que estão sendo mapeados. Desse modo, é possível identificar a posição e obter as estimativas dos parâmetros genéticos relativos a um dado QTL, sem o viés provocado pela influência de outros QTLs, ligados ou não. Esse controle é garantido pelo modelo adotado, que se baseia na regressão múltipla entre a observação fenotípica do caráter e o conjunto de marcadores, possuindo como uma de suas propriedades o fato de que o efeito eventualmente detectado no mapeamento em um intervalo depende tão somente dos efeitos nele presentes. Como vários marcadores são incluídos no modelo além dos flanqueadores do intervalo em análise, há um maior controle da variância genética residual, o que aumenta o poder do teste estatístico e a precisão do mapeamento. Os marcadores utilizados para essa finalidade são denominados de co-fatores. Além disso, permite o uso de dados obtidos em vários ambientes, para avaliar os efeitos da interação QTLs  $\times$  Ambientes.

A partir daí, a presença de um possível QTL, em um intervalo específico, pode ser determinada por testes de permutação como relatado por (DOERGE e CHURCHILL, 1996). O teste é realizado misturando-se os valores da característica quantitativa em questão e repetindo-se a análise. Tal procedimento é realizado até mil vezes, encontrando-se valores de LOD ou razão de

verossimilhança no mesmo número das repetições realizadas. Adota-se agora um determinado “ponto de corte” que é o valor mínimo de significância para que a probabilidade da presença de um QTL seja considerada em um intervalo determinado.

Kao *et al.* (1999) trouxeram considerável aprimoramento ao mapeamento de QTLs ao desenvolver o método de mapeamento por intervalo múltiplo, pois este incorporou parâmetros de epistasia ao modelo aditivo-dominante. Na presença de epistasia, esse método apresenta algumas vantagens potenciais como maior eficiência e precisão na identificação de QTLs, pois seus efeitos são estimados sem viés e, ao contrário de outros modelos, pode aumentar a eficiência da seleção assistida, considerando os efeitos epistáticos (BEARZOTI, 2000).

#### 2.4.3. Mapeamento de QTLs em soja

Estudos de mapeamento genético molecular, de características de qualidade em soja, têm resultado na obtenção de mapas genéticos detalhados (LARK *et al.*, 1993; SHOEMAKER *et al.*, 1992), inicialmente com a identificação de QTLs para um variado número de características (KEIM *et al.*, 1990; DIERS *et al.*, 1992; MANSUR *et al.*, 1993).

Na literatura constam inúmeros trabalhos envolvendo a identificação de QTLs ligados ao teor de óleo e proteína, utilizando diferentes tipos de marcadores moleculares, origens (cruzamentos entre progenitores adaptados ou não adaptados), tamanhos e tipos de populações (RILs, NILs, F<sub>2</sub>, F<sub>3</sub>, F<sub>n</sub> e de retrocruzamentos). Um dos primeiros trabalhos que detectaram dois QTLs nos GL I e GL E controlando a concentração de proteína foi desenvolvido por Diers *et al.* (1992) utilizando RFLP avaliado em uma população 60 progênies F<sub>2:3</sub> obtida a partir do cruzamento entre um genótipo não adaptado PI 468916 (*Glycine soja*) com alto teor de proteína e uma variedade comercial americana. Os aumentos proporcionados por estes QTLs foram de 1,7% e 2,4% para o GL I e para o GL E, respectivamente. Dando continuidade a esse trabalho, Sebolt *et al.* (2000) testaram os efeitos dos alelos destes dois QTLs provenientes do acesso de *Glycine soja* que aumentaram o teor protéico retrocruzando-o com o progenitor recorrente A81-356022. Uma população RC<sub>3</sub> foi obtida e avaliada para desempenho agrônômico e conteúdo de proteína e óleo em dois locais e em dois



anos. Os autores confirmaram o efeito do QTL do GL-I para teor de proteína, no entanto, não identificaram efeito significativo para o QTL no grupo GL-E, como informado por Diers *et al.* (1992). Entre os marcadores do GL-I testados, o QTL associado ao Satt127 obteve o maior valor de  $R^2$  para teor de proteína explicando uma variação de 3,0% e 1,2% nos anos de 1996 e 1997, respectivamente.

Até 2004, Hyten *et al.* (2004) afirmaram que haviam sido detectadas no mínimo 53 QTLs associados ao teor de óleo, 61 ao teor de proteína e 66 ao tamanho da semente, entretanto muitos desses QTLs não haviam sido confirmados.

Soares (2004) analisou uma população de 118 linhagens recombinantes endogâmicas cultivadas em dois locais (Viçosa-MG e São Gotardo-MG), utilizando marcadores SSRs, RAPDs e AFLPs. Foram obtidos 29 grupos de ligação pouco saturados, contendo 95 marcadores e mais 79 marcas não-ligadas. Foram encontrados 13 marcadores associados ao conteúdo de proteína total, que explicavam de 4,4 a 13,3% da variação total do teor de proteína. Também foram identificados 19 marcadores associados ao conteúdo de 7S e 16 marcadores associados ao conteúdo de 11S, explicando de 4,1 a 15,6% e de 4,2 a 9,7% da característica, respectivamente.

Panthee *et al.* (2005) realizaram um estudo de mapeamento de QTL para teor protéico em soja envolvendo uma população de 101 RILs em uma população de 101 RILs obtidas a partir do cruzamento entre N98-984 × TN93-99 e avaliadas em um único local na geração  $F_6$  e em experimento em blocos casualizados com três repetições. Após testar mais de 585 marcadores SSR, sendo 94 polimórficos, encontraram um QTL no grupo de ligação G que explicava cerca de 20% da característica teor de proteína e três QTLs nos grupos de ligação D1b, G e O que explicavam 11,8; 9,4 e 15% do teor de óleo respectivamente. Esse trabalho anterior foi ampliado por Panthee *et al.* (2007), que avaliaram as características agrônômicas em seis ambientes (três localidades e por dois anos). Três QTLs próximos aos marcadores Satt042 (GL-A1;  $R^2 = 14,6\%$ ), Satt412 (GL-G;  $R^2 = 13,3\%$ ) e Satt076 (GL-G;  $R^2 = 13,8\%$ ) foram associados com produtividade nesta população, e todos esses alelos foram contribuição do progenitor TN93-99. No entanto, apenas o QTL próximo ao Satt076 apresentou significância em todos ambientes. Juntos, esses três QTLs explicaram aproximadamente 41% da variação total para produção de sementes nesta população.

As populações de NILs (*Near Isogenic Lines* – linhagens quase isogênicas), também chamadas de BILs (*Backcross Inbred Lines*) são linhagens que contêm uma ou mais pequenas regiões de DNA de um progenitor doador (regiões que possuem os genes de interesse). Estas isolinhas são obtidas por sucessivas gerações de retrocruzamentos que se realizam concomitantemente com a seleção fenotípica ou genotípica via marcador molecular ligado ao QTL para a característica de interesse. Após o último retrocruzamento, a planta  $RC_nF_1$  é autofecundada por quantas gerações se desejar, podendo ter um comportamento de progênie F2 até como de uma RIL, só que segregando apenas para a característica para a qual ela foi selecionada. Populações de NILs são muito empregadas em mapeamento de alta resolução quando se objetivam a identificação e a clonagem de um gene candidato ou um QTL, sendo empregadas de cinco a sete gerações de retrocruzamentos (LYNCH e WALSH, 1998), mas não necessariamente podendo ser com um só retrocruzamento. Tipicamente, os estudos envolvendo NILs são realizados para avaliar os efeitos de um ou poucos QTLs, pois tão somente uma pequena porção do genoma do doador está representada entre as NILs desenvolvidas.

Kaeppler (1997) demonstrou que testes estatísticos para detectar QTL baseados em NILs possuem um poder inferior à detecção quando realizadas em RILs. Isto acontece principalmente quando cada linha da NIL é comparada individualmente com o progenitor recorrente, caso todas as linhas fossem comparadas simultaneamente o poder de detecção aumentaria. O autor afirma ainda que, embora o poder de detecção das RILs seja maior do que em NILs, estas apresentam a vantagem de oferecer uma maior acurácia na estimação dos efeitos dos QTLs detectados, pois nas RILs múltiplos QTLs estarão segregando.

Chung *et al.* (2003) mapearam um QTL entre os marcadores Satt496 e Satt239 para teor de proteína no grupo de ligação I na mesma região reportada por Sebolt *et al.* (2000) quando utilizaram uma população composta de 76 RILs que se encontrava na geração F5 a partir do cruzamento entre a PI437088A (um acesso de *G.max* com alto conteúdo de proteína, cerca de 48%) e a variedade ASGROW A 3733 (42%). O QTL encontrado explicou 45.5% da variação do teor de proteína no ano de 1996 e 27,7% em no ano de 1997. Eles observaram um aumento de 1,8% na concentração de proteína entre as linhagens homozigotas que continham o alelo da PI437088A em relação às outras.

Wang *et al.* (2004) desenvolveram um estudo de mapeamento de QTLs utilizando populações derivadas de retrocruzamentos entre *Glycine soja* (PI 468916) e *Glycine max* (cultivar IA2008), sendo a cultivar comercial o progenitor recorrente, objetivando encontrar QTLs oriundos da PI 468916 que aumentassem a produção. Destes retrocruzamentos, foram obtidas cinco populações constituídas por isolinhas RC<sub>2</sub>F<sub>4</sub>, sendo duas com 110 isolinhas e as outras três com 79,57 e 112 isolinhas, que foram avaliadas em dois locais por dois anos e genotipadas com marcadores SSR. Quatro QTLs para rendimento foram identificados nos grupos de ligação C2, E, K e M. O QTL no grupo de ligação C2 foi consistente em uma das populações para três ambientes. Esse QTL explicou 40% da variância fenotípica para produtividade nesta população. Em outra população, foram mapeados QTLs nos grupos de ligação E e M, que explicaram 28 e 29% da variação da característica produtividade, respectivamente. O QTL do grupo de ligação E foi significativo em três dos quatro ambientes, e o QTL do grupo de ligação M foi confirmado em todos os ambientes. O QTL do grupo de ligação K foi identificado em outras duas populações e foi significativo em todos os ambientes testados. Esse QTL, localizado no intervalo entre os marcadores Satt137 e Satt178, explicou em uma população 40% e em outra 38% da variação para produtividade. Todos os alelos dos marcadores associados aos QTLs que aumentam a característica produtividade foram provenientes da cultivar comercial IA2008.

Nichols *et al.* (2006), avaliando dois grupos de populações de NILs, um composto por três populações de isolinhas RC<sub>4</sub>F<sub>3,4</sub> derivadas da população RC3 desenvolvida por Sebolt *et al.* 2000 e o outro composto de quatro populações de isolinhas RC<sub>5</sub>F<sub>5</sub>, realizaram um estudo de mapeamento de QTL com marcadores SSR e AFLP direcionado para o grupo de ligação I. Encontraram nas populações do primeiro grupo QTLs para produtividade, concentração de proteína e óleo, maturidade e tamanho da semente. Esses QTLs foram mapeados no intervalo de 11cM entre as marcas Satt614 e Satt354. Os dados das populações do segundo grupo redefiniram a posição do QTL para proteína e óleo no intervalo de três cM entre os marcadores Satt239 e ACG9b, sendo que o efeito desse QTL explicou 15,4; 11,7; 17,7 e 21,1% da variação do teor de proteína nas quatro populações.

Guzman *et al.* (2007) desenvolveram três populações derivadas de retrocruzamentos para identificar QTLs relacionados com produtividade em soja. Na primeira população, o progenitor doador foi a PI 68658 e o progenitor

recorrente foi a variedade Lawrence, sendo obtidas 94 linhas RC<sub>3</sub>F<sub>2</sub> testadas em dois anos (2003 e 2004) em vários locais usando RC<sub>3</sub>F<sub>2</sub> e RC<sub>3</sub>F<sub>3</sub>. A segunda população teve como progenitor doador a PI 407720 e como progenitor recorrente a variedade Beeson, nestas populações foram obtidas 64 linhas RC<sub>2</sub>F<sub>5</sub> testadas nos mesmos anos usando RC<sub>2</sub>F<sub>5</sub> e RC<sub>2</sub>F<sub>6</sub>. A terceira população teve como progenitor doador a PI 391583 e como progenitor recorrente a variedade Kenwood, tendo sido obtidas 75 linhas RC<sub>1</sub>F<sub>5</sub> testadas nos mesmos anos usando RC<sub>1</sub>F<sub>5</sub> e RC<sub>1</sub>F<sub>6</sub>. Foram inicialmente testados 602 marcadores SSR para identificar polimorfismos entre os progenitores e depois foram avaliadas em cada população utilizando-se 30 na população derivada de Lawrence, 45 na população derivada de Beeson e 80 na população derivada de Kenwood. Foram encontrados 13 possíveis QTLs para produtividade nas três populações, alguns destes QTLs foram comuns em anos diferentes. A variância fenotípica para produtividade explicada por QTLs individuais na população derivada de Beeson variou de 10 a 51%, na população derivada de Kenwood variou de 5 a 23% e na população derivada de Lawrence variou de 8 a 26%, aproximadamente. Coletivamente três QTLs explicaram 53% na população derivada de Beeson, cinco QTLs explicaram 58% na população derivada de Kenwood e três QTLs explicaram 30% na população derivada de Lawrence.

#### **2.4.4. Seleção assistida por marcadores moleculares**

A seleção de indivíduos com características de interesse, assistida por marcadores moleculares, pode ajudar na obtenção de variedades com uma ou mais características em um menor espaço de tempo, otimizando o programa de melhoramento (BRUMMER *et al.*, 1997). Além disso, o melhor entendimento das bases genéticas de características como teor de proteína e óleo, pode ajudar na elucidação de estratégias de melhoramento, aumentando a possibilidade de aumento simultâneo destas características durante o processo de seleção até a obtenção de uma variedade.

Na seleção assistida, marcadores moleculares permitem uma avaliação de um maior número de linhagens, aumentam a eficiência de seleção, são não destrutíveis (a análise é feita a partir do DNA extraído de uma folha ou de parte da semente) e permitem a análise tanto de caracteres qualitativos quanto quantitativos (PATERSON *et al.*, 1988).

Como já mencionado anteriormente, esses marcadores de DNA possuem vantagens: são baseados na variação de seqüências de nucleotídeos e numericamente ilimitados; quando co-dominantes, permitem a identificação de locos em qualquer tipo de cruzamento; são fenotipicamente neutros; e não possuem interação epistática.

Potencialmente, o maior impacto desta tecnologia encontra-se realmente para a seleção indireta de genes mapeados determinantes de características quantitativas. Entretanto, a detecção de QTLs para tais características não é tarefa fácil, demandando bastante investimento em trabalhos prévios, para mapeamento e posicionamento desses genes.

Estudos realizados por Dudley (1993) mostram grande vantagem na utilização de marcadores moleculares para mapeamento e identificação de QTLs associados às várias características quantitativas de interesse econômico.

A obtenção de marcadores ligados a QTLs de interesse é um fator primordial em um projeto em que se busca a posterior seleção e a transferência de genes de características quantitativas, qualquer que seja a espécie submetida a um programa de melhoramento. Distâncias de mapa entre marcadores e genes de interesse inferiores a 10 cM são consideradas razoáveis para que esta marca seja utilizada como ferramenta de seleção para monitoramento da transferência do alelo marcado em um posterior cruzamento. A eficiência da seleção assistida por marcadores moleculares de DNA pode ser aumentada pelo uso simultâneo de marcadores em acoplamento e em repulsão, fortemente ligados e flanqueando o loco de interesse. O marcador em repulsão deverá estar presente nos indivíduos recessivos ou heterozigotos, e os marcadores em acoplamento estarão presentes nos indivíduos homozigotos dominantes ou heterozigotos (HALEY *et al.*, 1994).

A SAM é mais efetiva em gerações iniciais entre progênies de cruzamentos derivados de linhas endogâmicas (STROMBERG *et al.*, 1994), pois nestas gerações, a herdabilidade é baixa por causa do limitado número de repetições e pelo fato de o desequilíbrio de ligação ser maior do que em gerações mais avançadas.

Tanto a seleção individual como a de famílias podem ser realizadas com o emprego da teoria do índice de seleção, utilizando marcadores que maximizam a taxa de ganhos genéticos pela combinação das informações do polimorfismo nos locos dos marcadores genéticos com dados da variação fenotípica entre

indivíduos, pois o uso da SAM permite a identificação de combinações favoráveis de alelos, permitindo assim maiores ganhos com a seleção.

Em soja, a seleção assistida por marcadores (SAM) é principalmente utilizada para identificar genótipos resistentes à Podridão Vermelha das Raízes de Soja (PVR) causada pelo fungo *Fusarium solani* e ao Nematóide de Cisto da Soja (NCS). Cervigni (1999) selecionou plantas resistentes homozigotas à raça 3 do nematóide de cisto da soja (NCS), com microssatélites (SSR), obtendo uma eficiência de seleção (ES) de 92%. Esta ES é similar à informada por Concibido *et al.* (1996), de 83-90%, usando marcadores RFLP. Mudge *et al.* (1997) reportaram uma acuracidade de predição fenotípica de 98% para resistência à raça 3 do NCS usando um par de SSR. Schuster *et al.* (2001) reportaram decréscimo em média de 70% no índice de parasitismo (IP) para a raça 14 do NCS, usando marcadores SSR. Fronza (2003) não obteve sucesso na validação de QTLs associados à resistência a PVR em populações desenvolvidas no Brasil, quando utilizou *primers* microssatélites ligados a PVR desenvolvidos por ABU-TREDEIH *et al.* (1996), CHANG *et al.* (1996), HNETKOVSKY *et al.* (1996), KILO *et al.* (1997), PRABHU *et al.* (1999), IQBAL *et al.* (2001) e NJITI *et al.* (2002). Para características de qualidade em sementes de soja Fasoula *et al.* (2004) validaram diferentes QTLs em duas populações de diferentes cruzamentos para teor de proteína, de óleo e peso de sementes. Para esta última característica, os autores confirmaram os QTLs em apenas uma das populações, sugerindo que os QTLs não confirmados podem ser devidos a falsos positivos na população original de mapeamento e à influencia do ambiente.

De acordo com Silva (2005), a identificação e o uso de QTLs necessitam de boa acurácia dos dados fenotípicos, de grandes populações avaliadas em ambientes distintos, da verificação independente em populações paralelas, para determinar a significância de um QTL e da sua confirmação em vários *backgrounds* genéticos.

Lande e Thompson (1990) modelaram a eficiência da SAM no melhoramento de uma característica única, em comparação aos métodos tradicionais de melhoramento.

Segundo Guimarães *et al.* (2006), para avaliar a eficiência da SAM para características quantitativas, vários autores têm utilizado modelos matemáticos baseados em escores moleculares que, para um dado indivíduo, correspondem à soma dos efeitos aditivos da característica associada aos marcadores. O valor

genético de um indivíduo proposto por Moreau *et al.* (2004) pode ser descrito pelo modelo:

$$y_i = \mu + \sum a_m \theta_{mj} + g_i^* + \varepsilon_i ,$$

Em que  $y_i$  corresponde à performance da parcela  $i$  e  $\mu$  à média do ensaio. O efeito genético ( $g_i$ ) do indivíduo  $j$  na parcela  $i$  é subdividido em um escore molecular ( $\hat{M} = \sum a_m \theta_{mj}$ ) e o valor genético não explicado pelos marcadores ( $g_i^*$ ). No escore molecular,  $\theta_{mj}$  é o valor do genótipo  $j$  no loco marcador  $m$  (1, 0 ou -1 quando  $j$  é homozigoto para o alelo de um parental, heterozigoto ou homozigoto para o alelo do segundo parental no caso de indivíduos derivados do cruzamento entre duas linhagens),  $a_m$  é o efeito aditivo do alelo associado ao marcador  $m$  e  $\varepsilon_i$  é o efeito de parcela.

Existem muitos estudos na literatura de simulação sobre a eficiência esperada em plantas. Mas pouco é relatado sobre resultados experimentais, mesmo em soja, pois as informações sobre programas de melhoramento de empresas privadas, que são hoje, possivelmente, as principais executoras da SAM, não estão disponibilizadas para a comunidade científica.

### **3. OBJETIVOS E METAS**

#### **3.1. Objetivos**

a. Comparar diferentes estratégias de seleção para a predição de ganhos com referência a produção de grãos, teor de proteína e de óleo.

b. Comparar diferentes critérios de seleção por meio de seus ganhos estimados e das progênies selecionadas por meio da metodologia do índice de seleção.

c. Identificar marcadores ligados a QTLs que contribuam para o aumento de teor de proteína em uma população de linhagens isogênicas (NILs).

#### **3.2. Meta**

Obter linhagens de soja produtivas, com alto teor de proteína, níveis adequados de óleo, com ciclo mais precoce do que a variedade Monarca e resistentes ao herbicida glifosato.



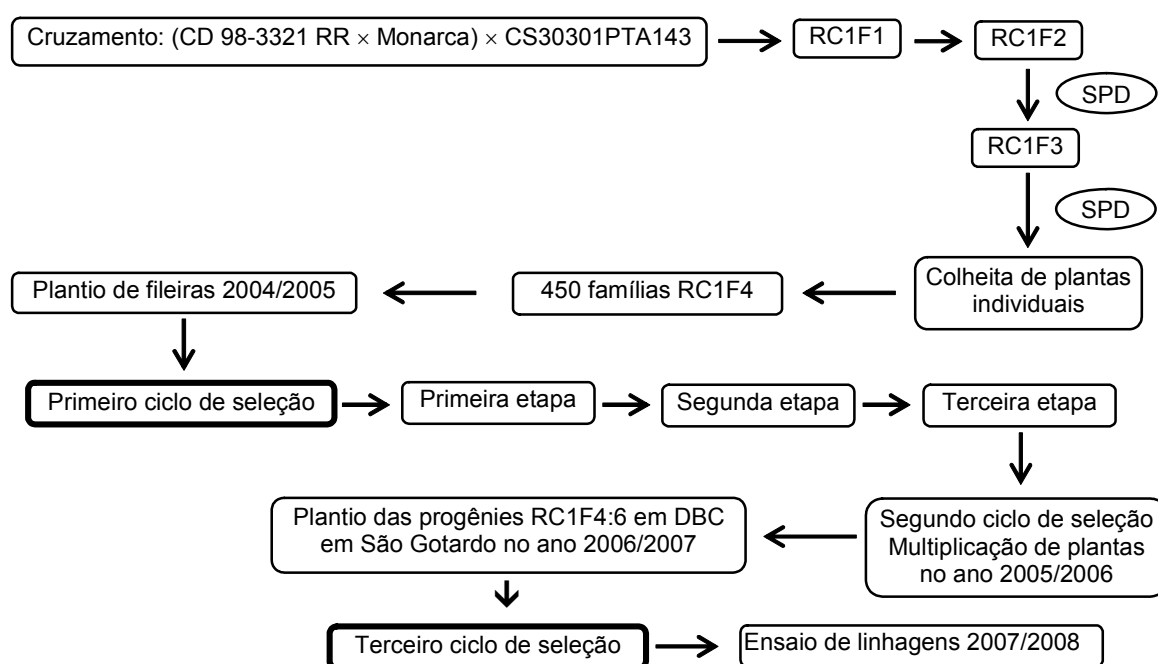
## 4. MATERIAL E MÉTODOS

### 4.1. Material genético e delineamento experimental

Neste trabalho, foram utilizadas duas populações, sendo uma para a seleção e a outra para mapeamento.

#### 4.1.1. População para fins de seleção

O desenvolvimento desta população foi iniciado no ano de 2002 com o cruzamento entre a linhagem CD98-3321RR da Coodetec resistente ao Glifosato e a variedade Monarca da Coopadap. As sementes F1 obtidas foram retrocruzadas com uma isolinha de alto teor protéico (CS3031PTA143) derivada de Monarca, produzindo 30 sementes  $RC_1F_1$ , que foram autofecundadas separadamente, inicialmente em vasos e posteriormente em fileiras até a geração  $RC_1F_{3:4}$ , utilizando-se o método SPD (*Single Pod Descent*) para avançar as gerações. CS3031PTA143 é uma isolinha derivada da variedade recorrente Monarca e da doadora de genes para alto conteúdo de proteína, BARC-8, desenvolvida pelo PMQS. Este trabalho foi dividido em ciclos e etapas de seleção de acordo com o ano agrícola e está resumido na Figura 3.



**Figura 3** - Esquema do cruzamento, condução da população e procedimentos de seleção.

**Primeiro ciclo:** Plantio, avaliação e seleção de famílias e/ou plantas utilizando experimento de linhas com testemunhas (progenitores) intercalares.

No ano de 2004, foram colhidas individualmente cerca de 450 plantas  $RC_1F_{3:4}$  (15 plantas/genótipo  $RC_1F_1$ ). Estas plantas deram origem a uma população composta de 450 famílias  $RC_1F_4$  que foram avaliadas no ano agrícola 2004/2005 no campo experimental da UFV, denominado Horta Nova, localizado em Viçosa, MG. O preparo do solo consistiu de uma aração seguida de três gradagens. A adubação foi feita no sulco de plantio, aplicando-se 90 kg/ha de  $P_2O_5$  e 45 kg/ha de  $K_2O$ . Foram feitas capinas para o controle de plantas daninhas, pulverizações para controle de pragas e doenças, além de irrigação suplementar.

