

MURILO VIOTTO DEL CONTE

**USO DE MODELOS MISTOS NO MELHORAMENTO DA SOJA PARA TEORES DE
ÓLEO E PROTEÍNA NOS GRÃOS**

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

VIÇOSA
MINAS GERAIS - BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

D345u
2017

Del Conte, Murilo Viotto, 1992-

Uso de modelos mistos no melhoramento da soja para teores de óleo e proteína nos grãos / Murilo Viotto Del Conte. – Viçosa, MG, 2017.

viii, 60f. : il. ; 29 cm.

Orientador: Felipe Lopes da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Soja - Melhoramento genético. 2. *Glycine max*. 3. Soja - Produtos. 4. Soja como alimento. I. Universidade Federal de Viçosa. Departamento de Fitotecnia. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

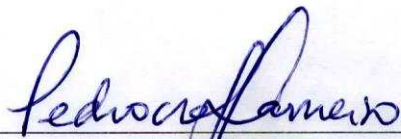
CDD 22 ed. 633.342

MURILO VIOTTO DEL CONTE

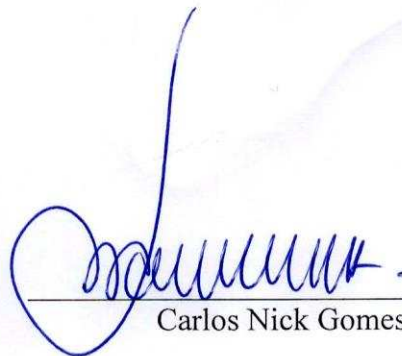
**USO DE MODELOS MISTOS NO MELHORAMENTO DA SOJA PARA TEORES DE
ÓLEO E PROTEÍNA NOS GRÃOS**

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

APROVADA: 17 de fevereiro de 2017



Pedro Crescêncio Souza Carneiro
(Coorientador)



Carlos Nick Gomes



Felipe Lopes da Silva
(Orientador)

A todos que de alguma forma contribuíram para meu aprendizado.

DEDICO

AGRADECIMENTOS

A Deus, pela graça da vida e por me guiar pelo caminho que me trouxeram até aqui.

Aos meus pais, Luzia Aparecida Viotto Del Conte e Anderson Del Conte; e ao meu irmão, Pedro Viotto Del Conte pelo amor incondicional, conselhos, apoio e suporte em todas as ocasiões.

À minha noiva Aline Biggi Maciel pelo amor, companheirismo e motivação investida em minhas escolhas.

À Universidade Federal de Viçosa, bem como ao Programa de Pós Graduação em Genética e Melhoramento e ao Departamento de Fitotecnia pelo acolhimento e pela oportunidade da realização do mestrado em Genética e Melhoramento.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao professor Felipe Lopes da Silva, pelo tempo dedicado às orientações, suporte para desenvolvimento dos trabalho e amizade.

Aos Professores Marcos Deon Vilela de Resende e Pedro Crescêncio Souza Carneiro pela coorientação e pelos ensinamentos.

Ao Professor Carlos Nick Gomes pela participação na banca e contribuições para este trabalho.

Aos estagiários, funcionários e demais colegas de pós-graduação do Programa Soja pela ajuda na condução dos experimentos, ideias compartilhadas, discussões técnico-científicas e amizade.

Ao BIOAGRO-UFV por viabilizar e contribuir com as análises realizadas neste estudo.

À Fundação de Apoio à Pesquisa de Minas Gerais (FAPEMIG) pelos recursos investidos no desenvolvimento do projeto.

“Transportai um punhado de terra todos os dias e fareis uma montanha”.

(Confúcio)

RESUMO

DEL CONTE, Murilo Viotto, M.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Uso de modelos mistos no melhoramento da soja para teores de óleo e proteína nos grãos.** Orientador: Felipe Lopes da Silva. Coorientadores: Marcos Deon Vilela de Resende e Pedro Crescêncio Souza Carneiro.