Na avaliação do material experimental, foi empregada uma derivação do delineamento de blocos aumentados de Federer (1955), em que as famílias eram intercaladas com testemunhas, procedimento este chamado de famílias com testemunhas intercaladas por Cruz (2006). Neste sistema, as testemunhas equivalem aos tratamentos comuns, e as famílias aos tratamentos regulares. Como testemunhas foram utilizados dois dos genitores da população que são BARC-8 e Monarca. O experimento foi composto de 15 blocos com 30 linhas de progênies  $RC_1F_4$  e duas linhas com os pais intercalares (BARC-8 e Monarca) em cada bloco. As parcelas constituíram-se de linhas de 2,0 m com 12 plantas por metro e espaçamento entre linhas de 0,5 m. De cada linha foram colhidas e avaliadas oito plantas para produção de grãos em gramas. Das 450 linhas plantadas foram inicialmente descartas 118 por apresentarem desuniformidade de estande, porte baixo ou acamamento.

As seguintes características foram avaliadas:

PROD: Produção de grãos (g/parcela) - peso total das sementes de cada planta ou peso médio da fileira.

PTN: teor de proteínas nos grãos - conteúdo percentual de proteína no grão com base em matéria seca avaliada pelo método Kjeldahl.

OL: teor de óleo nos grãos - conteúdo percentual de óleo no grão com base em matéria seca avaliada utilizando o extrator de Soxhlet.

**Segundo ciclo:** Abertura de linhas selecionadas na segunda etapa do primeiro ciclo.

Esta fase foi conduzida no campo experimental Prof. Diogo Alves de Mello, localizado no *Campus* da UFV em Viçosa, Minas Gerais, no ano agrícola 2005/2006. Esta fase teve como objetivo apenas a multiplicação das plantas selecionadas na geração anterior a fim de aumentar o volume das sementes para posterior avaliação em ensaios com repetições.

**Terceiro ciclo:** Plantio, avaliação e seleção de linhagens RC<sub>1</sub>F<sub>4.6</sub> selecionadas no primeiro ciclo e multiplicadas no segundo.

Esta fase foi conduzida na área experimental da Cooperativa Agropecuária do Alto Paranaíba (Coopadap) localizada no município de Rio Paranaíba, Minas Gerais, no ano agrícola 2006/2007, utilizando as famílias RC<sub>1</sub>F<sub>4.6</sub>. A constituição experimental utilizada foi o delineamento em blocos completos casualizados, com testemunhas adicionais e três repetições. Como testemunhas, foram utilizadas as variedades BARC-8 e Monarca. Cada parcela foi constituída de quatro fileiras de 5,0 m de comprimento, no espaçamento de 0,5 m entre fileiras, sendo utilizadas 15 sementes por metro. Foram consideradas área útil as duas linhas centrais, totalizando 5 m<sup>2</sup>.

As seguintes características foram avaliadas:

NDM - Número de dias para maturação, contados a partir da semeadura até estágio R<sub>8</sub> (quando 95% das vagens atingem a coloração típica de maturação).

APM - Altura da planta na maturação, altura em centímetros do nível do solo até o último nó da haste principal, por ocasião da maturação.

PG - Produtividade de grãos (kg/hectare), avaliada após a colheita, pelo peso dos grãos das parcelas, transformado para kg/ha.

PTN - Teor de proteínas nos grãos, conteúdo percentual de proteína no grão com base na matéria seca avaliada pela metodologia do infravermelho próximo (NIR).

OL - Teor de óleo nos grãos, conteúdo percentual óleo no grão com base em matéria seca avaliada pela metodologia do infravermelho próximo (NIR).

#### **4.1.2. População para fins de mapeamento de QTLs**

Foi utilizada uma população de 168 linhagens quase isogênicas (Near Isogenic Lines – NILs) obtidas por quatro ciclos de retrocruzamento e avançadas

até a geração F<sub>6</sub> pelo método SSD (*Single Seed Descent*) no estudo do mapeamento de QTLs ligados ao teor protéico e na determinação da herdabilidade da característica. Estas linhagens tiveram como progenitores o acesso BARC-8 (alto teor de proteína) e a variedade comercial Monarca. Foram feitos retrocruzamentos e seleções para teor de proteína até o terceiro ciclo, quando foram autofecundadas e selecionadas para teor protéico até a geração RC<sub>3</sub>F<sub>5</sub> e então novamente retrocruzadas com a variedade recorrente Monarca.

Foram obtidas quatro sementes RC<sub>4</sub>F<sub>1</sub> que viriam a formar quatro subpopulações que foram autofecundadas separadamente até a geração RC<sub>4</sub>F<sub>6</sub> pelo método SSD, não sendo mais selecionadas para teor protéico e permitindo que a característica segregasse nesta população semelhante a uma RIL, sendo as plantas da geração RC<sub>4</sub>F<sub>6</sub> colhidas individualmente, para em seguida serem avaliadas no campo. O delineamento experimental utilizado foi o de blocos completos casualizados com duas repetições e com testemunhas adicionais. Como testemunhas, foram utilizadas as variedades BARC-8 e Monarca. Cada parcela foi constituída de uma fileira de 2,0 m de comprimento contendo uma família RC<sub>4</sub>F<sub>6</sub>, no espaçamento de 0,5 m entre fileiras, sendo utilizadas 15 sementes por metro. Os tratos culturais foram os normalmente recomendados para a cultura da soja, sendo utilizada irrigação quando necessária.

#### 4.2. Correção ambiental dentro do experimento

No experimento em que as famílias foram avaliadas em ensaios sem repetições (primeiro ciclo), houve necessidade de fazer a correção dos dados de acordo com os efeitos ambientais que atuam de forma diferenciada sobre as fileiras. Esses efeitos ambientais são estimados a partir das informações dos progenitores, cujas fileiras foram intercaladas no ensaio. De acordo com Cruz (2006), os valores de cada família são corrigidos conforme o que se segue:

##### i) Efeito ambiental

Calculam-se, inicialmente, as médias dos progenitores do ensaio experimental, pelas expressões:

$\bar{Y}_{1j}$ : média do progenitor 1 na repetição (ou linha) j;

$\bar{Y}_{2j}$ : média do progenitor 2 na repetição (ou linha) j; e

$$\bar{Y} = \frac{\sum_{j=1}^r (\bar{Y}_{1j} + \bar{Y}_{2j})}{2r} \text{ que é a média geral dos dois progenitores.}$$

Estimou-se o efeito ambiental, denotado  $\varepsilon_f$ , por meio de:  $\varepsilon_f = \frac{1}{2}(\bar{Y}_{1j} + \bar{Y}_{2j}) - \bar{Y}$

#### ii) Correção dos dados

A correção dos dados para efeito ambiental no experimento foi realizada utilizando a seguinte expressão:

$$V_{Cx} = V_{ox} + \varepsilon_f + \frac{p_f - p_x}{p_f - p_i}(\varepsilon_f - \varepsilon_i)$$

em que:

$V_{Cx}$ : valor corrigido da família na fileira x;

$V_{ox}$ : valor original da família na fileira x;

$\varepsilon_f$ : valor do efeito ambiental que ocorre nos progenitores que se encontram na posição  $p_f$  das fileiras;

$\varepsilon_i$ : valor do efeito ambiental que ocorre nos progenitores que se encontram na posição  $p_i$  das fileiras;

$p_i$ : posição extrema inicial, onde se encontram os progenitores;

$p_f$ : posição extrema final, onde se encontram os progenitores; e

$p_x$ : posição ocupada pela família, cujo valor está sendo corrigido.

A correção dos valores de cada planta dentro de cada família foi realizada multiplicando-se o valor original da característica pelo fator de correção ambiental da família a que pertence a planta. Este fator ambiental foi obtido pela divisão da média da família corrigida pela média da família original.

Além da correção dos valores dos tratamentos não regulares, foram obtidas novas estimativas das herdabilidades e correlações, considerando que a correção proporciona novos valores da variância fenotípica entre as unidades de seleção.

### 4.3. Procedimentos de seleção

#### 4.3.1. Procedimentos de seleção no primeiro ciclo

Os procedimentos de seleção e a predição de ganhos foram divididos em três etapas seqüenciais em função dos objetivos do melhoramento, do tamanho da população e das características avaliadas. A seleção foi feita com base nos dados corrigidos para efeito ambiental.

1ª etapa: Seleção de progênies (famílias e/ou plantas) para máxima produção. População composta de 332 famílias com informação de produção de grãos por planta em nível de família e de indivíduo. Nesta população, as estratégias empregadas e o número de famílias e/ou indivíduos selecionados para predição de ganhos foram:

- i) Seleção entre famílias (205 e 60 famílias);
- ii) Seleção entre e dentro de famílias (60 famílias e 4 plantas por família);
- iii) Seleção massal (240 indivíduos); e
- iv) Seleção combinada (240 indivíduos).

2ª etapa: Seleção entre famílias com ênfase principalmente para teor de proteína, produção de grãos e manutenção do teor de óleo.

População composta de 205 famílias selecionadas na primeira etapa para produção de grãos, com informação de média por família para produção grãos por planta, teor de proteína e óleo. Na seleção das 40 melhores famílias e na obtenção da predição dos ganhos, os critérios de seleção utilizados foram os seguintes:

- i) Seleção direta e indireta;
- ii) Seleção simultânea: índice clássico proposto por Smith (1936) e Hazel (1943) com e sem restrição;
- iii) Seleção simultânea: índice baseado nos ganhos desejados (PESEK e BAKER, 1969) com e sem restrição.

3ª etapa: Seleção de progênies (famílias e/ou plantas) para ganhos tanto para teor de proteína como para produção.

População composta de 43 famílias selecionadas na segunda etapa para teor protéico com informação de produção de grãos por planta e teor de proteína em nível de família e de indivíduo. Nesta população, os critérios de seleção empregados e o número de famílias e/ou indivíduos selecionados para predição de ganhos foram:

*i)* Seleção entre e dentro: direta e indireta. Percentagem de seleção entre igual a 50% e dentro igual a 25%.

*ii)* Seleção simultânea: índice clássico proposto por Smith (1936) e Hazel (1943) com seleção de 40 indivíduos.

#### **4.3.2. Procedimentos de seleção no terceiro ciclo**

Na seleção entre famílias de uma população composta de 117 progênies RC<sub>1</sub>F<sub>4:6</sub> selecionadas na terceira etapa do primeiro ciclo do programa, utilizaram-se informação de média por família para produtividade de grãos (PG), altura de planta na maturação (APM), número de dias para maturação (NDM), teor de óleo (OL) e teor de proteína (PTN). As características consideradas principais foram PG e PTN, os pesos econômicos e os ganhos desejados variaram em função das características e dos seguintes objetivos: a) seleção de genótipos que maximizem ganhos para teor de proteína, mas com menor redução possível de produtividade de grãos e b) seleção de genótipos que maximizem ganhos para produtividade de grãos, mas com menor redução possível para teor protéico.

Para selecionar as 20 melhores famílias e estimar seus ganhos, empregaram-se os seguintes critérios de seleção:

*i)* Seleção direta e indireta;

*ii)* Seleção simultânea: índice clássico proposto por Smith (1936) e Hazel (1943) com e sem restrição; e

*iii)* Seleção simultânea: índice baseado nos ganhos desejados (PESEK e BAKER, 1969) com e sem restrição.

#### **4.4. Determinação do conteúdo de proteínas totais**

A determinação do conteúdo de proteínas das sementes no primeiro ciclo foi realizada utilizando o método Kjeldahl modificado, segundo as normas analíticas do Instituto Adolfo Lutz que requer um mínimo de três gramas (10 sementes) de soja moída. Na fase de digestão, após obtenção do material digerido, adiciona-se peróxido de hidrogênio 30%, levando a mistura ao aquecimento por mais 30 minutos. Na fase de destilação, recolhe-se a amônia liberada em solução de ácido bórico 4%. Obtém-se o teor de nitrogênio pela

titulação da amônia com ácido clorídrico 0,05 N. A partir do teor de nitrogênio, é calculada a porcentagem de proteína total da amostra, empregando-se o fator de 6,25. Os resultados dos teores de proteína foram expressos em porcentagem com base na matéria seca, constituindo a média de duas repetições da análise.

#### 4.5. Determinação do conteúdo de óleo

A determinação do teor de óleo no primeiro ciclo foi realizada em aparelho extrator de Soxhlet, utilizando éter de petróleo como solvente, segundo o procedimento descrito em Normas Analíticas do Instituto Adolfo Lutz (INSTITUTO ADOLFO LUTZ, 1985), com modificações.

#### 4.6. Determinação da composição do grão pela metodologia NIR

A determinação dos teores de proteína e óleo no terceiro ciclo foi realizada simultaneamente em espectrômetro de infravermelho próximo (NIR, equipamento Agrosystem, modelo instalab 600 product analyzer), utilizando sementes moídas e com granulometria adequada para a leitura de absorvância no equipamento NIR.

#### 4.7. Análises estatísticas e genéticas da população de seleção

##### 4.7.1. Análise de variância do experimento com linhas segregantes intercaladas pelos progenitores com informação entre e dentro de fileiras

Inicialmente foi realizada análise de variância para produção de grãos entre e dentro de 332 famílias com objetivo de atender à **Etapa 1**. O procedimento proporcionou resultados em relação aos pais e geração segregante, descritos a seguir:

O modelo estatístico básico adotado foi o seguinte:

$$Y_{ij} = \mu + f_i + e_i + p_{ij} + \delta_{ij}$$

em que:

$Y_{ij}$  : observação relativa à j-ésima planta da i-ésima família;

$\mu$  : média geral da geração (pai ou linha segregante);



$f_i$ : efeito genético atribuído à i-ésima família, sendo  $f_i \sim \text{NID}(0, \sigma_g^2)$ , com  $(i = 1, 2, \dots, f)$ ; para os pais este efeito é inexistente;

$e_i$ : efeito ambiental entre fileiras (de um pai ou de linhas segregantes);

$p_{ij}$ : efeito genético atribuído à j-ésima planta da i-ésima família; sendo  $p_{ij} \sim \text{NID}(0, \sigma_{gd}^2)$ , com  $(i = 1, 2, \dots, f)$  para os pais este efeito é inexistente; e

$\delta_{ij}$ : efeito ambiental entre plantas dentro de fileiras (de um pai ou de linhas segregantes), sendo  $\delta_{ij} \sim \text{NID}(0, \sigma_e^2)$ .

A análise de variância bem como os estimadores destas variâncias relativas às fontes de variação foram obtidos conforme esquema a seguir:

**a) Para famílias**

FV	GL	QM	E(QM)
Entre fileiras	f-1	QME <sub>f</sub>	$\sigma_{Df}^2 + k_1 \sigma_{Ef}^2$
Dentro de fileiras	N-1	QMD <sub>f</sub>	$\sigma_{Df}^2$

Sendo:

$$N = \sum_{i=1}^f n_i$$

$$k_1 = \frac{N - \left( \frac{1}{N} \sum_{i=1}^f n_i^2 \right)}{f - 1}$$

$$\sigma_{Ef}^2 = \sigma_{ge}^2 + \sigma_{ee}^2$$

$$\sigma_{Df}^2 = \sigma_{gd}^2 + \sigma_{ed}^2$$

em que:

$\sigma_{Ef}^2$ : variância fenotípica entre famílias

$\sigma_{Df}^2$ : variância fenotípica dentro de famílias

$\sigma_{gd}^2$ : variância genética entre plantas dentro de família;

$\sigma_{ge}^2$ : variância genética entre famílias;

$\sigma_{ed}^2$ : variância ambiental entre plantas dentro de famílias; e

$\sigma_{ee}^2$ : variância ambiental entre famílias.

**b) Para o Pai 1**

FV	GL	QM	E(QM)
----	----	----	-------

Entre fileiras	$p_1-1$	$QME_{p1}$	$\sigma_{Dp1}^2 + k_2\sigma_{Ep1}^2$
Dentro de fileiras	$N-p_1$	$QMD_{p1}$	$\sigma_{Dp1}^2$

Sendo:

$$N = \sum_{i=1}^{p_1} n_i \quad k_2 = \frac{N - \left( \frac{1}{N} \sum_{i=1}^{p_1} n_i^2 \right)}{p_1 - 1}$$

$$\sigma_{Dp1}^2 = \sigma_{ed}^2$$

$$\sigma_{Ep1}^2 = \sigma_{ee}^2$$

### c) Para o Pai 2

FV	GL	QM	E(QM)
Entre fileiras	$p_2-1$	$QME_{p2}$	$\sigma_{Dp2}^2 + k_3\sigma_{Ep2}^2$
Dentro de fileiras	$N-p_2$	$QMD_{p2}$	$\sigma_{Dp2}^2$

Sendo:

$$N = \sum_{i=1}^{p_2} n_i \quad k_3 = \frac{N - \left( \frac{1}{N} \sum_{i=1}^{p_2} n_i^2 \right)}{p_2 - 1}$$

$$\sigma_{Dp2}^2 = \sigma_{ed}^2$$

$$\sigma_{Ep2}^2 = \sigma_{ee}^2$$

Com a combinação da informação destas classes de material genético com avaliação de plantas individuais, foi possível estimar os seguintes parâmetros relativos à população segregante:

Variância fenotípica entre e dentro de famílias segregantes

$$\hat{\sigma}_{f.entre}^2 = \frac{QME_f - QMD_f}{k_1} \quad \hat{\sigma}_{f.dentro}^2 = QMD_f$$

Variância ambiental entre e dentro de famílias

$$\hat{\sigma}_{\varepsilon.entre}^2 = \frac{1}{2} \left( \hat{\sigma}_{ee(pai1)}^2 + \hat{\sigma}_{ee(pai2)}^2 \right) = \frac{1}{2} \left( \frac{QME_{p1} - QMD_{p1}}{k_2} + \frac{QME_{p2} - QMD_{p2}}{k_3} \right)$$

$$\hat{\sigma}_{\varepsilon.dentro}^2 = \frac{1}{2}(\hat{\sigma}_{ed(pai1)}^2 + \hat{\sigma}_{ed(pai2)}^2) = \frac{1}{2}(QMD_{p1} + QMD_{p2})$$

Variância genotípica entre e dentro de famílias

$$\hat{\sigma}_{g.entre}^2 = \hat{\sigma}_{f.entre}^2 - \hat{\sigma}_{\varepsilon.entre}^2 \quad \hat{\sigma}_{g.dentro}^2 = \hat{\sigma}_{f.dentro}^2 - \hat{\sigma}_{\varepsilon.dentro}^2$$

Herdabilidades no sentido amplo entre e dentro de famílias e total

$$h_{entre}^2 = \frac{\hat{\sigma}_{g.entre}^2}{\hat{\sigma}_{f.entre}^2}; \quad h_{dentro}^2 = \frac{\hat{\sigma}_{g.dentro}^2}{\hat{\sigma}_{f.dentro}^2}; \quad h_{total}^2 = \frac{\hat{\sigma}_{g.total}^2}{\hat{\sigma}_{f.total}^2}$$

A decomposição da variância genética total foi assim obtida:

FV	Nível de endogamia (F)	Geração F <sub>4</sub> (F = 0,75)
Entre fileiras	$2F\sigma_a^2 + F(1-F)\sigma_d^2$	$(3/2)\sigma_a^2 + (3/16)\sigma_d^2$
Dentro de fileiras	$(1-F)\sigma_a^2 + (1-F)\sigma_d^2$	$(1/4)\sigma_a^2 + (1/4)\sigma_d^2$
Total	$(1+F)\sigma_a^2 + (1-F^2)\sigma_d^2$	$(7/4)\sigma_a^2 + (7/16)\sigma_d^2$

Sendo:

$\sigma_a^2$ : variância genética aditiva e

$\sigma_d^2$ : variância genética devido aos desvios de dominância.

Com as equações acima, foram estimadas as seguintes variâncias:

$$\hat{\sigma}_a^2 = \frac{\hat{\sigma}_{ge}^2 - F\hat{\sigma}_{gd}^2}{F(1+F)} \quad \hat{\sigma}_d^2 = \frac{2F\hat{\sigma}_{gd}^2 - (1-F)\hat{\sigma}_{ge}^2}{F(1-F^2)}$$

As herdabilidades restritas entre e dentro de famílias e total foram assim estimadas:

$$h_{R.entre}^2 = \frac{\hat{\sigma}_{a.entre}^2}{\hat{\sigma}_{f.entre}^2} = \frac{2F\hat{\sigma}_a^2}{\hat{\sigma}_{f.entre}^2}$$

$$h_{R.dentro}^2 = \frac{\hat{\sigma}_{a.dentro}^2}{\hat{\sigma}_{f.dentro}^2} = \frac{(1-F)\hat{\sigma}_a^2}{\hat{\sigma}_{f.dentro}^2}$$

$$h_{R.total}^2 = \frac{\hat{\sigma}_{a.total}^2}{\hat{\sigma}_{f.total}^2} = \frac{(1+F)\hat{\sigma}_a^2}{\hat{\sigma}_{f.total}^2}$$

Os coeficientes de variação foram assim estimados:

$$\text{Coeficiente de variação experimental: } CV_e \% = \frac{\left(100\sqrt{\hat{\sigma}_{\varepsilon.\text{entre}}^2 + \hat{\sigma}_{\varepsilon.\text{dentro}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação fenotípico: } CV_f \% = \frac{\left(100\sqrt{\hat{\sigma}_{f.\text{entre}}^2 + \hat{\sigma}_{f.\text{dentro}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação genético: } CV_g \% = \frac{\left(100\sqrt{\hat{\sigma}_{g.\text{entre}}^2 + \hat{\sigma}_{g.\text{dentro}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação genético entre famílias: } CV_{ge} \% = \frac{\left(100\sqrt{\hat{\sigma}_{g.\text{entre}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação genético dentro famílias: } CV_{gd} \% = \frac{\left(100\sqrt{\hat{\sigma}_{g.\text{dentro}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação experimental entre famílias: } CV_{\varepsilon e} \% = \frac{\left(100\sqrt{\hat{\sigma}_{\varepsilon.\text{entre}}^2}\right)}{m_{\text{geral}}}$$

$$\text{Coeficiente de variação experimental dentro famílias: } CV_{\varepsilon d} \% = \frac{\left(100\sqrt{\hat{\sigma}_{\varepsilon.\text{dentro}}^2}\right)}{m_{\text{geral}}}$$

#### 4.7.2. Análise de variância do experimento com linhas segregantes intercaladas pelos progenitores com informação entre médias de fileiras

Foram realizadas análises de variância para produção de grãos por planta, teor de proteína e óleo nas 205 famílias segregantes, utilizando o delineamento com famílias e progenitores intercalares (testemunhas). Esta análise foi realizada para atender aos objetivos da **Etapa 2**. O modelo estatístico adotado será foi seguinte:

$$\text{Para as testemunhas, } Y_{ij} = \mu + T_i + \varepsilon_{ij},$$

em que  $Y_{ij}$  é o valor da característica da  $i$ -ésima testemunha na  $j$ -ésima repetição;  $\mu$  é a média geral das testemunhas;  $T_i$  é o efeito da  $i$ -ésima testemunha ( $i = 1, 2, \dots, t$ );  $\varepsilon_{ij}$  é o erro aleatório que incide sobre as testemunhas  $\varepsilon_{ij} \sim NID(0, \sigma^2)$ .

**Para as linhagens,  $Y_i = \mu + G_i + \varepsilon_i$ ,**

em que  $Y_i$  é o valor da característica da  $i$ -ésima linhagem;  $\mu$  é a média geral das linhagens;  $G_i$  é o efeito genético da  $i$ -ésima linhagem ( $i = 1, 2, \dots, g$ )  $G_i \sim NID(0, \sigma_g^2)$ ; e  $\varepsilon_i$  é o erro aleatório que incide sobre as linhagens,  $\varepsilon_i \sim NID(0, \sigma^2)$ .

A análise de variância foi realizada conforme esquema apresentado na Tabela 2.

Tabela 2 - Esquema da análise de variância, quadrados médios e esperança dos quadrados médios do experimento com linhas segregantes intercaladas pelos progenitores

FV	GL	QM	E(QM)
Famílias	f-1	QMF	$\sigma^2 + \sigma_g^2$
Testemunhas	t-1	QMTe	$\sigma^2 + k\phi_t$
Resíduo	t(r-1)	QMR	$\sigma^2$

Com  $m_t$ : média das testemunhas e  $m_f$ : média das famílias.

Sendo:

$$N = \sum_{i=1}^{te} r_i \qquad k = \frac{N - \left( \frac{1}{N} \sum_{i=1}^{te} r_i \right)}{te - 1}$$

As estimativas dos parâmetros foram assim obtidas:

**Para as testemunhas:**

$$\text{Variância fenotípica média: } \hat{\sigma}_f^2 = \frac{QMT_e}{k}$$

Componente quadrático que expressa a variabilidade genotípica:

$$\hat{\phi}_t = \frac{QMT_e - QMR}{k}$$

Coefficiente de determinação genotípico, baseado na média das testemunhas:

$$H^2 = \frac{\hat{\phi}_t}{\hat{\sigma}_f^2}$$

$$\text{Correlação intraclasse: } \hat{\rho} = \frac{\hat{\phi}_t}{\hat{\phi}_t + \hat{\sigma}^2}$$

$$\text{Coeficiente de variação genético: } CV_g \% = \frac{\left(100\sqrt{\hat{\phi}_t}\right)}{m_t}$$

$$\text{Coeficiente de variação experimental: } CV_e \% = \frac{\left(100\sqrt{\hat{\phi}_t}\right)}{m_t}$$

$$\text{Razão } \frac{CV_g}{CV_e} = \sqrt{\frac{\hat{\phi}_t}{\hat{\sigma}^2}}$$

**Para as famílias:**

$$\text{Variância fenotípica: } \hat{\sigma}_f^2 = QMF$$

$$\text{Variância genotípica: } \hat{\sigma}_g^2 = QMF - QMR$$

$$\text{Herdabilidade: } h^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_f^2}$$

$$\text{Coeficiente de variação genético: } CV_g \% = \frac{\left(100\sqrt{\hat{\sigma}_g^2}\right)}{m_f}$$

$$\text{Coeficiente de variação experimental: } CV_e \% = \frac{\left(100\sqrt{\hat{\sigma}^2}\right)}{mf}$$

$$\text{Razão } \frac{CV_g}{CV_e} = \sqrt{\frac{\hat{\sigma}_g^2}{\hat{\sigma}^2}}$$

$$\text{Coeficiente de variação experimental do ensaio: } CV \% = \frac{\left(100\sqrt{\hat{\sigma}^2}\right)}{m}$$

#### 4.7.3. Análise de variância das linhagens selecionadas e plantadas em 2006/2007

Foi usado o modelo indicado a seguir na análise de variância dos dados obtidos para os cinco caracteres avaliados, feita segundo o delineamento em blocos completos casualizados com testemunhas adicionais (progenitores):

$$Y_{ij} = \mu + T_i + B_j + \varepsilon_{ij},$$

em que:

$Y_{ij}$  é o valor observado da característica no  $i$ -ésimo tratamento (família ou testemunha) no  $j$ -ésimo bloco;  $\mu$  é a média geral;  $T_i$  é o efeito do  $i$ -ésimo tratamento. Para ( $i = 1, 2, \dots, g$ ), tem-se o efeito aleatório ou fixo para os genótipos e para ( $i = g + 1, g + 2, \dots, g + t$ ), tem-se o efeito fixo para as testemunhas;  $B_j$  é o efeito do  $j$ -ésimo bloco ( $j = 1, 2, \dots, r$ ); e  $\varepsilon_{ij}$  é o erro aleatório,  $\varepsilon_{ij} \sim NID(0, \sigma^2)$ .

O esquema da análise de variância, com base neste modelo, está apresentado na Tabela 3.

Tabela 3 - Esquema da análise de variância, quadrados médios e esperança dos quadrados médios do delineamento em blocos completos casualizados com testemunhas adicionais

FV	GL	SQ	QM	E(QM)
				Genótipo aleatório
Blocos	r-1	SQB	QMB	$\sigma^2 + T\sigma_b^2$
Tratamentos	T-1	SQT	QMT	$\sigma^2 + r\phi_T$
Genótipos (G)	g-1	SQG	QMG	$\sigma^2 + r\sigma_g^2$
Testemunhas (Te)	t-1	SQTe	QMTe	$\sigma^2 + r\phi_t$
G versus Te	1	SQGr	QMGr	-
Resíduo	(r-1)(T-1)	SQR	QMR	$\sigma^2$
Total	rT-1	SQTo		

Sendo  $T = g + t$ , o número total de tratamentos, incluindo as famílias e as testemunhas e  $Gr =$  grupo.

As médias e as somas dos quadrados e quadrados médios foram obtidas por meio das seguintes expressões:

$$\text{Média geral} = M = \frac{Y_{..}}{Tr} = \frac{\sum_{i=1}^T \sum_{j=1}^r Y_{ij}}{Tr}; C = \frac{Y_{..}^2}{Tr} = C_g + C_t$$

$$\text{Média dos genótipos} = M_g = \frac{Y_g}{gr} = \frac{\sum_{i=1}^g \sum_{j=1}^r Y_{ij}}{gr}; C_g = \frac{Y_g^2}{gr}$$

$$\text{Média das testemunhas} = M_t = \frac{Y_t}{tr} = \frac{\sum_{i=g+1}^{T+1} \sum_{j=1}^r Y_{ij}}{tr}; C_t = \frac{Y_t^2}{tr}$$

$$SQTo = \sum_{i=1}^T \sum_{j=1}^r Y_{ij}^2 - C; SQB = \frac{1}{T} \sum_{j=1}^r Y_{.j}^2 - C; e QMB = \frac{SQB}{r-1}$$

$$SQT = \frac{1}{r} \sum_{i=1}^T Y_{i.}^2 - C \text{ e } QMT = \frac{SQT}{T-1}$$

$$SQG = \frac{1}{r} \sum_{i=1}^g Y_{i.}^2 - C_g \text{ e } QMG = \frac{SQG}{g-1}$$

$$SQTe = \frac{1}{r} \sum_{i=g+1}^T Y_{i.}^2 - C_t \text{ e } QMTe = \frac{SQTe}{t-1}$$

$$SQGr = SQT - (SQG + SQTe) \text{ e } QMGr = SQGr$$

$$SQR = SQT_0 - (SQB + SQT) \text{ e } QMR = \frac{SQR}{(r-1)(T-1)}$$

Obtiveram-se as estimativas dos parâmetros da seguinte forma:

$$\text{Variância fenotípica média: } \hat{\sigma}_f^2 = \frac{QMG}{r}$$

$$\text{Variância ambiental média: } \hat{\sigma}_e^2 = \frac{QMR}{r}$$

$$\text{Variância genotípica média: } \hat{\sigma}_g^2 = \frac{QMG - QMR}{r}$$

$$\text{Herdabilidade para seleção baseada na média da família: } h^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_f^2}$$

$$\text{Correlação intraclasse: } \hat{\rho} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + QMR}$$

$$\text{Coeficiente de variação genético: } CV_g \% = \frac{(100\sqrt{\hat{\sigma}_g^2})}{M_g}$$

$$\text{Coeficiente de variação ambiental: } CV_e \% = \frac{(100\sqrt{\hat{\sigma}_e^2})}{M_g}$$

$$\text{Razão } \frac{CV_g}{CV_e} = \sqrt{\frac{\hat{\sigma}_g^2}{\hat{\sigma}_e^2}}$$

$$\text{Coeficiente de variação experimental do ensaio: } CV \% = \frac{(100\sqrt{\hat{\sigma}^2})}{M}$$



## 4.8. Esquema de seleção e predição de ganhos por seleção na população de seleção

### 4.8.1. Seleção entre famílias

Este procedimento foi utilizado para estimar ganhos diretos, indiretos e para seleção simultânea de caracteres entre médias de famílias utilizando-se as seguintes expressões:

a) Ganho de seleção direto pela metodologia proposta por Eberthart (1970).

$$GS_i = DS_i \times h_i^2$$

em que:

$GS_i$  = ganho direto predito no caráter  $i$ ;

$h_i^2$  = herdabilidade, em nível de média de família ou progênies, para o caráter  $i$ ;

$DS_i$  = diferencial de seleção do caráter  $i$ , dado por:  $DS_i = \bar{X}_{si} - \bar{X}_{oi}$

$\bar{X}_{si}$  : média das famílias selecionadas; e

$\bar{X}_{oi}$  : média original da população sob seleção.

Em termos percentuais,

$$GS_i \% = \frac{GS_i}{\bar{X}_{oi}} \times 100$$

b) ganho de seleção indireto

$$GS_{j(i)} = DS_{j(i)} \times h_j^2$$

em que:

$GS_{j(i)}$  = ganho indireto predito no caráter  $j$  pela seleção no caráter  $i$ ;

$h_j^2$  = herdabilidade em nível de média de família ou progênies para o caráter  $j$ ; e

$DS_i$  = diferencial de seleção indireto do caráter j, obtido em função da média daqueles indivíduos cuja superioridade foi evidenciada com base no caráter i, sobre a qual se pratica a seleção direta.

Em termos percentuais:

$$GS_{j(i)} \% = \frac{GS_{j(i)}}{\bar{X}_{oj}} \times 100$$

Essa expressão é vantajosa principalmente quando o número de progênes for pequeno e/ou as correlações entre os dados amostrados não refletirem as correlações entre as unidades selecionadas.

#### 4.8.2. Seleção entre e dentro de famílias

O procedimento foi realizado selecionando-se primeiramente as melhores famílias de acordo com as percentagens de seleção desejada para cada característica. Depois de selecionadas as melhores famílias, os melhores indivíduos dentro das famílias selecionadas foram escolhidos de acordo com a percentagem de seleção.

a) O ganho de seleção direto entre foi estimado pela seguinte expressão:

$$GS_{entre} = DS_e \times h^2_{R.entre}$$

Sendo:

$h^2_{R.entre}$  : herdabilidade no sentido restrito entre

$$DS_e = \bar{X}_{fs} - \bar{X}_o ,$$

em que:

$\bar{X}_{fs}$  : média das famílias selecionadas; e

$\bar{X}_{oi}$  : média original da população sob seleção.

b) O ganho de seleção dentro foi estimado pela seguinte expressão:

$$GS_{dentro} = DS_m \times h^2_{R.dentro}$$

sendo:

$h_{R.dentro}^2$  : herdabilidade no sentido restrito dentro; e

$DS_m$  : diferencial de seleção médio dentro.

$$DS_m = \frac{1}{s} \sum_{i=1}^s DS_i = \frac{1}{s} \sum_{i=1}^s \left( \bar{X}_{ind(i)} - \bar{X}_{fs(i)} \right),$$

em que:

s: número de famílias selecionadas;

$\bar{X}_{fs(i)}$  : média da i-ésima família selecionada; e

$\bar{X}_{ind(i)}$  : média dos indivíduos superiores dentro da família i.

c) O ganho de seleção total foi estimado pela seguinte expressão:

$$GS_{total} = GS_{entre} + GS_{dentro}$$

#### 4.8.3. Seleção massal

Neste procedimento, praticou-se a seleção dos melhores indivíduos, independentemente da família a que pertenciam.

O ganho de seleção entre foi estimado pela seguinte expressão:

$$GS_{massal} = DS_m \times h_{R.total}^2$$

em que:

$h_{R.total}^2$  : herdabilidade no sentido restrito total; e

$DS_m$  : diferencial de seleção, considerando os valores fenotípicos das plantas avaliadas no ensaio.

#### 4.8.4. Seleção combinada

A seleção foi realizada a partir de um índice obtido levando-se em consideração o desvio do indivíduo em relação à média da família à qual pertencia e o desvio da média da família em relação à média geral. Para cada desvio, foi dado um peso, objetivando minimizar as diferenças entre os elementos do vetor I, de índices de seleção e os elementos correspondentes do vetor G, de valores genéticos. A expressão para o índice de seleção combinada foi dada por:

$$I_{ij} = b_1(Y_{ij} - \bar{Y}_i) + b_2(\bar{Y}_i - \bar{Y}_{..})$$

em que:

$b_1$ : peso dado ao desvio do valor fenotípico do indivíduo ( $Y_{ij}$ ) em relação à média de sua família ( $\bar{Y}_i$ ); e

$b_2$ : peso dado ao desvio da média fenotípica da família ( $\bar{Y}_i$ ) em relação à média geral da população ( $\bar{Y}_{..}$ ).

Sendo:

$$b_1 = h_{Rdentro}^2 ;$$

$$b_2 = h_{Rentre}^2 .$$

A resposta à seleção é dada pelo diferencial de seleção no índice:

$$RS = DS_i .$$

#### 4.8.5. Seleção simultânea utilizando índices de seleção

Para a seleção das famílias foram utilizados o índice clássico de Smith (1936) e Hazel (1943) e o índice com base nos ganhos desejados de Pesek e Baker (1969) com e sem restrição. Com relação ao índice clássico, proposto por Smith (1936) e Hazel (1943), o índice (I) e o agregado genotípico (H) são assim descritos:

$$I = b_1x_1 + b_2x_2 + \dots + b_nx_n = \sum_{i=1}^n b_i x_i = b'x$$

$$H = a_1g_1 + a_2g_2 + \dots + a_ng_n = \sum_{i=1}^n a_i g_i = a'g$$

em que:

$n$  é o número de caracteres avaliados;  $b'$  é um vetor  $1 \times n$  dos coeficientes lineares a serem estimados;  $x$  é um vetor  $n \times 1$  de médias dos caracteres;  $a'$  é um vetor  $1 \times n$  de pesos econômicos previamente estabelecidos; e  $g$  é um vetor  $n \times 1$  de valores genéticos desconhecidos dos  $n$  caracteres.

Em função das estruturas do índice e do agregado genotípico, pode-se deduzir que:

$$\hat{b} = P^{-1}Ga$$

em que:

$P$  é a matriz de variâncias e co-variâncias fenotípicas; e

$G$  é a matriz de variâncias e co-variâncias genéticas entre os caracteres.

Na aplicação do índice clássico, consideraram-se como pesos econômicos os coeficientes de variação genética de cada caráter, conforme sugestão de Cruz (1990) e a proporcionalidade da composição do grão .

Para construção do índice de Pesek e Baker (1969), deve-se considerar a expressão do ganho desejado para os vários caracteres, definida por

$$\Delta Gd = \frac{G\hat{b}}{\hat{\sigma}_I}$$

em que  $\Delta Gd$  é o vetor de ganhos desejados,  $G$  é a matriz de variâncias e co-variâncias genéticas entre os caracteres,  $\hat{b}$  é o vetor  $n \times 1$  de coeficientes do índice e  $\hat{\sigma}_I$  a raiz quadrada da variância do índice. Como  $\hat{\sigma}_I$  é escalar e não afeta a proporcionalidade do índice, ele pode ser desprezado, obtendo-se:

$$\hat{b} = G^{-1}\Delta Gd$$

Assim, o índice seria estabelecido apenas com a definição do vetor de ganhos desejados e da matriz de variâncias e co-variâncias genotípicas. Apesar de o referido método dispensar o estabelecimento dos pesos econômicos, estes podem ser estimados, posteriormente, pela expressão:

$$\hat{a} = G^{-1}P\hat{b}$$

em que  $\hat{a}$  é o vetor  $n \times 1$  de pesos econômicos estimados e  $P$  é a matriz de variâncias e co-variâncias fenotípicas entre os caracteres.

Para construção do índice de Pesek e Baker (1969) inicialmente considerou-se como ganhos desejados o equivalente a um desvio padrão genético para cada caráter principal. Foram incluídas outras combinações de ganhos desejados e diferentes classificações quanto aos caracteres considerados como principais, visando as melhores distribuições de ganhos de acordo com os propósitos do presente trabalho.

Também foram estabelecidas restrições em relação ao índice clássico de Smith (1936) e Hazel (1943) e de Pesek e Baker (1969), a fim de maximizar os

ganhos nas características principais. Utilizou-se dos índices de Kempthorne e Nordskog (1959) e o de Tai (1977), sendo este uma generalização do índice de Pesek e Baker (1969).

Depois de selecionadas as famílias utilizando-se as características para compor o índice de seleção, à resposta à seleção foi estimada para cada característica isoladamente, utilizando-se procedimento semelhante à seleção individual, ou seja:

$$GS_X = DS_X \times h_{r^x}^2$$

#### 4.8.6. Seleção simultânea de caracteres entre e dentro de famílias

a) O ganho de seleção direto entre e dentro foi estimado conforme item 4.8.1.

b) O ganho de seleção indireto entre e dentro de famílias foi estimado pela seguinte expressão:

$$GS_{j(i)} = \hat{\beta}_{j(i)} GS_i$$

Sendo:

$GS_{j(i)}$  = ganho de seleção indireto predito no caráter j, pela seleção praticada no caráter i;

$GS_i$  = ganho de seleção direto em i;

$\hat{\beta}_{j(i)} = \frac{\hat{Cov}_g(x_i, x_j)}{\hat{\sigma}_{gi}^2}$ : coeficiente de regressão genético;

$\hat{Cov}_g(x_i, x_j)$  = co-variância genética (entre ou dentro) entre os caracteres i e j; e

$\hat{\sigma}_{gi}^2$  = variância genética (entre ou dentro) do caráter principal sobre o qual se pratica a seleção.

c) O ganho de seleção entre e dentro de famílias utilizando o índice clássico foi estimado pelas seguintes expressões:

c.1) Índice para seleção entre famílias dado por:

$$I = b_1 \bar{y}_1 + b_2 \bar{y}_2 + \dots + b_n \bar{y}_n = \sum_{i=1}^n b_i \bar{y}_i = y' b_e$$

$I$ : índice de seleção entre médias de famílias a ser estimado;

$y'$ : vetor (1xn) de médias de famílias; e

$b_e$ : vetor (nx1) dos coeficientes de ponderação do índice estimado por meio de:

$$P_e b_e = G_e a,$$

Sendo:

$P_e$ : matriz de variâncias e covariâncias fenotípicas entre famílias;

$G_e$ : matriz de variâncias e covariâncias genotípicas entre famílias; e

$a$ : vetor de pesos econômicos.

c.2) Índice para seleção dentro de famílias dado por:

$$I = b_1 y_1 + b_2 y_2 + \dots + b_n y_n = \sum_{i=1}^n b_i y_i = y' b_d$$

$I$ : índice de seleção dentro de famílias a ser estimado;

$y'$ : vetor (1 × n) de valores individuais das plantas dentro de famílias; e

$b_d$ : vetor (n × 1) dos coeficientes de ponderação do índice estimado por meio de:

$$P_d b_d = G_d a$$

Sendo:

$P_d$ : matriz de variâncias e covariâncias fenotípicas entre plantas dentro de famílias;

$G_e$ : matriz de variâncias e covariâncias genotípicas entre plantas dentro de famílias; e

$a$ : vetor de pesos econômicos, idêntico ao estabelecido para a seleção entre famílias.

O ganho esperado no caráter  $j$ , quando a seleção é praticada sobre o índice, foi expresso por:

$$\Delta_{j(i)} = DS_{j(i)} \times h^2_j$$

Neste caso, a seleção dentro foi realizada entre todos os indivíduos das famílias selecionadas e avaliadas no experimento.

#### 4.8.7. Ganhos de seleção em relação ao progenitor recorrente

Para cada procedimento de seleção e característica selecionada, foram calculados ganhos em relação ao progenitor Monarca. Para a determinação dos ganhos, utilizou-se a seguinte expressão:

$$GS_{(i)}(\%) = \frac{(\bar{X}_{fs(i)} - \bar{X}_{M(i)})}{\bar{X}_{M(i)}} \times 100$$

em que:

$\bar{X}_{fs(i)}$ : média das famílias selecionadas para a característica  $i$ ; e

$\bar{X}_{M(i)}$ : média geral do progenitor Monarca para a característica  $i$  no experimento.

### 4.9. Análises estatísticas e genéticas da população de mapeamento

#### 4.9.1. Teste de normalidade

##### 4.9.1.1. Teste de assimetria

O coeficiente de assimetria fornece o grau de desvio ou afastamento da simetria de uma distribuição de dados. Em uma distribuição simétrica, tem-se igualdade dos valores da média, mediana e moda. A estimativa do coeficiente de assimetria ( $s$ ) de uma variável  $X$ , numa amostra de tamanho  $n$ , é dada por:

$$\hat{s} = \frac{n}{(n-1)(n-2)} \cdot \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s(X)} \right)^3$$

Se o resultado for zero, a distribuição é simétrica, se o resultado for negativo, a distribuição é assimétrica negativa (inclinada para a esquerda) e se o resultado for positivo, a distribuição é assimétrica positiva (inclinada para a direita).

As hipóteses  $H_0: s = 0$  e  $H_a: s \neq 0$  serão testadas pelo teste t:



$$t = \frac{\hat{s}}{\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}}$$

O valor de t calculado será comparado ao de t tabelado, a nível  $\alpha$  de probabilidade, com infinitos graus de liberdade. Se  $|t_{\text{cal}}| \geq t_{\text{tab}}$ , rejeita-se  $H_0$ . (SCHUSTER e CRUZ, 2004).

#### 4.9.1.2. Teste de curtose

Curtose é o grau de achatamento de uma distribuição, considerado usualmente em relação à distribuição normal. Segundo Schuster e Cruz (2004), para estimar o grau de curtose (k) de uma variável X, utiliza-se a seguinte fórmula:

$$\hat{k} = \frac{n.(n+1)}{(n-1)(n-2)(n-3)} \cdot \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s(X)} \right)^4 - \frac{3.(n-1)^2}{(n-2)(n-3)}$$

Se o resultado for zero, a distribuição de freqüências é a própria distribuição normal, sendo chamada de mesocúrtica ( $k = 0$ ), se o resultado for negativo ( $k < 0$ ), a distribuição é achatada (alta variabilidade), e chamada de planicúrtica se o resultado for positivo ( $k > 0$ ) e a distribuição concentrada em torno da média (alta homogeneidade) é chamada de leptocúrtica.

Para testar a hipótese  $H_0: k = 0$  vs  $H_a: k \neq 0$  tem-se o teste t:

$$t = \frac{\hat{k}}{\sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}}$$

Este valor de t calculado será comparado ao valor de t tabelado, a nível  $\alpha$  de probabilidade, com infinitos graus de liberdade. Se  $|t_{\text{cal}}| \geq t_{\text{tab}}$ , rejeita-se  $H_0$ .

#### 4.9.1.3. Teste de Lilliefors

O teste de Lilliefors apresenta o mesmo princípio que o teste de qui-quadrado ( $\chi^2$ ), porém considera os dados individualmente, ou seja, dada uma

amostra de tamanho  $n$ , agrupam-se os dados em  $k$  classes, podendo-se formar uma distribuição de frequência com intervalos de classes iguais a 0,25, 0,50, 0,75, 1,0 do desvio padrão. As frequências esperadas ( $FE_i$ ) de cada classe são encontradas de acordo com a distribuição normal reduzida  $z$ , multiplicando-se a área de cada classe pelo número total de observações ( $n$ ), e as frequências observadas ( $FO_i$ ) são os números de observação de cada classe ( $n_i$ ). Para isso, obtêm-se os valores de  $z$  com a introdução de uma modificação, ampliando o seu uso para os casos em que a média e a variância não são conhecidas, mas sim estimadas por meio dos dados da amostra, como segue:

$$z_i = \frac{X_i - \bar{X}}{s},$$

em que:

$z_i$  = valor da variável normal padronizada da classe  $i$ ;

$X_i$  = valor máximo da classe  $i$ ;

$\bar{X}$  = média geral da amostra de dados; e

$s$  = desvio padrão da amostra de dados.

Na maioria das vezes, o teste de Lilliefors é mais poderoso que o  $\chi^2$ .

O teste de Lilliefors pressupõe o cálculo de todos os  $z_i$ , que devem ser ordenados para as seguintes considerações:

$F(z_i) = FE_i \leq z_i$  = valor da tabela de distribuição normal reduzida (área);

$S(z_i) = FO_i \leq z_i = n_i/n$ , em que:

$n_i$  = número de valores observados  $\leq z_i$ ; e

$n$  = número total de observações da amostra.

$D$  = Máximo  $|F(z_i) - S(z_i)|$ .

O teste é bilateral, como segue:

$H_0$ : é razoável estudar os dados por meio da distribuição normal; e

$H_a$ : não é razoável estudar os dados por meio da distribuição normal.

Rejeita-se a hipótese de nulidade quando  $D_{cal} \geq D_{tab}$ , em um nível  $\alpha$  de probabilidade com  $n$  observações, caso contrário não se rejeita  $H_0$ .

Deve-se lembrar, porém, que a não rejeição de  $H_0$  indica apenas que esta é uma razoável aproximação da distribuição conhecida.

#### 4.9.2. Análise de variância da população de mapeamento

Foi usado para avaliação do teor de proteína nas NILs o mesmo modelo de análise, que foi efetuada no item 4.7.3, realizada segundo o delineamento em blocos completos casualizados com duas repetições e com as mesmas testemunhas adicionais.

#### 4.10. Extração, amplificação e análise do DNA por marcadores SSR

Para a extração de DNA, foram trituradas cinco gramas de sementes de cada linhagem e dos parentais. Amostras de DNA das sementes foram extraídas pelo método descrito segundo McDonald *et al.* (1994) com algumas modificações. Em um tubo eppendorff de 1,5 mL contendo cerca de 50 mg de sementes, foram adicionados 1.000  $\mu$ L de tampão de extração contendo Tris-HCl 0,2 M, NaCl 0,28 M, EDTA 25 mM e SDS 10%. As amostras foram extraídas em Politrone e, em seguida, centrifugadas por 10 min por 14.000 rpm em microcentrífuga eppendorff 1.405. Os sobrenadantes foram transferidos para novos tubos, sendo, em seguida, acrescentados 10  $\mu$ L de proteinase K 10 mg/mL e 10  $\mu$ L de CaCl<sub>2</sub> 1 mM, e colocado em banho-maria a 55 °C por 30 min. Foram, então, adicionados às amostras 900  $\mu$ L de isopropanol e deixadas em repouso por 2 min. Após este tempo, as amostras foram centrifugadas por 10 min a 14.000 rpm. Os sobrenadantes foram descartados e os precipitados secos por 15 min. à temperatura ambiente, sendo, a seguir, ressuspensos em solução de Tris-EDTA (TE) 10 mM e 1 mM contendo RNase 60  $\mu$ g/mL e colocados em banho-maria por 90 min. As amostras foram precipitadas novamente pela adição de 900  $\mu$ L de isopropanol e deixadas em precipitação por 2 min. Logo após, as amostras foram novamente centrifugadas por 10 min. a 14.000 rpm e os sobrenadantes descartados. Os precipitados foram ressuspensos ao final em TE 10 mM e 1 mM. A concentração do DNA era estimada espectrofotometricamente por leitura da absorbância em 260 nm, sendo cada unidade de absorbância correspondente a 50  $\mu$ g/mL de DNA fita dupla (SAMBROOK *et al.*, 1989).