A soja é uma commodity de grande importância no agronegócio brasileiro e é amplamente cultivada em território nacional. Os teores de óleo e proteína fazem com que os grãos sejam de grande interesse na alimentação humana e animal. Apesar da importância que se dá a qualidade de grãos, os teores de óleo e proteína são negativamente correlacionados e dificultam os ganhos simultâneos para ambos os caracteres. Para o sucesso de programas de melhoramento, uma fase crucial é a escolha das melhores populações. No entanto, a literatura não apresenta muitos relatos de estudos sobre estratégias para seleção entre populações F_1 . O presente estudo teve como objetivo investigar estratégias de seleção para populações F_1 em soja para altos teores de óleo e proteína nos grãos, sendo estas, o uso de índices de seleção ou a seleção via capacidade geral de combinação; investigar as correlações genéticas entre componentes crescimentos, de produção e de qualidade de grãos; e as relações de causa e efeito de tais caracteres sobre o teor de óleo nos grãos. Para o teor de óleo nos grãos de soja não foi verificada variabilidade entre as capacidades de combinação das populações e, portanto a estratégia utilizada para escolha de populações foi o uso de índice de seleção. As cultivares TMG 132 RR, DM 5958 IPRO, M 8349 IPRO, P98Y30, M7739 IPRO e DM 6563 IPRO são recomendadas para serem utilizadas como genitores em um programa para ganhos em teor de óleo. O índice de coincidência apontou que 6 populações das populações foram selecionadas para alto teor de proteína por meio das duas estratégias. As cultivares BMX Apolo RR, BMX Desafio RR, P98Y30, TMG 7062 IPRO, M 8221 RR e P98N71 são recomendadas para serem utilizadas como genitores em um programa de melhoramento para ganhos no teor de proteína nos grãos. Notou-se uma tendência de plantas de ciclo mais prolongado apresentarem maior produção de grãos por planta pelas altas correlações dos caracteres dias para maturação x número de nós, número de nós x número de vagens por planta e número de vagens x massa de grãos por planta. Verificou-se pela análise de trilha a relação de causa e efeito entre o peso de grãos por planta e teor de óleo nos grãos. O caractere teor de proteína apresentou alta correlação negativa e efeito direto baixo de mesmo sinal sobre o teor de óleo, logo devem ser empregados índices de seleção multivariados para obter ganhos genéticos sobre o teor de óleo. A exclusão de caracteres da análise de trilha é a alternativa

mais simples para contornar os problemas com a multicolinearidade e neste estudo permitiu a obtenção das estimativas mais confiáveis.

ABSTRACT

DEL CONTE, Murilo Viotto, M.Sc., Universidade Federal de Viçosa, February, 2017. **Use of mixed models in soybean breeding to grain oil and protein content.** Advisor: Felipe Lopes da Silva. Co-advisors: Marcos Deon Vilela de Resende and Pedro Crescêncio Souza Carneiro.

Soybean is an important commodity in the Brazilian agribusiness and the crop is widely cultivated in this country. The oil and protein content make the grains of great interest in the human and animal feed. Despite of the importance given to the quality of grains, oil and protein content are negatively correlated and this makes it difficult to obtain simultaneous gains on both characters. To breeding programs being successful, a crucial step is the choice of the best populations. However, it is not found in the bibliography many reports about selection strategies of F_1 populations. This study aimed to investigate F_1 selection strategies of populations to improve the oil and protein content, these being, the use of selection index and populations selection by specific combining ability; investigate the genetics correlations among growth, yield and grains quality component; and direct and indirect effects of these characters on oil content in grains. Was not verified variability of specific combining ability among populations to oil content, therefore we used index selection as a strategy of population selection. The cultivars TMG 132 RR, DM 5958 IPRO, M 8349 IPRO, P98Y30, M7739 IPRO e DM 6563 IPRO are recommended to be used as genitors in a breeding program to improve oil content in grains. The coincidence index indicated that 6 of all populations were selected to high protein content in grains by both strategies. The cultivars: BMX Apolo RR, BMX Desafio RR, P98Y30, TMG 7062 IPRO, M 8221 RR e P98N71 are recommended to be used as genitors in breeding program that aims to improve the protein content in grains. There was a tendency of late plants to show higher grain yield per plant which is explained by high correlations of characters maturation days x number of nodes, number of nodes x number of pods per plant and number of pods x grain yield per plant. It was verified through the path analysis the relations of cause and effect among grain yield per plant and grain oil content. The character grain protein content showed high negative correlation and low negative indirect effect on grain oil content, wherefore, multivariate selection indexes must be used to obtain gains on grain protein content. The exclusion of characters of the analysis is a simpler alternative to work around multicollinearity problems and in this case it provided the more reliable.