As reações de amplificação de microssatélites foram feitas em um volume total de 15  $\mu$ L, contendo 12,5  $\mu$ M de Tris-HCl (pH 8,3), 62,5  $\mu$ M de KCl, 2,5 mM de MgCl<sub>2</sub>, 125  $\mu$ M de cada um dos deoxinucleotídeos (dATP, dTTP, dGTP e dCTP), 0,6  $\mu$ M de cada *primer*, uma unidade da enzima Taq-polimerase e 18 ng

de DNA. As amplificações foram realizadas em Termociclador Perkin Elmer 9.600 programado para uma etapa inicial de 7 min. a 72°C, seguida de 30 ciclos de 1 min. a 94°C, 1 min. a 50°C e 2 min. a 72°C. Por fim, uma etapa de 7 min a 72°C. Os fragmentos amplificados foram separados por eletroforese em gel não desnaturante de poliacrilamida 10% por 4 horas a 100 V, e corados com solução de brometo de etídio (0,2 µg/mL). O gel tem 16,5 × 12,5 cm e espessura de 0,5 mm. Terminada a eletroforese, os géis foram fotodigitalizados pelo sistema de fotodocumentação Eagle Eye II (Stratagene).

Foram testados aproximadamente 380 *primers* microssatélites com base no mapa integrado da soja (CREGAN *et al.*, 1999), e aqueles que se mostraram polimórficos nos bulks formados pelas 168 NILs foram amplificados em toda a população para obter os dados moleculares de cada genótipo.

#### 4.11. Obtenção do mapa de ligação

A segregação de cada fragmento polimórfico foi testada nas populações por meio do teste qui-quadrado. Apenas os fragmentos, cuja segregação se ajustou à proporção esperada de 1:1 ( $P < 0,05$ ), foram utilizados nas análises posteriores.

O mapeamento dos marcadores e o estabelecimento dos grupos de ligação foram feitos com a utilização do programa GQmol. Os marcadores foram agrupados, utilizando-se LOD mínimo de 3,0 e frequência máxima de recombinação de 0,30. A distância de mapeamento de Kosambi foi utilizada para a conversão da frequência de recombinação para centiMorgans (cM).

#### 4.12. Mapeamento de QTLs para conteúdo de proteína

A identificação de marcadores ligados a um QTL que controla o teor de proteína foi realizada analisando-se a co-segregação entre os marcadores amplificados na população e o fenótipo de cada linhagem, pelo método da análise de variância e por regressão linear e múltipla. A seleção de marcadores utilizando regressão múltipla foi feita utilizando o procedimento de eliminação *stepwise*, com nível de significância de 10% para a entrada e saída das variáveis independentes (marcadores) (SCHUSTER e CRUZ, 2004). As informações do mapa de ligação foram utilizadas para localizar e caracterizar os efeitos dos possíveis QTLs por

meio do mapeamento por intervalo simples baseado na análise de regressão segundo Haley e Knott (1992). As estimativas dos valores aditivos, o coeficiente de determinação do QTL correspondente ao pico de maior significância estatística do QTL e a posição do possível QTL foram declarados quando os valores da razão da máxima verossimilhança (*Likelihood Ratio* – LR) excederem os valores críticos de corte ( $\alpha = 0,05$ ) em cada grupo de ligação. O ponto de corte da razão de máxima verossimilhança (LR) para considerar a presença de um QTL pelo método de mapeamento por intervalo simples foi obtido a partir de um teste de permutação como relatado por Doerge e Churchill (1996). O teste é realizado misturando-se os valores da característica quantitativa em questão, repetindo-se a análise. Os valores de LR críticos foram determinados pela realização de 1.000 permutações.

#### 4.13. Recursos computacionais

Todas as análises estatísticas e genéticas utilizadas foram processadas, utilizando-se os programas Genes ([www.ufv.br/genes/genes.htm](http://www.ufv.br/genes/genes.htm)) e GQMOL ([www.ufv.br/dbg/gqmol/gqmol.htm](http://www.ufv.br/dbg/gqmol/gqmol.htm)).

## 5. RESULTADOS E DISCUSSÃO

### 5.1. Análise das testemunhas e correção dos dados da população de seleção

Inicialmente foi realizada uma análise de variância com todas as fileiras úteis do experimento (375 famílias e 15 blocos de testemunhas intercalares) com o objetivo apenas de correção dos dados, pelo fato de se necessitar de uma regularidade na disposição das famílias dentro dos blocos, pois isto não seria possível depois da eliminação de muitas famílias. Na Tabela 4, encontram-se o resumo da análise de variância e as estimativas dos parâmetros genéticos referentes às testemunhas intercalares para as três características avaliadas: produção de grãos por planta, teor de proteína e teor de óleo.

Observam-se diferenças significativas entre as testemunhas para todas as características. Além disso, a variação ambiental no experimento foi pouco pronunciada, mostrada pelo coeficiente de variação experimental que assumiu valores abaixo de 15%, sendo que isso irá refletir em uma menor discrepância entre os valores originais e os corrigidos.

Tabela 4 - Resumo da análise de variância para as testemunhas das características produção de grãos (PROD), teor de proteína (PTN) e teor de óleo (OL)

FV	GL	QM		
		PROD	PTN	OL
Testemunha	1	1729,154**	1493.896**	445.599**
Resíduo	28	3,141	1.173	1,055
Média		12,85	46,05	16,88
Coeficiente de variação – CV <sub>e</sub> (%)		13,78	2,352	6,084
Variância fenotípica		115,28	99,59	29,70
Variância ambiental		0,209	0,078	0,070
Variância genotípica		115,07	99,51	29,64
Coeficiente de determinação genotípica		99,81	99,92	99,76
Correlação intraclasses		97,34	98,83	96,56
Coeficiente de variação genético – CV <sub>g</sub>		83,46	21,66	32,25
Razão CV <sub>g</sub> /CV <sub>e</sub>		6,06	9,21	5,29

(\*\*) significativo a 1% de probabilidade pelo teste F.

### 5.2. Estimação de parâmetros genéticos e predição de ganhos para produção de grãos utilizando análise de gerações com informação entre e dentro de famílias

Esta análise foi realizada em uma população composta de 332 famílias, que inicialmente passou por uma seleção prévia para desuniformidade de estande, porte baixo e acamamento. Após esta seleção visual, os valores fenotípicos para produção de grãos por planta foram corrigidos, tanto em nível de família como de indivíduo dentro de família (oito plantas). Em seguida, procedeu-se à análise de variância de acordo com Cruz (2006a). Após definidos os parâmetros genéticos desta população, foram feitos os procedimentos de seleção conforme objetivos estabelecidos na primeira etapa do primeiro ciclo.

### 5.2.1 Análise de variância

Na Tabela 5, são apresentados os quadrados médios da variação entre e dentro de progênies e dos progenitores, as variâncias fenotípicas e genéticas entre e dentro de famílias, a variância aditiva, as herdabilidades, as médias e os coeficientes de variação para produção de grãos por planta. Inicialmente destaca-se a significância do quadrado médio entre e dentro de famílias, indicando a existência de variabilidade genotípica para produção de grãos por planta na população estudada, evidenciando a possibilidade de seleção de genótipos superiores quando utilizada alguma estratégia de seleção. A existência de variação estatisticamente significativa, entre e dentro de famílias, é entendida como possibilidade de esta característica ser melhorada pela seleção. No entanto, esta constatação nada indica sobre a quantidade de variabilidade genética dessa população. Portanto, para uma maior informação do conhecimento da magnitude relativa das variações genéticas e ambientais e da sua utilização, devem-se empregar estimativas de herdabilidades e dos coeficientes de variação genéticos e ambientais.

O coeficiente de variação experimental obtido próximo de 15%, encontra-se abaixo dos valores obtidos por Reis (2000), Farias Neto e Vello (2001), Costa *et al.* (2004) e Azevedo Filho *et al.* (1998) que utilizaram o mesmo tipo de delineamento aqui empregado. Os coeficientes de variação experimental entre e dentro de famílias apresentam valores próximos e de pequena magnitude. Além disso, as razões  $CV_g/CV_e$ ;  $CV_{ge}/CV_{\epsilon e}$  e  $CV_{gd}/CV_{\epsilon d}$  se apresentam superiores à unidade, indicando, conforme comentado por Cruz e Regazzi (1997), que possibilidades de ganhos com métodos relativamente mais simples de seleção

poderão ser obtidos tanto em nível de família como dentro, mas com superioridade da seleção quando praticada entre famílias, pois a relação  $CV_{ge}/CV_{\epsilon e}$  foi superior à relação  $CV_{gd}/CV_{\epsilon d}$ .

Tabela 5 – Estimativas dos quadrados médios em nível de médias de famílias, dentro de famílias e das testemunhas, das variâncias fenotípicas, genéticas e aditiva, das herdabilidades, das médias e dos coeficientes de variação para a característica produção de grãos por planta.

FV	Progênes		BARC-8		Monarca	
	GL	QM	GL	QM	GL	QM
Entre parcelas	331	141,704**	14	4,776	14	35,828
Dentro de parcelas	2324	14,889**	75	2,209	75	6,173
Resíduo entre	14	20,302				
Resíduo dentro	75	4,191				
Média		17,97		5,24		20,7
Variância fenotípica entre	17,713					
Variância fenotípica dentro	14,889					
Variância ambiental entre	3,384					
Variância ambiental dentro	4,191					
Variância genotípica entre	14,329					
Variância genotípica dentro	10,698					
Variância aditiva	4,804					
Herdabilidade ampla entre	0,809					
Herdabilidade ampla dentro	0,718					
Herdabilidade restrita entre ( $h^2_{R.entre}$ )	0,407					
Herdabilidade restrita dentro ( $h^2_{R.dentro}$ )	0,081					
Herdabilidade restrita total ( $h^2_{R.total}$ )	0,258					
Coeficiente de variação experimental ( $CV_e\%$ )	15,33					
Coeficiente de variação fenotípico ( $CV_f\%$ )	31,77					
Coeficiente de variação genético ( $CV_g\%$ )	27,84					
Coeficiente de variação genético entre famílias ( $CV_{ge}\%$ )	21,06					
Coeficiente de variação genético dentro famílias ( $CV_{gd}\%$ )	18,20					
Coeficiente de variação experimental entre famílias ( $CV_{\epsilon e}\%$ )	10,24					
Coeficiente de variação experimental dentro famílias ( $CV_{\epsilon d}\%$ )	11,39					
$CV_g\%/CV_e\%$	1,818					
$CV_{ge}\%/CV_{\epsilon e}\%$	2,058					
$CV_{gd}\%/CV_{\epsilon d}\%$	1,598					

(\*\*) significativo a 1% de probabilidade pelo teste F.

As herdabilidades tanto no sentido amplo e restrito entre e dentro de famílias encontram-se dentro da faixa de variação dos valores obtidos por outros autores, tais como Weber e Moorthy (1952), Hanson e Weber (1961), Mauro *et al.* (2000), Farias Neto e Vello (2001) e Reis *et al.* (2004), apesar da baixa magnitude encontrada para a herdabilidade no sentido restrito dentro. Esse valor



reduzido de herdabilidade restrita dentro pode ser explicado pela própria existência de uma menor fração de variância genética aditiva dentro de famílias, devido ao maior grau de endogamia destas famílias e provavelmente à subestimação desta variância genética ou à superestimação da variância fenotípica dentro, resultante de um menor grau de precisão na estimação dos parâmetros genéticos entre plantas dentro de famílias. A superioridade da herdabilidade no sentido restrito entre famílias em relação à herdabilidade restrita dentro de famílias refletirá em uma maior proporção do diferencial de seleção a ser transmitida à geração seguinte quando se considerar a média de família como unidade de seleção ao invés de considerar como unidade de seleção o indivíduo dentro da família, como poderá ser verificado ao comparar diferentes estratégias de seleção.

### **5.2.2. Predição de ganhos**

Conforme mencionado em material e métodos, foram dois os objetivos desta primeira etapa. O primeiro foi o de selecionar as 205 famílias com maior produção de grãos, com base nas informações de suas médias, sendo as selecionadas submetidas a uma nova seleção utilizando informações tanto da produção de grãos como do teor de óleo e proteína. O segundo objetivo foi o de comparar três diferentes estratégias de seleção que maximizassem os ganhos por seleção para produção de grãos quando selecionadas 18% desta população.

São apresentadas, na Tabela 6, as estimativas das herdabilidades restritas em nível de família, dentro e total, da média original, da média dos selecionados, do diferencial de seleção, do ganho esperado pela seleção, do ganho percentual e das eficiências de seleção, utilizando as estratégias de seleção entre, seleção dentro, seleção entre e dentro, seleção massal e seleção combinada para produção de grãos por planta.

A seleção de 205 famílias promoveu apenas um ganho genético de 3,8%, o que era esperado, pois a intensidade de seleção foi baixa, ou seja, alta percentagem de famílias selecionadas (61,7%). Já a seleção, quando praticada com maior intensidade, com 18% de indivíduos ou famílias selecionadas, promoveu maiores ganhos, sendo os maiores ganhos preditos proporcionados quando se praticou a seleção massal, seguida da seleção entre e dentro, da seleção entre e por fim a seleção combinada. Este fato contradiz as expectativas,

pois diversos autores já citados evidenciaram um maior ganho quando do emprego da seleção combinada.

Tabela 6 - Estimativas das herdabilidades restritas em nível de família, dentro de família e total, da média original ( $X_o$ ), da média dos selecionados ( $X_s$ ), do diferencial de seleção ( $DS$ ), do ganho esperado pela seleção ( $GS$ ), dos ganhos percentuais ( $GS\%$ ) e das eficiências de seleção, utilizando as estratégias de seleção entre (SE), seleção dentro (SD), seleção entre e dentro (SED), seleção massal (SM) e seleção combinada (SC) para produção de grãos por planta.

Parâmetros	Estratégias					
	SE	SE	Dentro (SD)	SED	SM	SC
Nº de selecionados	205	60	240 (60 × 4)	240	240	240
$h^2_R$	0,407	0,407	0,081		0,258	(0,407 e 0,081)
$\bar{X}_o$	17,97	17,97	24,23	17,97	17,97	17,97
$\bar{X}_s$	19,65	24,23	27,08		28,84	
$DS$	1,68	6,26	2,85		10,87	2,43
$GS$	0,68	2,54	0,23		2,80	2,43
$GS\%$	3,8	14,16	0,95	15,11	15,59	13,52
EF.SM/SE		1,10				
EF.SED/SE		1,07				
EF. SC/SE		0,95				

Um dos fatores determinantes deste resultado foi a pequena magnitude encontrada para herdabilidade restrita dentro, o que também proporcionou pequeno ganho de seleção dentro de famílias (0,95%). Pelos resultados apresentados, a pequena ou até mesmo a menor superioridade das estratégias de seleção que usam a informação de plantas individuais (SED, SM e SC) sobre a estratégia que utiliza apenas a informação de médias de famílias, não compensam o esforço despendido para obter essas informações, pelo menos na geração em que se encontra. Isto torna preferível a análise de um maior número de famílias, só com a informação da média entre, do que a análise de um menor número, mas com a informação dentro, ou seja, em nível de família e de indivíduo.

### 5.3. Estimação de parâmetros genéticos e predição de ganhos entre famílias para produção de grãos por plantas, teor de proteína e teor de óleo.

#### 5.3.1. Análise de variância e estimação de parâmetros genéticos

Os resultados da análise de variância para as características produção de grãos por plantas (PROD), teor de proteína (PTN) e teor de óleo (OL) utilizando o delineamento de famílias com testemunhas intercalares, são apresentados na Tabela 7, em que o teste F foi significativo a 1% de probabilidade para as três características avaliadas. A existência de variância genética altamente significativa sugere a possibilidade de obtenção de ganhos pela aplicação de seleção para as três características.

Os valores dos coeficientes de variação experimental podem ser considerados baixos, conforme os padrões normalmente encontrados para estas características em soja (SHORTER *et al.*, 1976; PRADO, 1994; SCOTT e KEPHART, 1997; FARIAS NETO e VELLO, 2001; MIRANDA, 2006), podendo-se admitir boa precisão na obtenção e análise dos dados, refletindo, portanto, numa maior confiabilidade dos resultados a serem obtidos. A relação  $CV_g/CV_e$  superior a uma unidade para as três características é um indicador da existência de possibilidades de sucesso na obtenção de ganhos genéticos, pois a variação genética supera a variação ambiental, favorecendo a seleção. Além da relação  $CV_g/CV_e$ , o conhecimento dos valores de herdabilidade permite indicar as possibilidades de ganhos genéticos com a seleção, pois desempenha importante função no entendimento da estrutura genética da população por mostrar a quantidade de variação existente entre famílias. Conforme pode ser observado na Tabela 7, as herdabilidades entre médias de famílias para a característica PROD apresentaram valor relativamente elevado, já as características PTN e OL apresentaram valores intermediários quando comparados aos de outros trabalhos realizados por Thorne e Fehr (1970), Shannon *et al.* (1972), Shorter *et al.* (1976), Mckendry *et al.* (1985), Wilcox (1998), Prado (1994), Mauro *et al.* (2000), Reis *et al.* (2004), Miranda (2006) e Costa *et al.* (2008).

O conhecimento das correlações entre caracteres é um aspecto genético de grande valor para o melhoramento de plantas, que deve receber atenção dos melhoristas. Seu conhecimento é importante porque mostra como a seleção para um caráter influencia a expressão de outros caracteres. Nos programas de melhoramento, geralmente, além de se visar ao aprimoramento de um caráter principal busca-se também manter ou melhorar a expressão de outros caracteres, sendo necessário distinguir duas causas nessa correlação: a genética e a ambiental. Somente a correlação genética envolve associações de natureza herdável e, por conseguinte, de real interesse num programa de melhoramento.

Tabela 7 - Resumo da análise de variância para as características produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL) do experimento de famílias F<sub>4</sub> de soja com testemunhas intercalares.

FV	GL	QM		
		PROD	PTN	OL
Famílias	204	12,363**	3,694**	4,022**
Testemunha	1	1228,685**	1048,64**	320,560**
Resíduo	18	2,326	1,367	1,266
Média das famílias		19,65	39,03	17,67
Média das testemunhas		13,14	45,92	16,75
Coeficiente de variação – CV <sub>e</sub> (%)		7,762	2,997	6,367
Variância fenotípica		12,363	3,694	4,022
Variância ambiental		2,326	1,367	1,266
Variância genotípica		10,036	2,327	2,756
Herdabilidade		81,18	62,99	68,53
Coeficiente de variação genético – CV <sub>g</sub>		16,122	3,909	9,395
Razão CV <sub>g</sub> /CV <sub>e</sub>		2,077	1,305	1,476

(\*\*) significativo a 1% de probabilidade pelo teste F.

Na Tabela 8, podem ser verificadas as estimativas das correlações fenotípica, genotípica e ambiental entre produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL). Valores já esperados negativos e significativos de correlações fenotípicas e genotípicas entre PTN e OL foram encontrados, só que de magnitude abaixo da maioria dos trabalhos aqui citados, tais como os de Miller e Fehr (1979), Openshaw e Hadley (1984), Scott e Kephart (1997), Wilcox e Shibles (2001) e Miranda (2006).

Este fato pode ser explicado, segundo Gomes (1995) e Rios (1997), pela baixa eficiência deste tipo de delineamento na estimação de parâmetros genéticos. A ausência ou presença de correlações genotípicas positivas de baixa magnitude entre PROD e PTN e PROD e OL indicam que estas duas características PTN ou OL podem ser melhoradas simultaneamente com a produção de grãos por planta, mas dificilmente as três conjuntamente. A ausência de correlação negativa entre PROD e PTN deve-se ao fato de esta população ter se originado de progenitores que apresentavam diferenças significativas entre teores de proteína, mas que apresentavam desempenho semelhante para produção de grãos. Fato contrário ao observado na maioria dos trabalhos já citados que encontravam sempre correlações negativas, pois o progenitor de alto teor protéico utilizado no cruzamento sempre possuía baixo desempenho agrônômico, por ser um genótipo geralmente não adaptado.

Tabela 8 - Coeficientes de correlação fenotípica ( $r_F$ ), genotípica ( $r_G$ ) e ambiental ( $r_E$ ) entre produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL) do experimento de famílias F4 de soja com testemunhas intercalares

Características	Correlações		
	$r_F$	$r_G$	$r_E$
PROD × PTN	0,103	0,133	0,031
PROD × OL	0,073	0,102	-0,012
PTN × OL	-0,557**	-0,398**	-0,864**

\*\* Significativo a 1% de probabilidade pelo teste t.

### 5.3.2. Predição de ganhos

Nesta segunda etapa do primeiro ciclo de seleção, o objetivo principal foi o de selecionar as 40 famílias que maximizassem os ganhos para teor protéico, proporcionassem ganhos de moderados a altos para produção de grãos e reduzissem o mínimo possível o teor de óleo. Na seleção destes genótipos e na estimativa dos ganhos preditos, foram empregados os seguintes critérios de seleção conforme citados por Cruz e Carneiro (2003): Seleção Direta e Indireta (SD); seleção baseada em índices de Smith e Hazel (SH), Kempthorne e Nordskog (KN), Pesek e Baker (PB) e Pesek e Baker generalizado por Tai (PB-Tai). É importante destacar que, dificilmente, se consegue elevar o teor de proteína sem a redução do teor de óleo e vice-versa. Como exemplo, pode ser citado o trabalho de Wilcox (1998) em que um aumento de 0,58% no teor de proteína reduziu em 2,3% a média para teor de óleo. No entanto, o uso de seleção simultânea baseada em índice pode atenuar a redução de uma dessas duas características ou até mesmo conseguir um pequeno ganho para as duas características simultaneamente como os resultados obtidos por Openshaw e Hadley (1984).

Na Tabela 9, são apresentadas a média original, a média dos selecionados, a estimativa da herdabilidade no sentido amplo entre famílias e os ganhos percentuais obtidos pela seleção direta e indireta para as características produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL). Como podem ser observados, os ganhos diretos foram maiores que os indiretos, pois as características eram pouco correlacionadas ou correlacionadas negativamente como no caso de PTN e OL. Além disso, a seleção direta para

PROD praticamente não interferiu na média das características PTN e OL, o que é extremamente desejável. É também interessante notar que ganhos indiretos similares para PROD foram obtidos quando o caráter principal sob seleção foi tanto PTN como OL. Os ganhos indiretos foram preditos pelo diferencial de seleção direta e não pela expressão que utiliza o coeficiente de correlação genotípica, assim, pode-se afirmar que as correlações genotípicas não foram estimadas incorretamente, como se havia suposto, pois se tivessem sido, os ganhos para PROD seriam diferentes em magnitude.

Tabela 9 – Estimativas de ganhos percentuais para produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL) avaliados entre médias de famílias RC<sub>1</sub>F<sub>4</sub> de soja por meio de seleção baseada em seleção direta e indireta

Características	$\bar{X}_o$	$\bar{X}_s$			$h^2_{A.entre}$	GS%*			Total
		PROD	PTN	OL		PROD	PTN	OL	
PROD	19,65	<b>24,80</b>	39,34	17,93	0,809	<b>21,20</b>	0,5	0,99	22,69
PTN	39,03	20,60	<b>41,79</b>	16,16	0,630	3,91	<b>4,45</b>	-5,85	2,51
OL	17,67	20,66	37,31	<b>20,32</b>	0,685	4,17	-2,74	<b>10,27</b>	11,70

( $\bar{X}_o$ ): média original; ( $\bar{X}_s$ ): média dos selecionados;  $h^2_{A.entre}$ : estimativa da herdabilidade no sentido amplo entre famílias; (GS%): ganhos percentuais; (\*) em negrito ganho direto.

Os resultados dos ganhos preditos quando da utilização dos índices de seleção encontram-se na Tabela 10. Foi empregado o índice clássico de Smith e Hazel (SH) com cinco combinações com diferentes pesos econômicos, inicialmente considerando as três características (PROD, PTN e OL) como principais e dando a elas pesos econômicos, sendo SH1 (1; 2,5; 5) baseado na proporcionalidade da composição; SH2 (16,12; 3,9; 9,39) baseado nos coeficientes de variação genéticos; SH3 (1; 1; 1) com iguais pesos econômicos; SH4 (1; 5;5) com pesos econômicos iguais para PTN e OL; e SH5 (1,10) que somente considerou PROD e PTN como principais e priorizou PTN. O índice restrito de Kempthorne e Nordskog (KN), tendo como características principais PROD e PTN, restringiu o ganho para OL como nulo ao se adotar a COV (escores do índice, valores genéticos de OL) = 0, considerando três variações deste índice: KN1 (1; 2,5; 0), KN2 (16,12; 3,9; 0), e KN3 (1; 10; 0). O índice baseado nos ganhos desejados de Pesek e Baker (PB) se apresenta com três variações: a primeira com base no desvio padrão genético das características

PROD, PTN e OL tidas como principais em PB1(3,16; 1,52; 1,66), a segunda e a terceira somente com PROD e PTN em PB2 (3,16; 1,52) e PB3(3,16; 3,04) como principais, não incluindo OL nem como secundária. E o índice de Pesek e Baker generalizado por Tai (PB-Tai) em que PROD e PTN são as características principais e OL secundária com ganho nulo em PB-Tai-1(3,16; 1,52; 0); PB-Tai-2(3,16; 3,04; 0).

Tablela 10 – Estimativas de ganhos percentuais para produção de grãos (PROD), teor de proteína (PTN) e teor de óleo (OL) avaliadas em famílias RC<sub>1</sub>F<sub>4</sub> de soja por meio de seleção baseada nos índices de Smith e Hazel (SH), Kempthorne e Nordskog (KN), Pesek e Baker (PB) e Pesek e Baker generalizado por Tai (PB-Tai).

Índice	PROD			PTN			OL			GS% Total
	$\bar{X}_o$	$\bar{X}_s$	GS%	$\bar{X}_o$	$\bar{X}_s$	GS%	$\bar{X}_o$	$\bar{X}_s$	GS%	
SH1	19,65	22,54	11,92	39,03	39,15	0,19	17,67	19,51	7,13	19,24
SH2	19,65	24,66	20,67	39,03	39,08	0,09	17,67	18,66	3,84	24,60
SH3	19,65	24,54	20,21	39,03	39,42	0,64	17,67	18,52	3,30	24,15
SH4	19,65	22,06	9,96	39,03	40,31	2,07	17,67	18,70	4,00	16,03
SH5	19,65	21,26	6,65	39,03	41,74	4,39	17,67	16,24	-5,55	5,49
KN1	19,65	23,72	16,8	39,03	40,91	3,04	17,67	16,78	-3,46	16,38
KN2	19,65	24,75	21,07	39,03	39,59	0,91	17,67	17,61	-0,22	21,76
KN3	19,65	21,57	7,91	39,03	41,54	4,06	17,67	17,07	-2,31	9,66
PB1	19,65	22,31	10,99	39,03	40,68	2,67	17,67	18,28	2,35	16,01
PB2	19,65	23,07	14,12	39,03	41,26	3,60	17,67	16,51	-4,48	13,24
PB3	19,65	21,74	8,64	39,03	41,67	4,26	17,67	16,37	-5,05	7,85
PB-Tai-1	19,65	23,29	15,02	39,03	41,12	3,38	17,67	16,93	-2,86	15,54
PB-Tai-2	19,65	21,71	8,49	39,03	41,63	4,2	17,67	16,72	-3,68	9,01

( $\bar{X}_o$ ): média original; ( $\bar{X}_s$ ): média dos selecionados;  $h_{A.entre}^2$ : estimativa da herdabilidade no sentido amplo entre famílias; (GS%): ganhos percentuais.

SH1 (1; 2,5; 5); SH2 (16,12; 3,9; 9,39); SH3 (1; 1; 1); SH4 (1; 5; 5); SH5 (1; 10); KN1 (1; 2,5; 0); KN2 (16,12; 3,9; 0); KN3 (1; 10; 0); PB1(3,16; 1,52; 1,66); PB2(3,16; 1,52); PB3(3,16; 3,04); PB-Tai-1(3,16; 1,52; 0); PB-Tai-2(3,16; 3,04; 0).

Os critérios baseados nos índices SH2, SH3 e KN2 que conduziram à maximização dos ganhos preditos para produção de grãos não puderam ser considerados na seleção dos genótipos para o próximo ciclo de seleção, pois não proporcionaram ganhos considerados satisfatórios para PTN. Já o grupo de genótipos que foram selecionados baseados nos índices SH1, SH4 e PB1 foram os que tiveram os ganhos mais equilibrados para PROD e OL e que mantiveram ou aumentaram em pequena proporção a média para PTN, mas que também não atingem os objetivos do presente trabalho. O índice SH2, que utiliza como pesos econômicos os CVg como sugerido por Cruz (1990), proporcionou o maior ganho total, o que também aconteceu nos trabalhos realizados por Costa *et al.* (2004 e

2008) quando avaliou oito características agronômicas conjuntamente utilizando tanto o índice de Smith e Hazel como o de Pesek e Baker.

Os ganhos obtidos pelos índices KN1, PB2 e PB-Tai-1 foram semelhantes e proporcionaram ganhos moderados tanto para PROD como para PTN, portanto, inferiores a SD, além de apresentarem uma redução significativa para OL. Um novo ciclo de seleção poderia ser praticado nos genótipos selecionados por estes índices a fim de melhorar ainda mais uma dessas duas características PROD ou PTN.

Em relação aos índices SH5, KN3, PB3 e PB-Tai-2, foram bem similares na seleção das famílias como mostrado na Tabela 11. Estes índices proporcionaram ganhos próximos à seleção direta para PTN, mas com a vantagem de proporcionar maiores ganhos para PROD e um ganho negativo menos pronunciado do que a seleção indireta provocou para OL. Os ganhos preditos obtidos pela seleção das 40 famílias quando se aplicou o índice PB-Tai-2 foram aqueles mais uniformemente distribuídos, tendo sido o índice mais indicado para selecionar progênies para a terceira etapa do primeiro ciclo de seleção.

#### **5.4. Estimação de parâmetros genéticos e predição de ganhos para produção de grãos por planta e teor de proteína utilizando análise de gerações com informação entre e dentro de famílias.**

Depois de identificadas as 40 famílias que apresentaram ganhos elevados para teor protéico, de moderados a altos para produção de grãos e que reduzissem o mínimo possível o teor de óleo, foram avaliadas individualmente para PROD e PTN oito plantas por família. Além destas 40 famílias selecionadas, foram incluídas mais três que apresentaram características de produção de grãos mais elevadas e teor de proteína moderado. Ao todo, foram obtidas informações de 344 plantas.



Tabela 11 – Famílias selecionadas em ordem decrescente pelos índices SH5, KN3, PB3 e PB-Tai-2 e amplitude de variação para as características PROD, PTN e OL

Índice	Famílias						PROD		PTN		OL	
							$\bar{X}_m$	$\bar{X}_M$	$\bar{X}_m$	$\bar{X}_M$	$\bar{X}_m$	$\bar{X}_M$
PB-Tai-2	102	99	145	138	149	143	15,15	30,90	39,16	45,16	13,36	19,86
	65	162	115	146	144	48						
	104	5	64	36	101	141						
	116	62	194	47	167	15						
	189	205	155	21	118	200						
	161	59	131	25	127	110						
	87	68	114	30								
SH5	102	99	138	149	162	115	15,15	26,53	40,13	45,16	13,11	19,79
	143	146	145	116	144	65						
	104	155	48	167	159	5						
	36	141	118	156	64	47						
	21	62	194	101	15	127						
	189	25	205	200	125	110						
	131	30	114	165								
KN3	102	145	99	65	143	144	15,15	30,90	39,16	45,19	13,36	19,86
	138	64	149	146	48	101						
	62	36	104	141	5	162						
	115	15	189	194	116	47						
	205	21	167	87	68	25						
	16	131	72	114	200	118						
	110	98	127	67								
PB3	102	99	145	138	149	143	15,15	30,90	39,16	45,16	13,36	19,79
	65	162	115	146	144	48						
	104	5	64	36	101	141						
	116	62	194	47	167	15						
	189	205	155	21	118	200						
	161	59	131	25	127	110						
	87	68	114	30								

$\bar{X}_m$  : menor média na população;  $\bar{X}_M$  : maior média na população.

#### 5.4.1. Análise de variância

Na Tabela 12, são apresentados os quadrados médios da variação entre e dentro de progênies e dos progenitores, as variâncias fenotípicas e genéticas entre e dentro de famílias, a variância aditiva, as herdabilidades, as médias e os coeficientes de variação para as características produção de grãos por planta (PROD) e teor de proteína (PTN). Consideraram-se como quadrados médios do resíduo entre e dentro para produção de grãos os estimados na primeira etapa deste trabalho, devido à sua maior precisão. Houve uma alta significância

( $P < 0,01$ ) do quadrado médio entre e dentro de famílias indicando ainda a existência de variabilidade genotípica para PROD e PTN nesta população reduzida, evidenciando a possibilidade de seleção de genótipos superiores quando utilizada alguma estratégia de seleção. Maior ênfase nesta análise será dada à característica PTN, pois para PROD as estimativas dos parâmetros genéticos mantiveram a mesma magnitude dos valores obtidos na primeira etapa deste trabalho, tendo sido já realizada uma abordagem mais detalhada.

Tanto os coeficientes de variação experimental total, como de variação experimental entre e dentro de famílias para PTN apresentaram-se entre 2 a 4%, demonstrando uma boa precisão experimental, equivalentes aos obtidos por Scott e Kephart (1997), Chung *et al.* (2003) e Soares *et al.* (2004). Verifica-se uma baixa magnitude das variâncias genotípicas entre e dentro e da variância genética aditiva para a característica PTN, indicativo de uma pequena variabilidade genética, refletindo uma maior dificuldade para promover uma melhoria desta característica nesta população quando selecionada. Para a característica PROD, houve uma redução já esperada para as variâncias fenotípicas e genéticas em virtude da seleção a que estas famílias já haviam sido submetidas em etapas anteriores. A constatação de uma relação  $CV_g/CV_e$ ;  $CV_{ge}/CV_{ee}$  próxima e maior que a unidade indica uma maior possibilidade de sucesso na seleção quando realizada entre famílias em relação a possíveis ganhos de seleção dentro de famílias, pois a relação  $CV_{gd}/CV_{ed}$  foi inferior à unidade, indicando que a maior parte da variação fenotípica dentro se deve a causas ambientais, o que dificulta a identificação dos genótipos mais promissores dentro das famílias.

As estimativas de herdabilidades no sentido amplo entre e dentro de famílias apresentam valores ou inferiores ou semelhantes aos encontrados por Kwon e Torrie (1964), Shannon *et al.* (1972), Shorter *et al.* (1976) e Wilcox (1998). Já as estimativas de herdabilidades no sentido restrito entre famílias apresentam valores superiores aos encontrados por Erickson *et al.* (1981), Mckendry *et al.* (1985) e Miranda (2006). A herdabilidade no sentido restrito dentro apresentou valor reduzido, provavelmente causado pela existência de uma menor fração de variância genética aditiva dentro de famílias, devido a seu maior grau de endogamia, como já comentado para a característica produção de grãos. Este fato também ficou evidenciado analisando a relação  $CV_{gd}/CV_{ed}$ . Novamente a superioridade da herdabilidade no sentido restrito entre famílias em relação à herdabilidade restrita dentro de famílias refletirá uma superioridade da seleção

baseada nas médias de famílias do que dentro de famílias, considerada a mesma unidade de seleção.

Tabela 12 - Estimativas dos quadrados médios em nível de média de famílias, dentro de famílias e das testemunhas, das variâncias fenotípicas, genéticas e aditiva entre e dentro de famílias, das herdabilidades, das médias e dos coeficientes de variação para as características produção de grãos por planta (PROD) e teor de proteína (PTN)

FV	PROD						PTN					
	Progênes		Monarca		BARC-8		Progênes		Monarca		BARC-8	
	GL	QM	GL	QM	GL	QM	GL	QM	GL	QM	GL	QM
Entre parcelas	42	89,45**	14	4,78	14	35,83	42	18,40**	4	5,19	4	2,97
Dentro de parcelas	301	9,62**	75	2,21	75	6,17	301	1,490**	24	0,54	24	1,45
Resíduo entre	14	20,30					4	4,082				
Resíduo dentro	75	4,19					24	0,994				
Média		20,53		20,69		5,24		41,04		38,00		53,00
Variância fenotípica entre		11,181						2,300				
Variância fenotípica dentro		9,618						1,491				
Variância ambiental entre		3,383						0,705				
Variância ambiental dentro		4,191						0,993				
Variância genotípica entre		7,797						1,595				
Variância genotípica dentro		5,429						0,498				
Variância aditiva		2,839						0,931				
Herdabilidade ampla entre		0,697						0,694				
Herdabilidade ampla dentro		0,564						0,334				
Herdabilidade restrita entre ( $h^2_{R.entre}$ )		0,38						0,607				
Herdabilidade restrita dentro ( $h^2_{R.dentro}$ )		0,074						0,156				
Herdabilidade restrita total ( $h^2_{R.total}$ )		0,239						0,430				

Continua...

Tabela 12, Continuação...

FV	PROD	PTN
Coefficiente de variação experimental ( $CV_e$ % )	13,34	3,175
Coefficiente de variação fenotípico ( $CV_f$ % )	22,22	4,743
Coefficiente de variação genético ( $CV_g$ % )	17,70	3,525
Coefficiente de variação genético entre famílias ( $CV_{ge}$ % )	13,59	3,077
Coefficiente de variação genético dentro famílias ( $CV_{gd}$ % )	11,34	1,719
Coefficiente de variação experimental entre famílias ( $CV_{\epsilon e}$ % )	8,95	2,045
Coefficiente de variação experimental dentro famílias ( $CV_{\epsilon d}$ % )	9,99	2,428
$CV_g$ % / $CV_e$ %	1,32	1,110
$CV_{ge}$ % / $CV_{\epsilon e}$ %	1,52	1,505
$CV_{gd}$ % / $CV_{\epsilon d}$ %	1,14	0,708

(\*\*) significativo a 1% de probabilidade pelo teste F.

#### 5.4.2. Predição de ganhos

Nesta etapa, foram analisadas 43 famílias  $RC_1F_{4:6}$  com informação de produção de grãos por planta (PROD) e teor de proteína em nível de família e de indivíduo e selecionados 21 genótipos. Os ganhos diretos e indiretos foram obtidos por meio de seleção massal entre e dentro de família, combinada e índice combinado (IC) para as duas características. Na Tabela 13, estão apresentadas as estimativas das herdabilidades restritas em nível de família, dentro de família e total, da média original, da média dos selecionados, do diferencial de seleção, do ganho percentual esperado pela seleção (GS%) e da eficiência da seleção massal sobre as outras utilizando as estratégias acima mencionadas.

Sem considerar a seleção simultânea, os maiores ganhos preditos diretos foram proporcionados quanto se praticou a seleção massal, seguida da seleção entre e dentro, da seleção entre e por fim da seleção combinada tanto para PROD como para PTN. Considerando que na primeira etapa o objetivo foi a seleção apenas para produção de grãos em uma população muito maior e que a média dos selecionados foi de 27,08 g/planta para a seleção entre e dentro e de 28,84 g/planta para a seleção massal, então não deveria ser priorizada a seleção direta para PROD, pois a maior média dos selecionados foi de apenas 26,34 g/planta para a seleção massal, bem inferior às médias obtidas na primeira etapa. Além disso, a seleção direta, utilizando as estratégias de seleção entre e dentro, massal e combinada, não promoveu nenhum ganho indireto para PTN. Também se analisados os ganhos indiretos para PROD quando a seleção é realizada para PTN, observa-se um ganho irrisório para a produção de grãos, e pelo fato de a produtividade ser a principal característica a se desejar em uma variedade, então também não se pode utilizar a seleção direta para teor de proteína nesta população. Portanto, o uso do índice combinado de seleção, no caso o índice clássico de Smith (1936) e Hazel (1943), foi a melhor alternativa para obter ganhos proporcionais para as duas características utilizando as informações em nível de famílias e de indivíduos.

Como pode ser verificado na Tabela 13, o uso do  $IC_1$  proporcionou um ganho predito de 4,17% para PROD e de 2,00% para PTN. Já o  $IC_2$ , que priorizou

Tabela 13 - Estimativas das herdabilidades restritas em nível de família, dentro de família e total, da média original, da média dos selecionados, do diferencial de seleção, do ganho percentual esperado pela seleção e das eficiências de seleção, utilizando as estratégias de seleção entre (E) e dentro (D) de famílias, massal (M), combinada (C) e índice combinado (IC)

Características sob seleção	Estratégia e N <sup>os</sup> de selecionados	PROD(*)					PTN(*)				
		$\bar{X}_o$	$\bar{X}_s$	DS	$h^2_R$	GS%	$\bar{X}_o$	$\bar{X}_s$	DS	$h^2_R$	GS%
PROD	Entre (21)	<b>20,54</b>	<b>23,35</b>	<b>2,81</b>	<b>0,381</b>	<b>5,21</b>	<u>41,04</u>	<u>40,74</u>	<u>-0,20</u>	<u>0,607</u>	<u>-0,30</u>
	Dentro (84)	<b>23,35</b>	<b>26,69</b>	<b>3,34</b>	<b>0,074</b>	<b>1,06</b>	<u>40,74</u>	<u>41,86</u>	<u>1,19</u>	<u>0,156</u>	<u>0,46</u>
	Total (ED)					<b>6,27</b>					<u>0,16</u>
PTN	Entre (21)	<u>20,54</u>	<u>20,23</u>	<u>-0,31</u>	<u>0,381</u>	<u>-0,58</u>	<b>41,04</b>	<b>42,35</b>	<b>1,30</b>	<b>0,607</b>	<b>1,93</b>
	Dentro (84)	<u>20,23</u>	<u>21,70</u>	<u>1,47</u>	<u>0,074</u>	<u>0,54</u>	<b>42,35</b>	<b>43,42</b>	<b>1,07</b>	<b>0,156</b>	<b>0,39</b>
	Total (ED)					<u>-0,04</u>					<b>2,32</b>
PROD	Massal (84)	<b>20,54</b>	<b>26,34</b>	<b>5,80</b>	<b>0,238</b>	<b>6,75</b>	<u>41,04</u>	<u>41,32</u>	<u>0,28</u>	<u>0,43</u>	<u>0,30</u>
	PTN	<u>20,54</u>	<u>21,50</u>	<u>0,96</u>	<u>0,238</u>	<u>1,11</u>	<b>41,04</b>	<b>43,57</b>	<b>2,53</b>	<b>0,43</b>	<b>2,6</b>
PROD	Combinada (84)	<b>20,54</b>		<b>1,04</b>		<b>5,11</b>	<u>41,04</u>	<u>41,07</u>	<u>0,03</u>	<u>0,43</u>	<u>0,03</u>
PTN	Combinada (84)	<u>20,54</u>	<u>20,25</u>	<u>-0,29</u>	<u>0,238</u>	<u>-0,34</u>	<b>41,04</b>		<b>0,72</b>		<b>1,75</b>
PROD e PTN	IC <sub>1</sub> (84)	<b>20,54</b>	<b>24,13</b>	<b>3,60</b>	<b>0,238</b>	<b>4,17</b>	<b>41,04</b>	<b>42,95</b>	<b>1,91</b>	<b>0,43</b>	<b>2,00</b>
PROD e PTN	IC <sub>2</sub> (84)	<b>20,54</b>	<b>22,73</b>	<b>2,19</b>	<b>0,238</b>	<b>2,54</b>	<b>41,04</b>	<b>43,30</b>	<b>2,26</b>	<b>0,43</b>	<b>2,37</b>
EF. SED/SE						<b>1,20</b>					<b>1,20</b>
EF. SM/SED						<b>1,08</b>					<b>1,12</b>
EF. SM/IC						<b>1,32</b>					<b>1,49</b>
EF. SM/IC <sub>1</sub>						<b>1,62</b>					<b>1,30</b>
EF. SM/IC <sub>2</sub>						<b>2,65</b>					<b>1,10</b>

( $\bar{X}_o$ ) média original; ( $\bar{X}_s$ ) média dos selecionados; DS diferencial de seleção;  $h^2_R$  herdabilidade restrita; (GS%) ganhos percentuais.

EF: eficiência de seleção.

IC<sub>1</sub>: pesos genotípicos = 1; 2,5, para PROD e PTN respectivamente.

IC<sub>2</sub>: pesos genotípicos = 1; 5, para PROD e PTN respectivamente.

(\*) Em negrito é o ganho direto, sublinhado é o ganho indireto.

teor de proteína, pois utilizou como pesos econômicos 1 para PROD e 5 para PTN, proporcionou um ganho de 2,54% para produção de grãos e de 2,37% para PTN. Assim, o IC<sub>2</sub> foi o indicado para selecionar indivíduos para o próximo ciclo de seleção, pois, além de aumentar significativamente o teor de proteína da população selecionada que é o objetivo principal deste trabalho, também proporcionou ganhos moderados para produção de grãos. Os ganhos para PTN são compatíveis com os ganhos obtidos por Miller e Fehr (1979), Brim e Burton (1979), Openshaw e Hadley (1984), Scott e Kephart (1997) e Wilcox (1998) que variaram desde 0,5% até 4,0% por ciclo de seleção.

### **5.5. Avaliação do desempenho e predição de ganhos em progênies RC<sub>1</sub>F<sub>4</sub>:6 em experimento realizado no ano agrícola 2006/2007**

Este experimento foi conduzido em delineamento em blocos casualizados conforme descrito em Material e Métodos, na estação de pesquisa da Coopadap no ano agrícola (2006/2007). O material genético utilizado foram 117 famílias RC<sub>1</sub>F<sub>4:6</sub> originadas de plantas RC<sub>1</sub>F<sub>4</sub> selecionadas na terceira etapa do primeiro ciclo de seleção deste programa. Este ensaio objetivou prever ganhos genéticos para as características Produtividade de Grãos (PG), Altura da Planta na Maturação (APM), Número de Dias para Maturação (NDM), Teor de Proteínas nos Grãos (PTN) e Teor de Óleo nos Grãos (OL), utilizando diferentes critérios de seleção baseados em índices e priorizando o aumento do teor de proteína nos grãos, aumento da produtividade e redução do ciclo da planta e minimizando a redução do teor de óleo. A seguir são apresentadas as estimativas dos parâmetros genéticos dessa população e a predição dos ganhos utilizando os diferentes critérios de seleção.

#### **5.5.1. Análise de variância**

Os resultados da análise de variância para as características Produtividade de Grãos (PG), Altura da Planta na Maturação (APM), Número de Dias para Maturação (NDM) Teor de Proteínas nos Grãos (PTN) e Teor de Óleo nos Grãos (OL) utilizando o delineamento em blocos casualizados, são apresentados na Tabela 14, em que a aplicação do teste F revelou significância a



1% de probabilidade para as características APM, NDM, PTN e OL e a 5% para a característica PG. A existência de variância genética significativa sugere a possibilidade de obtenção de ganhos pela aplicação de seleção para as cinco características.

Tabela 14 - Resumo da análise de variância para as características PG, APM, NDM, PTN e OL de progênies RC<sub>1</sub>F<sub>4:6</sub> de soja avaliadas no experimento em blocos casualizados no agrícola 2006/2007

FV	QM					
	GL	PG	APM	MDN	PTN	OL
Blocos	1	136306,72	103,57	904,60	5,994	0,201
Tratamentos	118	97055,44**	104,83**	94,58**	9,455**	3,920**
Progênies RC <sub>1</sub> F <sub>4:6</sub>	116	75478,65*	86,43**	80,28**	7,492**	2,834**
Testemunhas	1	1856815,02**	1369,00**	1296,00**	104,336**	17,918**
Progênies vs Testemunhas	1	840204,25**	974,78**	551,49**	142,331**	115,871**
Resíduo	118	50583,03	28,61	14,41	2,378	1,055
Média Geral		1933,12	79,98	139,14	39,29	23,26
Média das Prog.		1940,89	80,24	139,34	39,19	23,35
Média das Test.		1478,68	64,50	127,5	45,21	17,92
Média da Monarca		2160,00	83,00	145,50	40,10	20,04
Média de BARC-8		797,36	46,00	109,5	50,32	15,80
C.V. %		11,63	6,69	2,73	3,92	4,42

(\* e \*\*) significativo a 5 % e 1% de probabilidade pelo teste F, respectivamente.

Os coeficientes de variação experimental obtidos para todas as características apresentaram valores satisfatórios próximos aos encontrados por Prado (1994), Carvalho *et al.* (2002), Soares *et al.* (2004) e Miranda (2006) que utilizaram o mesmo tipo de delineamento aqui empregado, evidenciando uma boa precisão para estimação dos parâmetros genéticos.

Nesta população, foram estimados a variância genética total, os coeficientes de herdabilidade no sentido restrito entre médias de famílias, o coeficiente de variação genotípico e a razão entre o coeficiente de variação genotípico e o coeficiente de variação ambiental. Os resultados encontram-se descritos na Tabela 15.

A menor estimativa de herdabilidade foi para a característica produtividade de grãos, com valor aproximado de 33%. Para as outras características, as estimativas de herdabilidade variaram de 62,26 a 82,05%, podendo ser consideradas de médias a altas, estando próximas aos valores encontrados por Handon e Weber (1961), Kwon e Torrie (1964), Shorer *et al.*

(1976) e Mckendry *et al.* (1985), indicando que o processo seletivo com base em tais características poderá ser realizado com grande eficiência. Juntamente com as estimativas de herdabilidades, o conhecimento da relação entre o coeficiente de variação genético e o coeficiente de variação experimental proporcionam um melhor entendimento da estrutura genética da população por mostrarem a quantidade de variação existente entre famílias. Como pode ser observado, a relação  $CV_g/CV_e$  foi inferior à unidade para PG e OL e para PTN foi próximo da unidade. Esses resultados refletem uma maior dificuldade em obter ganhos genéticos, principalmente para PG, pois a maior proporção da variação do fenótipo está sendo explicada por causas ambientais.

Tabela 15 - Estimativa de parâmetros genéticos para as características PG, APM, NDM, PTN e OL avaliadas em progênies  $RC_1F_{4:6}$  de soja.

Parâmetros	Características				
	PG	APM	NDM	PTN	OL
$\hat{\sigma}_f^2$	37739,32	43,21	40,14	3,75	1,42
$\hat{\sigma}_g^2$	12447,81	28,90	32,94	2,56	0,89
$h^2$	32,98	66,90	82,05	68,26	62,77
$CV_g\%$	5,75	6,70	4,12	4,08	4,04
$CV_g\%/CV_e\%$	0,496	1,00	1,51	1,04	0,92

$\hat{\sigma}_f^2$ : estimativa da variância fenotípica entre famílias;  $\hat{\sigma}_g^2$ : estimativa da variância genética entre famílias  $h^2$ : estimativa da herdabilidade no sentido restrito;  $CV_g$  - coeficiente de variação genética entre famílias;  $CV_g/CV_e$  - relação entre coeficiente de variação genética e coeficiente de variação ambiental entre famílias.