SUMÁRIO

1 INTRODUÇÃO GERAL	1
1.1 Cultura da soja	1
1.2 Aptidão da cultura da soja para teor de óleo e proteína.....	2
1.3 Melhoramento genético da soja para produtividade e teor de óleo	4
1.4 Uso da seleção indireta de caracteres e análise de trilha no melhoramento para teor de óleo e proteína nos grãos	6
2 REFERÊNCIAS BIBLIOGRÁFICAS	8
3 CAPÍTULO I – ESTRATÉGIAS DE SELEÇÃO PARA COMPONENTES DE QUALIDADE DO GRÃO EM POPULAÇÕES F1 DE SOJA COM O USO DE MODELOS MISTOS.....	15
3.1 RESUMO.....	15
3.2 ABSTRACT	16
3.3 INTRODUÇÃO.....	17
3.4 MATERIAL E MÉTODOS	18
3.5 RESULTADOS E DISCUSSÃO.....	22
3.6 CONCLUSÕES	31
3.7 REFERÊNCIAS BIBLIOGRÁFICAS	32
4 CAPÍTULO II – CORRELAÇÕES GENÉTICAS E ANÁLISE DE TRILHA PARA TEOR DE ÓLEO EM GRÃOS DE SOJA VIA VALORES GENOTÍPICOS PREDITOS (BLUP)..	36
4.1 RESUMO.....	36
4.2 ABSTRACT	37
4.3 INTRODUÇÃO.....	38
4.4 MATERIAL E MÉTODOS	39
4.5 RESULTADOS E DISCUSSÃO.....	42
4.6 CONCLUSÕES	55
4.7 REFERENCIAL BIBLIOGRÁFICO	56
5 CONCLUSÕES GERAIS.....	60

1 INTRODUÇÃO GERAL

1.1 Cultura da soja

A soja foi pioneiramente descrita por Linnaeus (1754) como *Phaseolus maxe Dolichos soja* com base em espécies descritas pelo próprio autor. Posteriormente, foi classificada como pertencente ao reino *Plantae*, divisão *Magnoliophyta*, classe *Magnoliopsida*, ordem *Fabales*, família *Fabaceae*, subfamília *Faboideae*, gênero *Glycine*, espécie *Glycine max* (L.) Merrill.

De acordo com Fukuda (1933), estudos baseados na literatura Chinesa antiga, na distribuição geográfica das espécies dos ancestrais, nos níveis de diversidade genética entre variedades de soja e nas evidências arqueológicas, indicam que a China é o centro diversidade da espécie. Xu et al. (1989) destacam a região centro-sul da china como centro primário e a região da Manchúria, onde a soja foi domesticada, como centro secundário.

Os primeiros registros encontrados na literatura sobre a soja, datam o período entre 2883 e 2838 aC, quando, ao lado do arroz, do trigo, da cevada e do milho, a soja era considerada um grão sagrado. A mais antiga referência à soja consta no herbário Pen Ts'ao Kong Mu, que descrevia as plantas da China ao Imperador Sheng-Nung (Medeiros, 2004).

O primeiro registro do cultivo da soja na Europa data de 1739 em um jardim de Paris para estudos botânicos, sendo disseminada por intermédio de missionários e navegadores (Bornetti, 1981). Em 1765, ocorreu o primeiro relato do cultivo dessa leguminosa em território norte americano, nos Estados Unidos, na Pensilvânia, com finalidade forrageira. A área cultivada para grãos só ultrapassou a destinada a cultivo da espécie como forrageira no ano de 1941 (Chung e Singh, 2000; Miyasaka e Medina, 1981). Em 1958, iniciaram-se os primeiros cultivos comerciais no México, no estado de Sonora e, em 1964, no estado de Sinaloa (Bonetti, 1981; Sedyama et al., 2005).