São apresentadas na Tabela 16 as estimativas das correlações fenotípica, genotípica e ambiental entre as características PG, APM, NDM, PTN e OL. Uma correlação tanto fenotípica como genotípica altamente negativa foi encontrada entre PTN e OL, de magnitude acima da maioria dos trabalhos já citados em itens anteriores. Esta altíssima correlação negativa pode ter sido reflexo do processo seletivo a que esta população foi submetida. A presença de correlações genotípicas positivas de baixa magnitude entre PG e OL e a ausência de correlações genotípicas significativas entre PG e PTN indicam que a seleção para PG favorecerá indiretamente a característica OL, mas, quando for realizada

seleção para PTN, a característica PG poderá ser ou não favorecida. Como já comentado, a ausência de correlação negativa significativa entre PG e PTN deve-se ao fato de esta população ter se originado de progenitores que apresentavam diferenças significativas entre teores de proteína, mas que apresentavam desempenho semelhante para produção de grão. A existência de uma correlação positiva entre PTN e NDM desfavoreceu o objetivo deste programa que era o de aumentar o teor de proteína e reduzir o ciclo das linhagens selecionadas, o mesmo tendo ocorrido entre PG e NDM, em que a seleção para um menor ciclo reduziu a produtividade. No entanto, a fim de minimizar estes efeitos contrários e de favorecer ganhos distribuídos de acordo com os objetivos do programa, principalmente para teor de proteína e produtividade, é que foi empregada, além da seleção direta e indireta, a seleção simultânea baseada nos índices de seleção de Smith e Hazel (SH), Kempthorne e Nordskog (KN), Pesek e Baker (PB) e Pesek e Baker generalizado por Tai (PB-Tai).

Tabela 16 - Coeficientes de correlação fenotípica ( $r_F$ ), genotípica ( $r_G$ ) e ambiental ( $r_E$ ) entre as características PG, APM, NDM, PTN e OL avaliadas em progênies RC<sub>1</sub>F<sub>4:6</sub> de soja.

Correlações	$r_F$	$r_G$	$r_E$
PG × APM	0,049	0,008	0,095
PG × NDM	0,128	0,300**	-0,081
PG × PTN	0,014	-0,089	0,121
PG × OL	0,085	0,212*	-0,007
APM × NDM	0,368**	0,465**	0,098
APM × PTN	0,189*	0,259**	0,044
APM × OL	-0,290**	-0,342**	-0,188
NDM × PTN	0,041	0,059	-0,010
NDM × OL	-0,098	0,104	-0,116
PTN × OL	-0,787**	-0,953**	-0,456

(\* e \*\*) significativo a 5% e 1% de probabilidade pelo teste t, respectivamente.

### 5.5.2. Predição de ganhos

As características consideradas principais neste experimento foram PG e PTN, sendo que os pesos econômicos e os ganhos desejados variaram em função dessas características e dos seguintes objetivos: a) seleção de 20 famílias que maximizassem ganhos para produtividade de grãos, mas com menor redução possível para teor protéico e b) seleção de 20 famílias que maximizassem ganhos

para teor de proteína, mas com menor redução possível de produtividade de grãos.

Estão apresentados na Tabela 17 os resultados dos ganhos preditos diretos e indiretos quando selecionados para PG e PTN, dos ganhos baseados nos índices de seleção e dos ganhos em relação ao progenitor Monarca para as características: Produtividade de Grãos (PG); Número de Dias para Maturação (NDM); Altura da Planta na Maturação (APM); Teor de Proteínas nos Grãos (PTN) e Teor de Óleo nos Grãos (OL).

Os pesos econômicos, os ganhos desejados, a classificação da característica como principal ou secundária e a restrição foram distribuídos da seguinte maneira: para o índice clássico de Smith e Hazel (SH), consideraram-se apenas PG e PTN como características principais e para os pesos econômicos a relação foi de 1:10; para o índice de Pesek e Baker (PB) também se consideraram apenas PG e PTN como características principais, sendo os ganhos desejados baseados no desvio padrão destas características; e para o índice de Pesek e Baker generalizado por Tai (PB-Tai), consideraram-se PG, NDM e PTN como características principais e OL secundária com ganho nulo em duas situações de ganhos desejados, sendo para PB-Tai-1 (111,56; -5,7; 1,59;0) e PB-Tai-2 (3,-3,3,0). Já para o índice restrito de Kempthorne e Nordskog (KN) consideraram-se PG e PTN como características principais e como pesos econômicos a relação de 1:5, mas restringindo o ganho para PG como nulo ao adotar a COV (escores do índice, valores genéticos de OL) = 0.

A seleção direta para produtividade de grãos (PG) promoveu um ganho de 4,96% para esta característica em relação à população original e de 3,36% em relação ao progenitor Monarca, mas provocou indiretamente um ganho negativo de 0,53% para a característica teor de proteína (PTN) nesta mesma população. Já a seleção direta para teor de proteína promoveu um ganho de 4,68% para esta característica em relação à população original, mas provocou indiretamente uma redução de 1,17% em relação à população original e de 13,33% em relação ao progenitor Monarca para a característica PG. Entretanto, apesar de a seleção direta ser eficiente para promover ganhos máximos para a característica principal, ela é inadequada quando se deseja manter ou melhorar simultaneamente outras características, que foi o propósito deste trabalho. Portanto, foram empregados os critérios de seleção baseados nos índices.

Verifica-se que, quando foram empregados os índices SH-1, PB-1 e PB-Tai-1, houve ganhos significativos para PG e ainda promoção de um pequeno ganho para teor de proteína, com superioridade para o índice PB-1. O interessante é destacar o uso do índice PB-Tai-1 que, além de promover ganhos moderados para PG, ficando a média dos selecionados acima da variedade Monarca, também reduziu o ciclo (NDM) em -2,56%, o que é muito desejável, pois os produtores de soja têm dado preferência às variedades de ciclo mais precoce. Este foi, portanto, o índice escolhido para selecionar as progênes para o próximo ciclo de seleção, pois essas famílias selecionadas com base nesse índice apresentarão produtividade média superior à variedade padrão Monarca, ciclo mais precoce e possível resistência ao glifosato.

O emprego dos índices PB-Tai-2 e KN1 que favoreciam mais a característica PTN, proporcionou ganhos preditos para PTN de 4,29 e 4,59%, respectivamente, e ganhos em relação à variedade Monarca de 3,93 e 4,31% para o teor protéico, além da manutenção da média da característica PG na população selecionada. Quando se compara o ganho em relação à variedade Monarca, ocorre uma redução de 9,31 e 10,19% na média dos selecionados, o que deve ser analisado com certa preocupação, pois no próximo ciclo de seleção deverá ser selecionado um menor número de linhagens, mais produtivas e de maior teor protéico. A vantagem da seleção de progênes utilizando o índice PB-Tai-2 em relação ao KN1 está no fato de reduzir o ciclo da população a ser selecionada, o que minimizaria a perda em produtividade, pois variedades mais precoces são em geral menos produtivas. Este será o segundo índice a ser utilizado neste trabalho para selecionar progênes que apresentarão teor protéico superior e ciclo mais precoce do que a variedade padrão Monarca, além da resistência ao glifosato e que serão utilizadas para o próximo ciclo de seleção. As progênes selecionadas pelos índices PB-Tai-1 e PB-Tai-2 encontram-se na Tabela 18.

103

Tablela 17 – Estimativas da média das progênes selecionadas ( $\bar{X}_s$ ), dos ganhos percentuais (GS%) e do ganho em relação ao progenitor Monarca ( $GS_M^%$ ) para as características PG, APM, NDM, PTN e OL avaliadas em progênes RC<sub>1</sub>F<sub>4:6</sub> por meio da seleção direta (SD) e indireta para PG e PTN e por meio da seleção baseada nos índices de Smith e Hazel (SH), Kempthorne e Nordskog (KN), Pesek e Baker (PB) e Pesek e Baker generalizado por Tai (PB-Tai)

Critério de seleção*	$\bar{X}_s$					GS%					$GS_M^%$	
	PG	APM	NDM	PTN	OL	PG	APM	NDM	PTN	OL	PG	PTN
<b>SD-PG</b>	<b>2232,65</b>	79,75	139,35	38,89	23,52	<b>4,96</b>	-0,41	0,0	-0,53	0,46	<b>3,36</b>	<b>-3,02</b>
<b>SD-PTN</b>	1872,16	81,13	138,00	<b>41,88</b>	21,95	-1,17	0,73	-0,79	<b>4,68</b>	-3,75	<b>-13,33</b>	<b>4,43</b>
<b>SH<sub>1</sub></b>	2232,00	80,25	139,85	39,34	23,40	<b>4,95</b>	0,01	0,30	0,25	0,13	3,33	-1,90
<b>PB-1</b>	2217,05	81,13	141,13	39,96	23,06	<b>4,69</b>	0,73	1,05	1,34	-0,79	2,64	-0,36
<b>PB-Tai-1</b>	2173,03	77,13	135,00	39,61	23,46	3,94	-2,6	<b>-2,56</b>	0,72	0,29	0,60	-1,23
<b>PB-Tai-2</b>	1958,65	79,63	136,13	41,66	22,05	0,30	-0,52	<b>-1,89</b>	<b>4,29</b>	-3,5	-9,31	3,93
<b>KN<sub>1</sub></b>	1939,80	81,13	139,63	41,83	21,95	-0,02	0,73	0,17	<b>4,59</b>	-3,76	-10,19	4,31

Número de progênes selecionadas: 20.

(\*) Em negrito ganho direto.

Sendo: SH<sub>1</sub>(1,10); PB-1 (111,56; 1,59); PB-Tai-1 (111,56; -5,7; 1,59; 0) e OL secundária; PB-Tai-2 (3,-3,3,0) e OL secundária; KN<sub>1</sub> (0,5).

Tabela 18 - Progênes selecionadas e valores médios das características PG, APM, NDM, PTN e OL pelos índices PB Tai 1 e PB Tai 2

	PB Tai 1					PB Tai 2						
	Progênie	PG	APM	NDM	PTN	OL	Progênie	PG	APM	NDM	PTN	OL
104	107	2672,50	67,50	136	40,29	22,82	5	2479,50	90,00	140	43,31	22,53
	5	2479,50	90,00	140	43,31	22,53	107	2672,50	67,50	136	40,29	22,82
	35	2133,00	75,00	128	38,91	24,05	90	1913,50	72,50	121	41,19	22,10
	90	1913,50	72,50	121	41,19	22,10	94	1869,00	87,50	138	42,98	21,81
	40	2292,00	77,50	137	38,39	24,59	74	1818,00	85,00	135	42,46	21,46
	44	2192,50	85,00	139	40,71	23,40	78	1585,00	77,50	141	43,98	21,33
	37	2160,00	70,00	134	39,24	24,04	75	1827,00	77,50	133	41,61	21,15
	3	2326,00	80,00	136	37,47	22,78	104	2046,00	82,50	137	41,17	21,64
	86	2175,00	82,50	136	39,65	23,69	76	1857,50	77,50	132	41,38	22,26
	1	2232,00	82,50	136	38,50	23,93	15	1847,00	82,50	134	41,42	21,63
	110	2289,50	80,00	141	38,52	23,67	73	1829,00	80,00	130	41,19	22,59
	39	2079,00	72,50	136	39,34	24,45	14	1793,50	82,50	136	41,80	21,69
	89	1936,00	77,50	123	37,94	23,38	19	2150,50	65,00	143	41,24	22,35
	19	2150,50	65,00	143	41,24	22,35	91	1944,00	82,50	145	42,01	21,34
	72	1882,50	77,50	128	40,64	23,34	72	1882,50	77,50	128	40,64	23,34
	83	2168,50	80,00	140	39,38	23,16	44	2192,50	85,00	139	40,71	23,40
	104	2046,00	82,50	137	41,17	21,64	4	2123,50	87,50	140	40,42	22,77
	42	2027,50	75,00	135	40,20	23,03	93	1779,50	82,50	138	41,72	21,93
	46	2202,50	82,50	143	39,11	23,82	79	1647,50	80,00	141	42,25	21,06
							20	1915,50	70,00	139	41,39	22,83

## 5.6. Avaliação Genética e Estatística da População de Mapeamento

### 5.6.1. Testes de normalidade

A característica teor de proteína da população de NILs composta de 168 isolinhas apresentou distribuição aproximadamente normal ( $P < 0,01$ ), como pode ser observado na Figura 4 por apresentar estimativas não significativas de curtose e valor D do teste de Lilliefors, apesar do teste de assimetria ter sido significativo. Este fato indica a existência de variação contínua e a possibilidade de que vários genes estejam envolvidos no controle desta característica nesta população, possibilitando o emprego de métodos de estimação de QTLs e testes estatísticos de significância que pressupõem uma distribuição normal dos dados.

Tabela 19 – Estimativas dos valores de curtose, simetria e D para a distribuição do teor de proteína do grão em uma população de NILs (isolinhas RC<sub>4</sub>F<sub>6</sub>) em soja.

	Curtose	Simetria	D <sup>(1)</sup>
Estimativa	0,3153 <sup>ns</sup>	0,6124*	0,0742 <sup>ns</sup>
Variância	0,1388	0,0351	
Valor de t	0,4106	3,2689	
Probabilidade (%)	68,445	0,1288	>1,00%<5,00%

(1) Estatística D para o Teste de Lilliefors.

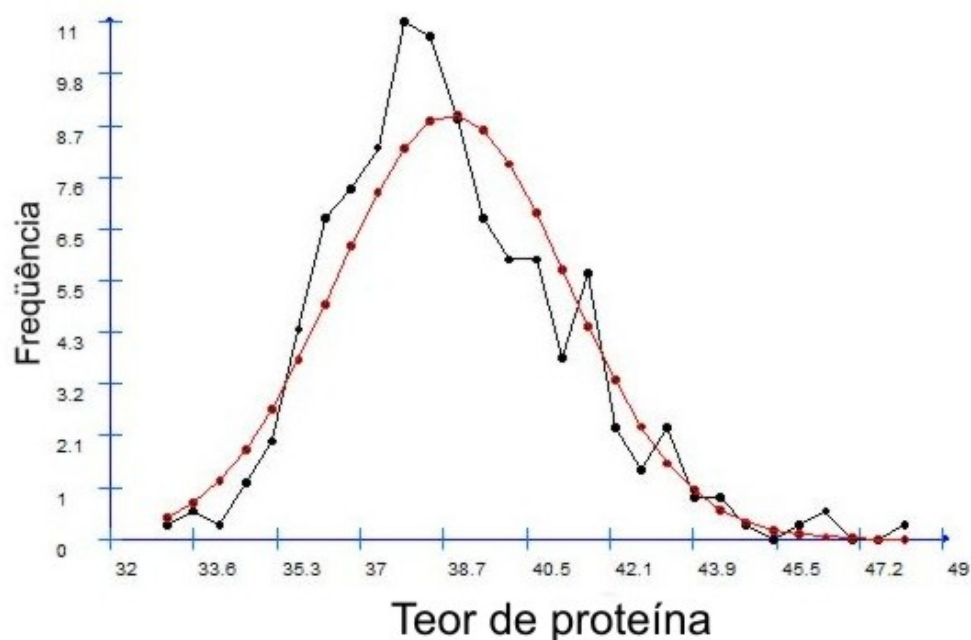
(\*) Significativo a 1% de probabilidade pelo teste t; (ns) – Não significativo pelo teste t.

### 5.6.2. Análise de variância

O resultado da análise de variância para a característica teor de proteína no grão utilizando o delineamento em blocos casualizados é apresentado na Tabela 20, em que a aplicação do teste F a 1% de significância revelou existência de variância genética significativa nesta população, evidenciando a existência de variabilidade genética entre as NILs (isolinhas RC<sub>4</sub>F<sub>6</sub>), possibilitando o estudo da associação Marcador-QTL.

O coeficiente de variação experimental para a característica teor de proteína apresentou valor satisfatório próximo aos encontrados por Prado (1994), Chung *et al.* (2003), Soares *et al.* 2004 e Miranda (2006), que utilizaram o mesmo tipo de delineamento aqui empregado, evidenciando uma boa precisão experimental para estimação dos parâmetros genéticos.





**Figura 4** – Distribuição dos valores do teor de proteína no grão (%) em isolinhas RC<sub>4</sub>F<sub>6</sub> de soja em comparação com distribuição normal de uma variável com média 38,97 e variância 3,91.

São também apresentados os parâmetros genéticos: médias, variâncias fenotípica, genética e ambiental, coeficiente de herdabilidade no sentido restrito entre médias das NILs, coeficiente de variação genético e a razão entre o coeficiente de variação genético e o coeficiente de variação experimental. Apesar de a população apresentar variabilidade genética significativa para teor de proteína, pré-requisito básico para o estudo de mapeamento de QTLs. As estimativas moderadas de herdabilidade e o valor menor do que a unidade para a relação  $CV_g/CV_e$  refletem uma maior dificuldade para diferenciar QTLs de pequeno efeito em função de que grande parte da variação fenotípica existente entre as NILs deve-se a causas ambientais. Como não se pode aumentar a variabilidade genética desta população, recomenda-se então em uma próxima avaliação o aumento da precisão experimental, pelo aumento do número de repetições e do aumento do tamanho da parcela.

Tabela 20 – Resumo da análise de variância para a característica teor de proteína do grão em uma população de NILs de soja.

FV	GL	SQ	QM
BLOCOS	1	2,098	2,098
Tratamentos	169	1602,230	9,481**
Genótipos	167	1302,640	7,800**
Testemunha	1	183,200	183,200**
Genótipos vs Testemunha	1	116,410	116,410**
Resíduo	169	579,820	3,431
-----			
Média geral		39,030	
Média genótipos		38,970	
Média testemunhas		44,390	
CVe (%)		4,750	
Variância fenotípica		3,900	
Variância ambiental		1,715	
Variância genotípica		2,185	
Herdabilidade		56,020	
Coeficiente de variação genético (%)		3,793	
Razão CVg/CVe		0,798	

(\*\*) Significativos a 1% de probabilidade pelo teste F.

### 5.6.3. Detecção de polimorfismos, análise de segregação e formação dos grupos de ligação.

Foram testados 380 pares de *primers* de microssatélites publicados por Song *et al.* (2004) para inicialmente detectar polimorfismos entre o genótipo doador BARC-8 e o genótipo recorrente Monarca. Desses, 23 não amplificaram, 262 apresentaram fragmentos monomórficos e 95 polimórficos. Os *primers* de microssatélites polimórficos foram então testados em quatro *bulks* formados da reunião de DNA de cada isolinha RC<sub>4</sub>F<sub>6</sub> da mesma subpopulação, ou seja, que tiveram a mesma planta RC<sub>4</sub>F<sub>1</sub> como ancestral. Esse procedimento foi necessário, pois poderia haver heterozigose em algum loco do *primer* de microssatélite na planta RC<sub>3</sub>F<sub>5</sub> que deu origem às quatro subpopulações. Caso ocorresse em alguma dessas populações e se não fossem mantidas separadas, não seria possível identificar essa diferença no padrão de segregação, como de fato ocorreu com alguns *primers* em algumas subpopulações, que neste caso não foram amplificados nesta subpopulação específica e considerados como dados perdidos para fins de análise. Dos 95 *primers* testados, apenas 20 foram polimórficos nos *bulks*, sendo sua segregação depois avaliada em toda a população de NILs.

A segregação de cada marcador foi testada por meio do teste qui-quadrado ( $P < 0,05$ ), utilizando o critério de Proteção Bonferroni. Apesar de ocorrerem falhas de amplificação, nenhum dos 20 marcadores apresentou distorção da proporção 1:1 (Tabela 21).

A obtenção dos grupos de ligação (GL) foi um pré-requisito para a análise de QTLs pelo método de mapeamento por intervalo simples. Todos os 20 marcadores foram empregados para o cálculo da distância genética em cM pela função de Kosambi. O agrupamento dos marcadores nos grupos de ligação foi verificado, utilizando-se os critérios de  $LOD_{\text{mínimo}} = 3$  e  $r_{\text{máximo}} = 0,30$ . O LOD é considerado um teste de significância para testar a hipótese de ligação entre dois locos. O valor elevado de LOD permitiu que fossem identificados apenas grupos de ligação com intervalo entre os marcadores inferior a 30 cM, o que torna os grupos de ligação mapeados úteis na identificação de QTLs, pois, segundo Lee (1995), o intervalo útil para a identificação preliminar de QTLs está entre 15 a 20 cM ou valores um pouco maiores.

Foram obtidos cinco grupos de ligação definidos pelo agrupamento dos 20 marcadores microssatélites, como pode ser observado na Figura 5, ficando nove marcadores não ligados. Os grupos obtidos representaram segmentos genômicos dos grupos de ligação B2, C2, E, F. Para três grupos de ligação (B2, E e F), houve coincidência dos marcadores com o mapa consenso da soja (SONG *et al.*, 2004). No GL C2, o marcador Satt422 que pertence ao GL H foi incluído e no outro grupo, ficou indefinido, pois os marcadores Satt436 que pertencem ao GL D1a e os Satt468 que pertencem ao GL A2 ficaram ligados.

Esse baixo padrão de polimorfismo já era esperado por se tratar de isolinhas derivadas de retrocruzamentos, em que há uma recuperação do genoma do progenitor recorrente, diminuindo assim a variabilidade para outros genes que não estejam relacionados com o teor protéico. Devido a isso, apenas cinco grupos de ligação foram obtidos e ainda ficaram pouco saturados, representando apenas 38,62 cM de um total de 2523,6 cM do mapa da soja (SONG *et al.*, 2004). Essa é a primeira etapa de um processo que visa ao mapeamento de QTLs que controlam a característica analisada nesta população de NILs, ficando para trabalhos futuros uma maior saturação quando se dispuser de mais marcadores polimórficos SSR ou SNP.

Tabela 21 - Marcadores polimórficos, Grupo de Ligação (GL), número de isolinhas com o mesmo do genótipo Monarca (0), número de isolinhas com o mesmo genótipo da Barc-8 (2), dados perdidos (9), número de isolinhas heterozigotas (1), Hipótese, valor do teste de Qui-quadrado e Probabilidade para os marcadores.

Marcador	GL	Classes				Hipótese	Qui-quadrado	Probabilidade
		0	1	2	9*			
Satt294	C1	57		59	52	1:1	0,034	85,2684 <sup>ns</sup>
Satt193	F	25	3	33	107	1:1	1,103	29,3511 <sup>ns</sup>
Satt239	I	85	3	76	4	1:1	0,503	47,8139 <sup>ns</sup>
Satt147	D1a	80	4	80	4	1:1	0,000	100,00 <sup>ns</sup>
Satt070	B2	59		58	51	1:1	0,009	92,6340 <sup>ns</sup>
Satt309	G	82		82	4	1:1	0,000	100,0000 <sup>ns</sup>
Satt422	H	84		77	7	1:1	0,304	58,1169 <sup>ns</sup>
Sat_112	E	93		74	1	1:1	2,162	14,1490 <sup>ns</sup>
Satt146	F	93		69	6	1:1	3,556	5,9346 <sup>ns</sup>
Satt520	C2	90		78	0	1:1	0,857	35,4539 <sup>ns</sup>
Satt281	C2	95	1	72	0	1:1	3,168	7,5110 <sup>ns</sup>
Satt384	E	90	1	76	1	1:1	1,181	27,7209 <sup>ns</sup>
Satt545	A1	42	3	40	83	1:1	0,049	82,5199 <sup>ns</sup>
Satt474	B2	31	1	33	103	1:1	0,063	80,2587 <sup>ns</sup>
Satt436	D1a	82		62	24	1:1	2,778	9,5581 <sup>ns</sup>
Satt338	C1	41	1	50	76	1:1	0,890	34,5448 <sup>ns</sup>
Satt556	B2	38		52	78	1:1	2,178	14,0017 <sup>ns</sup>
Satt193	F	22		13	133	1:1	2,314	12,8190 <sup>ns</sup>
Satt276	A1	47		55	66	1:1	0,627	42,8292 <sup>ns</sup>
Satt468	A2	84	1	73	10	1:1	0,771	38,0000 <sup>ns</sup>

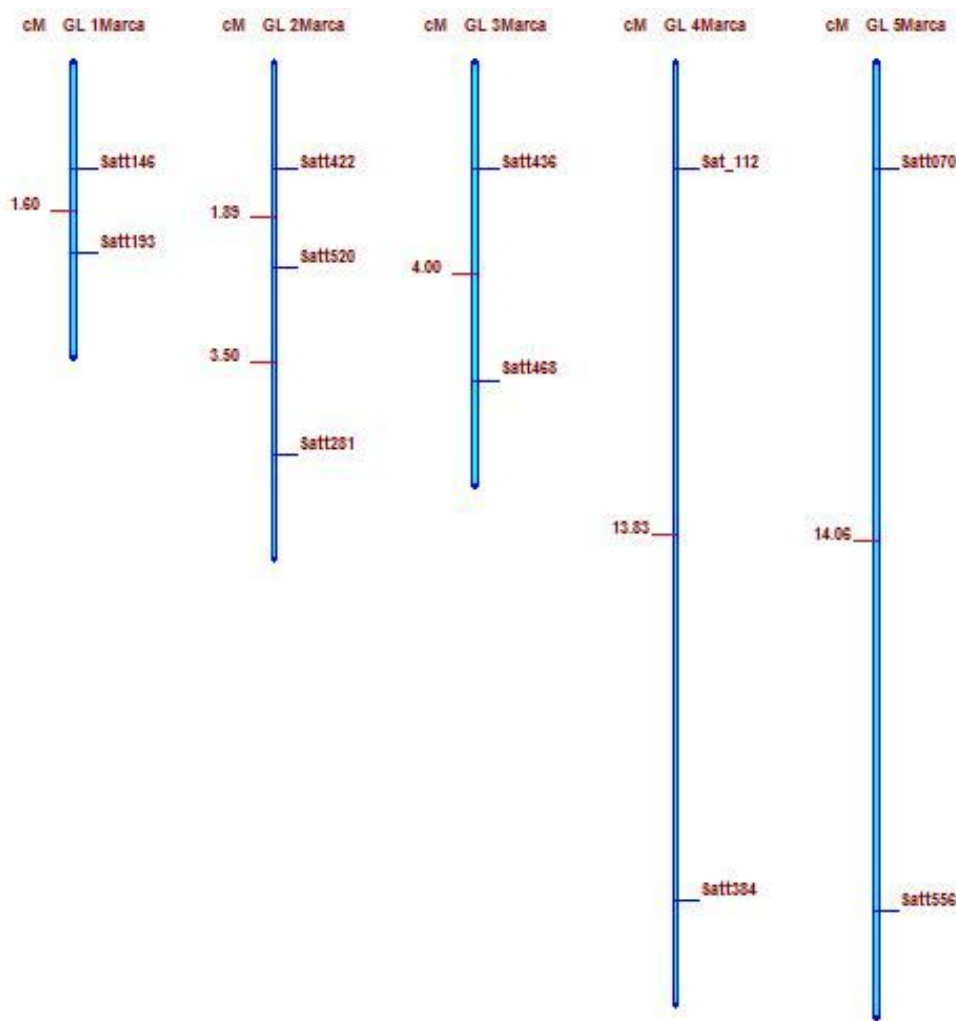
Nível crítico: 0,258%

Número total de testes: 20

Critério: Proteção Bonferroni

Número de distorções: 0

(\*) dado perdido



**Figura 5.** Grupos de ligação obtidos utilizando uma população de 168 isolinhas  $RC_4F_6$  derivadas da variedade Monarca. Os grupos de ligação foram obtidos adotando-se os critérios de  $LOD_{mínimo} = 3$  e  $r_{máximo} = 0,30$ . Os valores da esquerda representam as distâncias entre as marcas baseados na função de Kosambi e à direita estão os nomes dos marcadores.

#### 5.6.4. Identificação de QTLs

A identificação de QTLs foi realizada inicialmente pela análise de associação utilizando modelos lineares simples pelo método da ANOVA e por meio de regressão linear. Esses métodos não necessitam das informações dos mapas de ligação, e a distribuição dos valores fenotípicos é analisada para cada marcador separadamente. Portanto, por utilizar apenas diferenças entre as médias dos marcadores, o poder estatístico que este método apresenta é baixo, não sendo possível estimar a magnitude do efeito do QTL, nem sua posição no genoma (LYNCH e WALSH, 1998). Mas sabe-se que as associações detectadas

por este método podem ser atribuídas a QTLs de grande efeito e relativamente distantes do marcador, como também a QTLs de pequeno efeito, porém próximos ao marcador. Utilizando essa metodologia, foram encontrados dois marcadores (Satt 384 e Satt 239) associados a QTLs que explicavam 7,42% e 19,04% da variação fenotípica do teor protéico, determinado pelo coeficiente de determinação ( $R^2$ ). Esses resultados encontram-se na Tabela 22.

Tabela 22 – Análise de associação de marcas moleculares com teor protéico avaliadas em uma população de NILs pela ANOVA e Regressão Linear

Marca	Proteína Total				
	Prob, F	Média (2)	Média (0)	$R^2_{\%}$	Grupo de ligação
Satt384	0,0004**	39,58a	38,50 b	7,42	E
Satt239	0,0000**	39,82a	38,09 b	19,04	NL

\*\* Significativo a 1% pelo teste F; NL – Não ligado (pertence ao GL I no mapa integrado da soja).

Numa segunda fase, as ligações entre marcadores e QTLs foram avaliadas por meio de um modelo de regressão linear múltipla utilizando o procedimento de eliminação *stepwise* para seleção dos marcadores associados a QTLs. Essas análises foram efetuadas por meio do programa GQMOL, utilizando-se para escolha inicial dos marcadores a serem incluídos no modelo de regressão múltipla aqueles que se mostraram associadas à característica teor de proteína com um coeficiente de correlação parcial superior a 0,10 e tendo sido adotados como critério de entrada e permanência de marcadores no modelo um valor de probabilidade da estatística F inferior a 0,10. Novamente, apenas os marcadores Satt384 e Satt239 permaneceram no modelo, indicando associações com QTLs que explicaram conjuntamente 26,01% da variação do fenótipo determinado pelo  $R^2$  conjunto, como apresentado na Tabela 23.

Por fim, utilizou-se o procedimento de mapeamento por intervalo simples baseado em análise de regressão para identificar QTLs nos grupos de ligação. Nessa análise, o valor crítico para a razão de verossimilhança e o LOD em cada grupo de ligação foram determinados por meio de 1000 permutações dos dados para um nível de 5% significância. O mapeamento por intervalo permitiu estimar o efeito aditivo do QTL na expressão do caráter e determinar sua distância em relação ao marcador no grupo de ligação quatro (E), ao qual o marcador Satt384

pertence. As estimativas do valor aditivo, do coeficiente de determinação do QTL correspondente ao pico de maior significância estatística do QTL e a sua possível posição no grupo de ligação encontram-se na Tabela 24.

Tabela 23 - Análise de Variância considerando o modelo final de regressão múltipla utilizando o método de eliminação *stepwise* para seleção de marcadores associados às QTLs em uma população de NILs.

FV	GL	SQ	QM	F	Probabilidade (%)
Regressão	2	161,18599	80,59299	27,95394	0,0000
Satt239	1	118,21741	118,21741	41,00409	0,0000
Satt384	1	46,30573	46,30573	16,06129	0,0091
Desvio	159	458,4072	2,88306		
-----					
Total	161	619,5932			
R <sup>2</sup> : 0,2601		R <sup>2</sup> ajustado: 0,2508		AIC: 174,5062	
Coef. Reg	Valor	Desvio	t (bilateral)	Probabilidade (%)	
B(1)	0,86332	0,13482	6,40344	0,00000	
B(2)	0,53741	0,1341	4,00765	0,01456	
Constante	37,63				

AIC (Critério de Informação de Akaike).

Tabela 24 - QTL associado ao teor de proteína em grãos de soja presente no grupo de ligação E.

Característica	GL	Intervalo/Região	r	LR	R <sub>%</sub> <sup>2</sup>	a
Teor de proteína	E	Satt384-Sat_112	0,1349	13,24**	7,51	0,5425

GL = Grupos de ligação denominados segundo Song *et al.* (2004). R<sub>%</sub><sup>2</sup> = Coeficiente de determinação; a = efeito aditivo estimado para o QTL. (\*\*) valor significativo de LR que excedeu o valor crítico de corte a 1%.

A utilização de NILs na identificação de QTLs pelo mapeamento por marca simples minimiza o problema abordado por Bearzoti (2000) em que o método confunde efeito e posição de QTLs, pois não há distinção entre a ocorrência de um QTL de pequeno efeito situado próximo ao marcador e de um QTL de grande efeito só que distante do marcador. Assim, no caso de NILs, apenas marcadores próximos a QTLs serão identificados, não importando se são de pequeno ou grande efeito. A sensibilidade em identificar QTLs de pequeno efeito deve-se mais à precisão do delineamento estatístico empregado na avaliação fenotípica da população e do número de indivíduos utilizados. De acordo com Kaepler (1997), outra vantagem do uso de NILs no mapeamento de QTLs é o fato de esta população oferecer uma maior acurácia na estimação dos

efeitos dos QTLs detectados, pois, em outros tipos de populações, múltiplos QTLs estarão segregando. Essa maior acurácia pode ser comprovada, pois o QTL associado ao marcador Satt239 presente no GL I coincidiu com a região e magnitude que os autores Chung *et al.* (2003) e Nichols *et al.* (2006) encontraram utilizando diferentes backgrounds genéticos, demonstrando também a presença de genes conservados nesta região, possivelmente envolvidos no controle da expressão de alguma proteína. Outra evidência da acurácia do uso de NILs utilizando marca simples comparado ao mapeamento por intervalo composto e com diferentes populações foi quando se comparou o resultado do presente trabalho com os encontrados por Soares (2004) que avaliou uma população de RILs testada em dois ambientes e por Rodrigues (2008) que avaliou uma população  $F_{3:4}$ , ambas de mesmo background genético desta população de NILs. Eles encontraram esse marcador associado a um QTL que, no primeiro trabalho, explicou 14,74% e 11,34% da variação do fenótipo e 18,83% no segundo trabalho, próximo ao encontrado neste trabalho que foi de 19,04%. Isso demonstra que o emprego de mapeamento por marca simples quando utilizado NILs é tão eficiente quanto o emprego de outros tipos de populações e metodologias mais complexas, mas com a vantagem de demandar uma menor quantidade de *primers* para a genotipagem da população.

O QTL associado ao marcador Satt 384 presente no GL E coincidiu com o encontrado por Soares (2004) neste mesmo grupo e associado a um QTL que explicava 8,8% da variação da concentração da subunidade protéica beta.

A posição ocupada pelo marcador Satt384 no grupo de ligação E indica que ele está muito próximo ou acima desse QTL que é responsável em parte pela expressão dessa proteína, sendo, portanto, útil no processo de seleção assistida, juntamente com o marcador Satt 239, que é muito conservado.



## 6. CONCLUSÕES

Os resultados obtidos permitiram as seguintes conclusões:

A população que foi submetida ao primeiro ciclo de seleção apresentou variabilidade genética significativa para as características produção de grãos por planta (PROD), teor de proteína (PTN) e teor de óleo (OL), permitindo a predição de ganhos utilizando diferentes estratégias e critérios de seleção. Os valores de herdabilidade restrita entre famílias foram bem superiores à herdabilidade dentro de famílias tanto para PROD como para PTN, reduzindo a superioridade das estratégias que utilizam a informação em nível de indivíduo (seleção entre e dentro e seleção combinada) em relação àquelas que só utilizam a informação de média de famílias (seleção entre).

A herdabilidade entre médias de famílias com valores acima de 50% permitiu ganhos tanto para produção de grãos como para teor protéico. A correlação genética entre PTN × OL foi sempre altamente correlacionada negativamente, enquanto a correlação entre PG × PTN não foi significativa, o que permitiu obter ganhos tanto para PROD como para PTN, mas nunca para PTN e OL, simultaneamente.

A utilização da seleção simultânea com base nos índices de seleção proporcionou a maximização dos ganhos para teor protéico, ganhos de moderado a alto para produção de grãos e minimizou a redução dos ganhos para o teor de óleo, com destaque para o índice de Pesek e Baker generalizado por Tai.

A seleção praticada no terceiro ciclo de seleção realizada no ano agrícola de 2006/2007 com as linhagens previamente selecionadas no primeiro ciclo, quando utilizado índices como critérios de seleção, permitiu a seleção de dois grupos de linhagens, uma com máxima produtividade e redução no ciclo, maior do que a variedade comercial Monarca e com menor ciclo, além da resistência ao glifosato. O segundo grupo formado teve ganhos maximizados para teor de proteína e moderados para produtividade de grãos, havendo também uma redução do ciclo, além de possuir resistência ao glifosato, entretanto, apresentou ganhos negativos para produtividade em relação à variedade Monarca.

Na população de mapeamento, foram encontrados dois marcadores, Satt 239 e Satt 384 presentes nos GL I e GL E do mapa integrado da soja, associados a QTLs que explicavam 19,04% e 7,42% da variação do fenótipo, e

quando se empregou o modelo de regressão linear múltipla utilizando o procedimento de eliminação *stepwise*, esses mesmos QTLs foram mantidos e conjuntamente explicavam 26,01% da variação fenotípica da característica teor de proteína.

Utilizando-se o mapeamento por intervalo simples no GL E, foi estimado o efeito aditivo do QTL na expressão do caráter (+ 0,54) e determinada a posição do marcador em relação ao QTL, mostrando-se muito próximo ou coincidente ao QTL, o que resulta em uma marca extremamente confiável para futuros trabalhos de seleção assistida. Isso demonstra que o emprego de mapeamento por marca simples quando se utiliza NILs é tão eficiente quanto o emprego de outros tipos de populações e metodologias mais complexas, haja vista os resultados semelhantes obtidos por outros autores que utilizaram a mesma fonte de alto teor de proteína e outros tipos de populações. No entanto, o uso de NILs tem a vantagem de demandar menor quantidade de *primers* para a genotipagem da população.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABIOVE. **Dados do Complexo Soja.** Disponível em: <<http://www.abiove.com.br/>>. (Acessado em: Maio de 2008), 2008.
- AKHTER, M.; SNELLER, C.H. Yield and yield components of early maturing soybean genotypes in the Mid-South. **Crop Science**, v.36, p.877-882, 1996a.
- AKHTER, M.; SNELLER, C.H. Genotype x planting date interaction and selection of early maturing soybean genotypes. **Crop Science**, v.36, p.883-889, 1996b.
- AKKAYA, M.S.; BHAGWAT, A.A.; CREGAN, P.B. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics**, v.132, p.1131-1139, 1992.
- ALMEIDA, L.A.; KIIHL, R.A.S.; ABDELNOOR, R.V. Melhoramento da soja. In: Simpósio sobre Atualização em Genética e Melhoramento de Plantas. ABREU, A.F.B.; GONÇALVES, F.M.A.; MARQUES Jr., O.G.; e RIBEIRO, P.H.E. (eds.). Lavras, MG, UFLA-GEN, 1997. p.9-55.
- ALT, B.J.; FEHR, W.R.; WELKE, G.A. Selection for large seed and high protein in two and three parent soybean populations. **Crop Science**, v.42, p.1876-1881, 2002.
- ANAND, S.C.; TORRIE, J.H. Heritability of yield and other traits and interrelationships among traits in the F<sub>3</sub> and F<sub>4</sub> generations of three soybean crosses. **Crop Science**, v.3, p.508-511, 1963.
- APUYA, N.R.; FRAZIER, B.L.; KEIM, P.; ROTH, E.J.; LARK, K.G. Restriction fragment length polymorphisms as genetic markers in soybean *Glycine max* (L.) Merrill. **Theoretical and Applied Genetics**, v.75, p.889-901, 1988.
- AZEVEDO FILHO, J.A.; VELLO, N.A.; GOMES, R.L.F. Estimativas de parâmetros genéticos de populações de soja em solos contrastantes na saturação de alumínio. **Bragantia**, v.57, p.227-239, 1998.
- BACKES, R. L. **Estimativas de parâmetros genéticos, correção do efeito ambiental e predição de ganhos por seleção em populações F5 e F6 de soja (Glycine max (L.) Merrill)** Viçosa, MG:UFV, 2000. 92p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa.
- BACKES, R.L.; REIS, M.S.; CRUZ, C.D.; SEDIYAMA, T.; SEDIYAMA, C.S. Correlation estimates and assessment of selection strategies in five soybean populations. **Crop Breeding and Applied Biotechnology**, v.3, p.107-116, 2003.
- BASTEN, C.J.; WEIR, B.S.; ZENG, Z.B. **QTL cartographer; version 1.13.** Raleigh, NC: Department of Statistics; North Carolina State University, 1999. 132p.
- BEARZOTI, E. Mapeamento de QTL. In: PINHEIRO, J.B., CARNEIRO, I.F (Eds.). **Análise de QTL no melhoramento de plantas.** Goiânia: FUNAPE, 2000. p. 63-223.

BECKMANN, J.S.; SOLLER, M. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. **Theoretical and Applied Genetics**, v.67, p.35-43, 1983.

BONATO, E.R.; BERTAGNOLLI, P.F.; LANGE, C.E.; RUBIN, S.A.L. Teor de óleo e proteína em genótipos de soja desenvolvidos após 1990. **Pesquisa Agropecuária Brasileira**, v.35, n.12, p.2391-2398, 2000.

BONETTI, L.P. Distribuição da soja no mundo. In: MIYASAKA, S., MEDINA, J.C. (Eds.). A soja no Brasil. Campinas: ITAL, 1981. p.1-6.

BRIM, C.A.; BURTON, J.W. Recurrent selection in soybeans II. Selection for increased percent protein in seeds. **Crop Science**, v.19, p.494-498, 1979.

BRUMMER, E.C.; GRAEF, G.L.; ORF, J.; WILCOX, J.R.; SHOEMAKER, R.C. Mapping QTL for seed protein and oil content in eight soybean populations. **Crop Science**, v.37, n.1, p.370-378, 1997.

BURR, B.; BURR, F.A.; THOMPSON, K.H.; ALBERTSON, M.; STUBER, C.W. Gene mapping with recombinant inbreds in maize. **Genetics**, v.118, n.4, p.519-526, 1988.

CARVALHO, C.G.P.; ARIAS, C.A.A.; TOLEDO, J.F.F.; ALMEIDA, L.A.; KIIHL, R.A.S.; OLIVEIRA, M.F. Interação genótipo x ambiente no desempenho produtivo da soja no Paraná. **Pesquisa Agropecuária Brasileira**, v.37, p.989-1000, 2002.

CECON, P.R.; MORAIS, R.A.; SEDIYAMA, C.S. Obtenção da herdabilidade e das correlações genotípicas, fenotípicas e de ambiente nas gerações F<sub>2</sub> e F<sub>3</sub> em cruzamentos fatoriais em soja. **Pesquisa Agropecuária Brasileira**, v.28, p.1399-1406, 1993.

CERVIGNI, G.D.L. **Mapeamento de genes de resistência à raça 3 do nematóide de cisto da soja (*Heterodera glycines* Ichinohe)**. 1999. 52p. Dissertação (Mestrado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.

CHANG, S.J.C.; DOUBLER, T.W.; KILO, V.; SUTTNER, R.; KLEIN, J.; SCHMIDT, M.E.; GIBSON, P.T.; LIGHTFOOT, D.A. Two additional loci underlying durable field resistance to soybean sudden death syndrome (SDS). **Crop Science**, v.36, p.1684-1688, 1996.

CHUNG, J.; BABKA, H.L.; GRAEF, G.L.; STASWICK, P.E.; LEE, D.J.; CREGAN, P.B.; SHOEMAKER, R.C.; SPEECH, J.E. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop Science**, v.4, p.1053-1067, 2003.

CIANZIO, S.R.; FEHR, W.R. Genetic variability for soybean seed composition in crosses between high and low protein parents. **Journal of Agriculture of University of Puerto Rico**, v.66, p.123-129, 1982.

COBER, E.R.; VOLDENG, H.D. Developing high protein, high-yield soybean populations and lines. **Crop Science**, v.40, n.1, p.39-42, 2000.

COE, H.E.; NEUFFER, M.G.; HOISINGTON, D.A. The genetics of corn. In: SPRAGUE, G. F.; DUDLEY, J.W. (Eds.). **Corn and corn improvement**. 1988. p.81-237.

CONAB. Companhia Nacional de Abastecimento. **Central de Informações Agropecuárias: Safras 2008**. Disponível em: <<http://www.conab.gov.br/>>. Acessado em: Maio de 2008.

CONCIBIDO, V.C.; DENNY, R.L.; LANGE, D.A.; ORF, J.H.; YOUNG, N.D. RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. **Crop Science**, v.36, n.6, p.1643-1650, 1996.

COSTA, M.M.; Di MAURO, A.O.; UNÊDA-TREVISOLI, S.H.; ARRIEL, N.H.C.; BÁRBARO, I.M.; da SILVEIRA, G.D.; MUNIZ, F.R.S. Analysis of direct and indirect selection and índices in soybean segregating populations. **Crop Breeding and Applied Biotechnology**, v.8, p.47-55, 2008. Brazilian Society of Plant Breeding.

COSTA, M.M.; MAURO, A.O.; UNÊDA-TREVISOLI, S.H.; ARRIEL, N.H.C.; BÁRBARO, I.M.; MUNIZ, F.R.S. Ganho genético por diferentes critérios de seleção em populações segregantes de soja. **Pesquisa Agropecuária Brasileira**, v.39, p.1095-1102, 2004.

CREGAN, P.B.; JARVIK, T.; BUSH, R.C.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VANTOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPEXHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, v.39, p.1464-1490, 1999.

CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. 1990. 188p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

CRUZ C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 2a ed. Viçosa: UFV, 1997. 390p.

CRUZ, C.D. **Programa Genes**: versão Windows; aplicativo computacional em genética e estatística. Viçosa, MG: Editora UFV, 2001. 648p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos Aplicados ao Melhoramento Genético**. Viçosa, MG: UFV, v.2, 2003. 585p.

CRUZ, C.D. **Programa Genes: Estatística experimental e matrizes**. Viçosa MG: Editora UFV, 285p. 2006.

CRUZ, C.D. **Programa Genes: Biometria**. Viçosa MG: Editora UFV, 382p. 2006.

DEMBOGURSKI, N.M.S.S. **Determinação do preço da soja para trituração e obtenção do óleo com base na qualidade do grão**. Ijuí: UNIJUI, 2003. Dissertação (Mestrado em Modelagem Matemática) – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, 2003.

- DOERGE, R.W.; CHURCHILL, G.A. Permutation tests for multiple loci affecting a quantitative character. **Genetics**, v.142, n.1, p.285-294, 1996.
- DIERS, B.W., KEIM, P., FEHR, W.R., SHOEMAKER, R.C. RFLP analysis of soybean seed protein and oil content. **Theoretical and Applied Genetics**, v.83, n.5, p.608-612, 1992.
- DUDLEY, J.M.; MOLL, R.H. Interpretation and use of estimates of heritability and genetics variances in plant breeding. **Crop Science**, Madison, v.9, n.3, p.257-262, 1969.
- DUDLEY, J.W. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. **Crop Science**, v.33, p.660-668, 1993.
- EBERHART, S.A. Factors affecting efficiencies of breeding methods. **African Soils**, v. 5, p. 669-672, 1970.
- EMBRAPA. **Tecnologia de produção de soja para a região central do Brasil**. Londrina: Embrapa, Embrapa Agropecuária Oeste, 2007. 228p.
- EMBRAPA. Disponível em [www.embrapa.br](http://www.embrapa.br). Acessado em 11/2007.
- ERICKSON, L.R.; VOLDENG, H.D.; BEVERSDORF, W.D. Early generation selection for protein in *Glycine max* × *Glycine soja* crosses. **Canadian Journal of Plant Science**, v.61, p.901-908, 1981.
- FALCONER, D.S. **Introdução à genética quantitativa**. 1ª ed. (Trad. de Silva, M.A.). Viçosa: UFV, 1987. 279p.
- FARIAS NETO, J.T.; VELLO, N.A. Avaliação de progênies F<sub>4:3</sub> e F<sub>5:3</sub> e estimativas de parâmetros genéticos com ênfase para porcentagem de óleo e produtividade de grãos e óleo em soja. **Ciência e Agrotecnologia**, Lavras, v.25, n. 4, p.812-820, 2001.
- FASOULA, V.A.; HARRIS, D.K.; BOERMA, R.H. Validation and designation of quantitative trait loci for seed protein, seed oil and seed weight from two soybean populations. **Crop Science**, v.44, p.1218 -1225, 2004.
- FEDERER, W.T. **Experimental design: theory and application**. New York: MacMillan Company, 1955. 544p.
- FEHR, W.R.; WEBER, C.R. Mass selection by seed size and specific gravity in soybean populations. **Crop Science**, v.8, p.551-554, 1968.
- FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas**. 3.ed. Brasília: EMBRAPA; CENARGEN, 1998. 220p.
- FREY, K.J., HORNER, T. Heritability in standard units. **Agronomy Journal**, v.49, n.1, p.59-62, 1957.

FRONZA, V. **Genética da reação da soja a *Fusarium solani* f.sp. *glycines*.**, 2003. Tese (Doutorado em Agronomia). 154p. Escola Superior de Agricultura "Luiz de Queiroz"/USP, Piracicaba, SP, Brasil.

GARCIA, A.A.F.; SOUZA Jr., C.L. Comparação de índices de seleção não paramétricos para a seleção de cultivares. **Bragantia**, v.58, p.253-267, 1999.

GOMES, R.F.L. **Análise genética de progênes F<sub>6</sub> e F<sub>7:6</sub> de soja obtidas de cruzamentos dialélicos.** Piracicaba, 1995. 140p. Tese (Doutorado) – ESALQ/USP.

GOMES, F.P. **Curso de estatística experimental.** 13.ed. Piracicaba: Nobel, 1990. 468p.

GQMOL. Viçosa: 05.07.2008. (<http://www.ufv.br/dbg/gqmol/gqmol.htm>).

GRANATE, M.J.; CRUZ, C.D.; PACHECO, C.A.P. Predição de ganho genético com diferentes índices de seleção no milho pipoca CMS-43. **Pesquisa Agropecuária Brasileira**, v.37, p.1001-1008, 2002.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, Baltimore, v.137, p.1121-1137, 1994.

GUIMARÃES, C.T.; MOREIRA, M.A. Genética molecular aplicada ao melhoramento de plantas. In: BORÉM, A. (Ed). **Melhoramento de espécies cultivadas.** Ed. UFV, Viçosa-MG, p.715-740, 1999.

GUIMARÃES, C.T.; SCHUSTER, I.; MAGALHÃES, J.V.; SOUZA Jr., C.L. Marcadores moleculares no melhoramento de plantas. In: BORÉM, A.; CAIXETA, E.T. (Eds). **Marcadores moleculares.** Ed. UFV, Viçosa-MG, p.107-144, 2006.

GUZMAN, P.S.; DIERS, B.W.; NEECE, D.J.; MARTIN, S.K.S.; LEROY, A.R.; GRAU, C.R.; HUGHES, T.J.; NELSON, R.L. QTL associated with yield in three backcross-derived populations of soybean. **Crop Sci.** v.47, p.111-122. 2007.

HALEY, C. S.; KNOTT, S. A. A simple regression method for mapping quantitative trait loci in the line crosses using flanking markers. **Heredity**, v.69, p.315-324. 1992.

HALEY, C.S., KNOTT, S.A., AND ELSEN, J.M. Mapping quantitative trait loci in crosses between outbred lines using least squares. **Genetics**, v.136, p.1195-1207, 1994.

HALLAUER, A.R.; MIRANDA FILHO, J.B. **Quantitative genetics in maize breeding.** Ames: Iowa State University, 1981. 468p.

HANSON, C.H.; ROBINSON, H.F.; COMSTOCK, R.E. Biometrical studies of yield in segregating populations of Korean lespedeza. **Agronomy Journal**, v.48, p.268-272, 1956.

HANSON, W.D.; WEBER, C.R. Resolution of genetic variability in self-pollinated species with an application to the soybean. **Genetics**, v.45, p.1425-1434, 1961.

HANSON, W.D. Heritability. In: SYMPOSIUM ON STATISTICAL GENETICS AND PLANT BREEDING, 1961, Raleigh. **Proceedings...** Washington: NAS/NRC, 1963. p.125-140.

HARTWIG, E.E. Breeding soybean for high seed yield and seed protein. In: **Genetic Improvement. Proceedings: World Soybean Research Conference V.** Kasetsart University Press and WSRCV, p.74-83, 1994.

HAYWARD, M.D.; McADAM, N.J.; JONES, J.G.; EVANS, C.; EVANS, G.M.; FORSTER, J.W.; USTIN, A.; HOSSIN, K.G.; QUADER, B.; STAMMERS, M.; WILL, J.K. Genetic markers and the selection of quantitative traits in forage grasses. **Euphytica**, v.77, n.3, p.269-275, 1994.

HAZEL, L.N. The genetic basis for constructing selection indexes. **Genetics**, v.28, p.476-490, 1943.

HELMS, T.C.; ORF, J.H. Protein, oil and yield of soybean lines selected for increased protein. **Crop Science**, v.38, p.707-711, 1998.

HIROMOTO, D.M.; VELLO, N.A. Genetic base of Brazilian soybean cultivars. **Revista Brasileira de Genética**, v.9, n.2, p.295-306, 1986.

HNETKOVSKY, N.; CHANG, S.J.C.; DOUBLER, T.W.; GIBSON, P.T.; LIGHTFOOT, D.A. Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). **Crop Science**, v.36, p.393-400, 1996.

HOLBROOK, C.C.; BURTON, J.W.; CARTER JÚNIOR, T.E. Evolution of recurrent restricted index selection for increasing yield while holding seed protein constant in soybean. **Crop Science**, v.29, p.324-329, 1989.

HYTEN, D.L.; PANTALONE, V.R.; SAMS, C.E.; SAXTON, A.M.; LANDAU-ELLIS, D.; STEFANIAK, T.R.; SCHMIDT, M.E. Seed Quality QTL in a Prominent Soybean Population. **Theoretical and Applied Genetics**, v.109, p.552-561, 2004.

INSTITUTO ADOLFO LUTZ. **Normas analíticas do Instituto Adolfo Lutz, Métodos químicos e físicos para análise de alimentos.** 3ª ed., São Paulo, v.1, p.533, 1985.

IQBAL, M.J.; MEKSEM, K.; NJITI, V.N.; KASSEM, M.A.; LIGHTFOOT, D.A. Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex X Forrest RILs. **Theoretical and Applied Genetics**, v.102, p.187-192, 2001.

JANSEN, R.C. A general mixture model for mapping quantitative trait loci by using molecular marker. **Theoretical and Applied Genetics**, v.85, p. 252-260, 1992.

JANSEN, R.C. Interval mapping of multiple quantitative trait loci. **Genetics**, v. 135, p. 205-211. 1993.



JAMAGO, J.M.; NELSON, R.L. **Relationship Between Seed Yield and Protein in High Protein Soybean Populations**. ASA-CSSA-SSSA International Meetings. Indianapolis, 2006.

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

JOHNSON, H.W.; ROBINSON, H.F.; CONSTOCK, R.E. Genotypic and phenotypic correlations in soybeans and their implications in selection. **Agronomy Journal**, v.47, p.477-483, 1955.

JOHNSON, H.W.; BERNARD, R.L. Soybean genetics and breeding. In: NORMAN, A.G. (Ed.). **The soybean genetics, breeding, physiology, nutrition, management**. New York: Academic Press, 1963. p.1-73.

KAEPLER, S.M. Quantitative trait locus mapping using sets of near-isogenic lines: relative power comparisons and technical considerations. **Theoretical and Applied Genetics**, v. 95, p. 384-392. 1997.

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

KEIM, P.; DIERS, B.W.; OLSON, T.C.; SHOEMAKER, R.C. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. **Theoretical and Applied Genetics**, v.126, n.3, p.735-742, 1990.

KEIM, P.; SCUPP, J.M.; TRAVIS, S.E.; CLAYTON, K.; WEBB, D.M. A high-density soybean genetic map based upon AFLP markers. **Crop Science**, v.37, n.2, p.537-543, 1997.

KEMPTHORNE, O.; NORDSKOG, A.W. Restricted selection indexes. **Biometrics**, Alexandria, v.15, p:10-19, 1959.

KILO, V.Y.; ABU-THREDEIH, J.; SUTTNER, R.J.; CHANG S.J.C.; GIBSON, P.T.; LIGHTFOOT, D.A. Coinheritance of resistance to Sudden Death Syndrome (SDS) and Cyst Nematode (SCN) in soybean. **Soybean Genetics Newsletter**, v.24, p.126-127, 1997.

KNAPP, S.J. Mapping quantitative trait loci using molecular markers: multilocus estimators of backcross, recombinant inbred, and doubled haploid parameters. **Theoretical and Applied Genetic**, v.81, p.333-338, 1991.

KWOK, P.Y.; DENG, Q.; ZAKERI, H.; NICKERSON, D.A. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. **Genomics**, v.31; p.123-126. 1996.

KWON, S.H.; TORRIE, J.H. Heritability and interrelationships among traits of two soybean populations. **Crop Science**, v.4, n.2, p.196-198, 1964.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**; v.124, p.743-756, 1990.

LANDER, E.S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M.J.; LINCOLN, S.E.; NEWBURG, L. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics**, v.1, n.1, p.174-181, 1987.

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

LARK, K.G.; WEISEMANN, J.M.; MATTHEWS, B.F.; PALMER, R.; CHASE, K.; MACALMA, T. A genetic map of soybean (*Glycine max* L. Merr) using an intraspecific cross of two cultivars: 'Minosy' and 'Noir 1'. **Theoretical and Applied Genetics**, v.86, n.8, p.901-906, 1993.

LEE, M. DNA markers and plant breeding programs. **Advances in Agronomy**, San Diego, v.55, p.265-344, 1995.

LIU, B.H. **Statistical genomics: linkage, mapping, and QTL analysis**. Boca Raton, Flórida, USA: CRC Press.1998. 611p.

LUSH, J.L. **Animal breeding plans**. 3.ed. Ames : State College, 1945. 443p.

LUSH, J.L. Family merit and individual merit as bases for selection. **American Naturalist**, v.81, p.241-261, 1947.

LUSH, J.L. **Melhoramento dos animais domésticos**. Rio de Janeiro: CEDEGRA, 1964. 566p.

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

MAHMUD, I.; KRAMER, H.H. Segregation for yield, height, and maturity following a soybean cross. **Agronomy Journal**, v.43 n.12, p.605-609, 1951.

MANSUR, L.M.L.; LARK, K.G.; KROSS, H.; OLIVEIRA, A. Interval mapping of quantitative trait loci for reproductive, morphological and seed traits of soybean (*Glycine max* L. Merr.). **Theoretical and Applied Genetics**, v.86, n.8, p.907-913, 1993.

MAURO, A.O.Di.; OLIVEIRA, R.C.; MARCONDES, A.F.; SEDIYAMA, T. Ganho genético por seleção em linhagens de soja. **Rev. Ceres**, v. 47, n. 270, p. 135-144, 2000.

McDONALD, M.B.; ELLIOT, E.J.; SWEENEY, P.M. DNA extraction from dry seeds for RAPD analyses in varietal identification studies. **Seed Science and Technology**. v.22, p.171-176, 1994.

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, p.603-607, 1985.

- MILLER, J.E.; FEHR, W.R. Direct and indirect recurrent selection for protein in soybeans. **Crop Science**, v.19, p.101-106, 1979.
- MIRANDA, F.D. **Produção, conteúdo de proteína e óleo no grão da soja: herdabilidades, correlações e seleção de genótipos superiores**. Viçosa, MG: UFV, 2006. 76p. Tese (Doutorado em Genética e melhoramento) – Universidade Federal de Viçosa, 2006.
- MORAIS, O.P. **Análise multivariada da divergência genética dos progenitores, índices de seleção e seleção combinada numa população de arroz oriunda de intercruzamentos, usando macho-esterilidade**. Viçosa, 1992. 251p. Tese (Doutorado em Genética e Melhoramento) - UFV, 1992.
- MOREAU L.; CHARCOSSET A.; GALLAIS A. Experimental evaluation of several cycles of marker-assisted selection in maize. **Euphytica**, v.137, p.111-118, 2004.
- MORGANTE, M.; RAFALSKI, A.; BIDDLE, P.; TINGEY, S.; OLIVIERI, A.M. Genetic mapping and variability of seven soybean simple sequence repeat loci. **Genoma**, v.37, n.5, p.763-769, 1994.
- MORRISON, M.J.; VOLDENG, H.D.; COBER, E.R. Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. **Agronomy Journal**, v.92, p.780-784, 2000.
- MUDGE, J.; CREGAN, P.B.; KENWORTHY, J.P.; KENWORTHY, J.W.; ORF, J.H.; YOUNG, N.D. Two microsatellite markers that flank the major soybean cyst nematode resistance locus. **Crop Science**, v.37, n.3, p.1611-1615, 1997.
- NICHOLS, D.M.; GLOVER, K.D.; CARLSON, S.R.; SPECHT, J.E.; DIERS B.W.. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. **Crop Science**, 46:834-839. 2006
- NJITI, V.N.; MEKSEM, K.; IQBAL, M.J.; JOHNSON, J.E.; ZOBRIK, K.F.; KILO, V.Y.; LIGHTFOOT D.A. Common Loci Underlie Field Resistance to Soybean Sudden Death Syndrome in Forrest, Pyramid, Essex, and Douglas. **Theoretical and Applied Genetics**, v.104, p. 294- 300, 2002.
- OLIVEIRA, A.C.B.; SEDIYAMA, C.S.; CRUZ, C.D. Selection for later flowering in soybean (*Glycine max* L. Merrill) F<sub>2</sub> populations cultivated under short day conditions. **Genetics and Molecular Biology**, v.22, p.243-247, 1999.
- OPENSHAW, S.J.; HADLEY, H.H. Selection indexes to modify protein concentration of soybean seeds. **Crop Science**, v.24, p.1-4, 1984.
- PALMER, R.G.; HYMOWITZ, T.; NELSON, R.L. Germoplasm diversity within soybean. In: **Soybean genetics, molecular biology and biotechnology**. Ohio, USA:CAB INTERNACIONAL, p.1-35, 1996.
- PANTHEE, D.R.; PANTALONE, V.R.; WEST, D.R.; SAXTON, A.M.; SAMS, C.E. Quantitative trait loci for seed protein and oil concentration and seed size in soybean. **Crop Science**, v. 45, p. 2015-2022, 2005.

PANTHEE, D.R.; PANTALONE, V.R.; SAXTON, A.M.; WEST, D.R.; SAMS, C.E. Quantitative trait loci for agronomic traits in soybean. **Plant Breeding**, 126:51-57. 2007.

PATERSON, A.H.; LANDER, E.S.; HEWITT, J.D.; PATERSON, S.; LINCOLN, S.E.; TANKSLEY, S.D. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphism. **Nature**, v.335, n.6193, p.721-726, 1988.

PATERSON, A.H.; DAMON, S.; HEWITT, J.D.; ZAMIR, D.; RABINOWITCH, H.D.; LINCOLN, S.E.; LANDER, E.S.; TANKSLEY, S.D. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. **Genetics**, v.127, n.1, p.181-197, 1991.

PAULA, R.C. de; PIRES, I.E.; BORGES, R. de G.; CRUZ, C.D. Predição de ganhos genéticos em melhoramento florestal. **Pesquisa Agropecuária Brasileira**, v.37, p.159-165, 2002.

PESEK, J.; BAKER, R.J. Desired improvement in relation to selected indexes. **Canadian Journal of Plant Science**, v.49, p.803-804, 1969.

PIRES, I.E. **Eficiência da seleção combinada no melhoramento genético de *Eucalyptus* spp.** 1996. 116f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa.

PRABHU, R. R.; NJITI, V. N.; BELL JOHNSON, B.; JOHNSON, J.E.; SCHIMIDT; M.E.; KLEIN, J.H.; LIGHTFOOT, D.A. Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. **Crop Science**, v. 39, p. 982-987, 1999.

PRADO, L.C. Perspectivas para o melhoramento de linhagens homozigóticas de soja, com ênfase em produção de grãos e teor de proteína. Piracicaba, 1994. 128p. Dissertação (Mestrado) – ESALQ/USP.

PROBST, A.H.; JUDD, R.W. Origin U.S. history and development and world distribution. In: CALDEWELL, B.E. **Soybeans, improvement, production and uses**. Madison ASA, 1973. 618p.

RAMALHO, M.A.P.; VENCOVSKY, R. Estimação dos componentes da variância genética em plantas autógamas. **Ciência e Prática**, v.2, n.2, p.117-140, 1978.

RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. **Experimentação em genética e melhoramento de plantas**. Lavras : UFLA, 2000. 303p.

REIS, E.F. Ganhos preditos e realizados, por diferentes estratégias de seleção, em populações de soja (*Glycine max* (L.) Merrill). Viçosa, MG: UFV, 2000. 120p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2000.

REIS, E.F.; REIS, M.S.; CRUZ, C.D.; SEDIYAMA, T. Comparação de procedimentos de seleção para produção de grãos em populações de soja. **Ciência Rural**, Santa Maria, v. 34, n. 3, p. 685-692, 2004.

RESENDE, M.D.V.; HIGA, A.R. Maximização da eficiência da seleção em testes de progênies de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Boletim de Pesquisa Florestal**, Curitiba, n.28/29, p.37-56, 1994.

RESENDE, M.D.V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Colombo: EMBRAPA Floresta; Brasília: Informação Tecnológica, 2002. 975p.

RICK, C.M.; YODER, J.I. Classical and molecular genetics of tomato highlights and perspectives. **Annual Review of Genetics**, v.22, p.281-300, 1988.

RIOS, M.C.D. **Alternativas de procedimentos estatísticos para avaliação de genótipos em programas de melhoramento genético vegetal**. Piracicaba, 1997. 59p. Dissertação (Mestrado) – ESALQ/USP.

SAFRAS & MERCADO. Agências Safras – Últimas Notícias. Disponível em <<http://www.safras.com.br/>>. Acessado em: Março de 2008.

SAGHAI MAROOF, M. A.; BIYASHEV, R. M.; YANG, G. P.; ZHANG, Q.; ALLARD, R.W. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations and population dynamics. **Proceedings National Academy Science of United States of America**, v.91, p.5466-5470, 1994.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. 2.ed. New York: Cold Spring Harbor Laboratory, 1989. v.3, p.8B.

SANTOS, C.A.F.; REIS, M.S.; CRUZ, C.D.; SEDIYAMA, C.S.; SEDIYAMA, T. Adequação de modelos no estudo do coeficiente de trilha nos componentes primários e secundários de progênies F6 de soja (*Glycine max* (L.) Merrill). **Revista Ceres**, v. 42, p.111-121. 1995.

SAX, K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**. v.8, p.552-560, 1923.

SCHUSTER, I.; ABDELNOOR, R.V.; MARIN, S.R.R.; CARVALHO, V.P.; KIIHL, R.A.S.; SILVA, J.F.V.; SEDIYAMA, C.S.; BARROS, E.G.; MOREIRA, M.A. Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*). **Theoretical and Applied Genetics**, v.102, p.91-96, 2001.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. 1.ed. – Viçosa: UFV, 2004. 568p.

SCOTT, R.A.; KEPHART, K.D. Selection for yield, protein, and oil in soybean crosses between adapted and introduced parents. **Field Crops Research**, v.49, p.177-185, 1997.

SEBOLT, A.M.; SHOEMAKER, R.C.; DIERS, B.W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein. **Crop Science**, v.40, p:1438-1444, 2000.

SEDIYAMA, T.; TEIXEIRA, R.C.; REIS, M. S. Melhoramento de soja. In: BORÉM, A. (Ed.) **Melhoramento de espécies cultivadas**. Viçosa – MG, 1999. p.487-534.

SHANNON, J.G.; WILCOX, J.R.; PROBST, A.H. Estimated gains from selection for protein and yield in the F<sub>4</sub> generation of six soybean populations. **Crop Science**, v.12, p.824-830, 1972.

SHOEMAKER R. C.; GUFFY, R.D.; LOREZEN, L.L.; SPEECH, J.E. Molecular genetic mapping of soybean: map utilization. **Crop Science**, n.32, v.5, p.1901-1098, 1992.

SHOEMAKER, R.C. RFLP map of soybean. In: PHILLIPS, R.L., VASIL, I.K., (Eds). **DNA-based markers in plants**. [s.l.]: Kluwer Academic, p.2990-3009, 1994.

SHOEMAKER R. C.; POLZIN, K.; LABATE, J.; SPECHT, J.; BRUMMER, E. C.; OLSON, T.; YOUNG, N.; CONCIBIDO, V.; WILCOX, J.; TAMULONIS, J. P.; KOCHERT, G.; BOERMA, H.R. Genome duplication in soybean (*Glycine* subgenus soja). **Genetics**, v.144, p.329-338, 1996.

SHOEMAKER, R.C.; SCHLUETER, J.A.; CREGAN, P.; VOLKIN, L. The status of soybean genomics and its role in the development of soybean biotechnologies. **AgBioforum**, v.6, p.4-7, 2003.

SHORTER, R.; BYTH, D.E.; MUNGOMERY, V.E. Estimates of selection parameters associated with protein and oil content of soybean seeds (*Glycine max* (L.) Merrill). **Australian Journal of Agricultural Research**, v.28, p.211-222, 1976.

SILVA, L.C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. Viçosa, MG: UFV, 2005. 120p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa.

SINGH, R.J.; HYMOWITZ, T. The genomic relationship between *Glycine max* (L.) Merr. And *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. **Theoretical and Applied Genetics**, v.76, p.705-711, 1988.

SMITH, H.F. A discriminant function for plant selection. **Annual Eugenics**, v.7, p.240-250, 1936.

SMITH, J.D.; KINMAN, M.L. The use of parent-offspring regression as an estimator of heritability. **Crop Science**, v.5, n.6, p.595-596, 1965.

SMITH, R.R.; WEBER, C.R. Mass selection by specific gravity for protein and oil in soybean populations. **Crop Science**, v.8, p.373-377, 1968.

SOARES, T.C.B. **Mapeamento de locos associados ao conteúdo de proteínas de reserva em soja**. Viçosa, MG:UFV, 2004. 113p. Tese (Doutorado em Bioquímica Agrícola) – Universidade Federal de Viçosa, Viçosa.

SOARES, T.C.B.; PIOVESAN, N.D.; SCHUSTER, I.; CRUZ, C.D.; BARROS, E.G. de; MOREIRA, M.A. Quantitative genetic analysis of storage proteins in soybean. **Crop Breeding and Applied Biotechnology**, v.4, p.317-324, 2004.

SONG, Q.J.; MAREK, L.F.; SHOEMAKER, R.C.; LARK, K.G.; CONCIBIDO, V.C.; DELANNAY, X.; SPECHT J.E.; CREGAN, P.B. A new integrated genetic linkage map of soybean. **Theoretical and Applied Genetics**, 109:122-128. 2004.

SPEHAR, C.R. Field screening of soya bean (*Glycine max* (L.) Merrill) germoplasm for aluminium tolerance by the use of augmented design. **Euphytica**, v.76, p.203-213, 1994.

STROMBERG, L.D.; DUDLEY, J.W.; RUFENER, G.K. Comparing conventional early generation selection with molecular marker assisted selection in maize. **Crop Science**, v.34, p.1221-1225, 1994.

STUBER, C.W.; EDWARDS, M.D.; WENDEL, J.F. Molecular-marker facilitated investigations of quantitative trait loci in maize. **Crop Science**, v. 27, n. 4, p. 639-648, 1987.

TAI, G.C.C. Index selection with desired gains. **Crop Science**, Madison, v.17, p:182-183, 1977.

TANKSLEY, D.S.; YOUNG, N.D.; PATERSON, A.H.; BONIERBALE, M.W. RFLP mapping in plant breeding: new tools for an old science. **Biotechnology**, v.7, n.1, p.257-264, 1989.

TANKSLEY, S.D. Mapping polygenes. **Annu Rev Genet** 27, 205-33. 1993.

TAWARE, S.P.; HALVANKAR, G.B.; RAUT, V.M.; PATIL, V.P. Variability, correlation and path analysis in soybean hybrids. **Soybean Genetics Newsletter**, v.24, p.96-98, 1997.

TEIXEIRA, A.I. **Herdabilidades e correlações entre concentrações de proteína em soja avaliadas por diferentes metodologias**. Viçosa, MG: UFV, 2003. 53p. Dissertação (Mestrado em Bioquímica Agrícola) – Universidade Federal de Viçosa, 2003.

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, p.652-655, 1970.

VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E.; VIÉGAS, G.P. **Melhoramento e produção de milho**. Campinas: Fundação Cargill, 1987. p. 137-209.

VERNETTI, F.J. Origem da espécie, introdução e disseminação no Brasil. In: VERNETTI, F.J. (Ed.) **Soja: genética e melhoramento**. Campinas: Fundação Cargill, 1983. p.3-17.

VOLDENG, H.D.; COBER, E.R.; HUME, D.J.; GILLARD, C.; MORRISON, M.J. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. **Crop Science**, Madison, v.37, p.428-431, 1997.

WAKAWKAR, S.M.; YADAV, L.N.; KELKAR, G.M. Path coefficient analysis for some characters in soybean JNKVV. **Research Journal**, v.8, p.196-201, 1974. **Plant Breeding Abstracts**, v.46, p.905, 1976.

WANG, D.; GRAEF, G.L.; PROCOPIUK, A.M.; DIERS, B.W. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. **Theoretical and Applied Genetics**, v.108, p.458-467, 2004.

WARNER, J.N. A method for estimating heritability. **Agronomy Journal**, v.44, n.2, p.427-430, 1952.

WEBER, J.K., MAY, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **American Journal Human Genetic**, v.44, n.4, p.388-397, 1989.

WEBER, C.R.; MOORTHY, B.R. Heritable and nonheritable relationships and variability of oil content and agronomic characters in the F<sub>2</sub> generation of soybean crosses. **Agronomy Journal**, v.44, p.202-209, 1952.

WEBER, W.E. Selection in segregating of autogamous species. I. Selection response for combined selection. **Euphytica**, v.31, n.2, p.493-502, 1982.

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, p.927-931, 1987.

WILCOX, J.R. Increasing seed protein in soybean with eight cycles of recurrent selection. **Crop Science**, v. 38, p.1536-1540, 1998.

WILCOX, J.R.; CAVINS, J.F. Backcrossing higher seed protein to a soybean cultivar. **Crop Science**, v.35, n.4, p.1036-1041, 1995.

WILCOX, J.R.; GUODONG, Z. Relationship between seed yield and protein in determinate and indeterminate soybean population. **Crop Science**, v.37, p.361-364, 1997.

WILCOX, J.R.; SHIBLES, R.M. Interrelationships among seeds quality attributes in soybean. **Crop Science**, v.41, p.11-14, 2001.

XINHAI, L.; JINLING, W.; QINGKAI, Y.; SHAOJIE, J.; LIMING, W. The effect of selection method on the association of yield and seed protein with agronomic characters in an interspecific cross of soybean. **Soybean Genetics Newsletter**, v.26, p.11-19, 1999.



XU, H.; WILCOX, J.R. Recurrent selection for maturity and percent seed protein in *Glycine max* based on S<sub>0</sub> plant evaluations. **Euphytica**, v.62, p.51-57, 1992.

YAKLICH, R. W.; VINYARD, B.; CAMP, M.; DOUGLASS, S. Analysis of seed protein and oil from soybean northern and southern region uniform tests. **Crop Science**, v.42, p.1504-1515, 2002.

ZENG, Z.B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings National Academy Science of United States of America**, v.90, p.10972-10976, 1993.

ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, v.136, n.9, p.1457-1468, 1994.