No Brasil, a cultura da soja foi introduzida inicialmente por Gustavo D'utra, em 1882, no estado da Bahia. Em 1908, imigrantes japoneses trouxeram consigo variedades de soja do Japão para o Estado de São Paulo. Entretanto, foi no Rio Grande do Sul onde a soja obteve maior êxito a partir da década de 1960, após sua introdução em 1901, devido à localização geográfica de latitude semelhante à região sul dos EUA, de onde era proveniente a maior parte do material genético disponível no país (Embrapa, 2004; Miyasaka e Medina, 1981).

Até a década de 70, a região Sul brasileira foi a maior produtora do grão em território nacional. No entanto, a partir da década de 80, o cultivo da soja estendeu-se ao cerrado brasileiro, que atualmente, é a região de maior produção do grão (Paludzyszyn Filho et al., 1993). Isso só foi possível após o empenho de melhoristas que desenvolveram cultivares de soja adaptadas, sobretudo, às condições de fotoperíodo do Centro-Oeste brasileiro por meio da incorporação de genes que atrasam o florescimento sob essas condições, resultando em cultivares com período juvenil longo, ao contrário das cultivares adaptadas à região sul do Brasil que possuem período juvenil curto. (Campelo et al., 1999; Medeiros, 2004; Wysmierski, 2010).

Atualmente a cultura da soja (*Glycine max* (L.) Merrill) é uma *commodity* de grande importância comercial para a economia brasileira (Silva et al., 2006). Além da posição do destaque que o grão ocupa no mercado, o desenvolvimento de tecnologias para a produção e a alta demanda e liquidez, culminaram no expressivo aumento de área e produtividade. De acordo com dados da – Companhia Nacional de Abastecimento (Conab), 33,2 milhões de hectares do território brasileiro foram cultivados com a cultura da soja na safra 2015/16 e 95,4 milhões de toneladas do grão foram colhidas, com produtividade média de 2.870 kg ha⁻¹ (Conab, 2016)

Entre as cinco regiões brasileiras, as maiores produtoras de soja são as regiões Centro-Oeste e Sul, as quais, na safra 2015/16, foram responsáveis por 45,8% e 36,9% da produção brasileira de grãos, respectivamente.

1.2 Aptidão da cultura da soja para teor de óleo e proteína

De acordo com a Associação Brasileira das Indústrias de Óleos Vegetais (Abiove), da soja produzida em território brasileiro, 42,1 % é destinada ao processamento, sendo aproximadamente 31,9 % transformada em farelo e outros 8,33 % em óleo. Das 8,05 mil toneladas de óleo de soja produzidos no Brasil, 6,5 mil toneladas são destinadas ao uso doméstico (Abiove, 2016).

O grão de soja apresenta em sua composição em média 40% de proteínas e 20% de óleo (Lazzarotto e Hirakuri, 2010; Bezerra, 2015) e apresenta uma composição química quase completa para alimentação, pois é fornecedor de proteínas, ácidos graxos saturados e insaturados, algumas vitaminas e composto polifenólicos, tais como as isoflavonas (Ávila et al., 2007).

De acordo com dados da Abiove (2016) a produtividade de óleo de soja (tonelada de óleo produzido por tonelada de grãos colhidos) não tem evoluído ao longo dos últimos anos. Dados da Conab confirmam a estagnação da produtividade de óleo de soja, já que a importação do produto processado tem aumentado proporcionalmente ao aumento do consumo nacional (Conab, 2016).

Um dos fatores que tem contribuído para a ocorrência deste fato é o estreitamento da base genética devido ao emprego de um pequeno número de genótipos como genitores no desenvolvimento de cultivares de soja. A base genética das cultivares de soja semeadas em território brasileiro no ano de 1985 foram representadas por 26 ancestrais, sendo que 89% do conjunto gênico foi representado por apenas 11 destes (Hiromoto e Vello, 1986). Wysmierski e Vello (2013) confirma, com estudos mais recentes, o estreitamento da base genética da soja no Brasil.

As variações encontradas para os teores de óleo e proteína em grãos de soja podem ser devidos às condições de cultivo ou o genótipo cultivado. Fatores ambientais tais como temperatura e estresse hídrico estão relacionados com o teor de óleo e proteína nos grãos, de maneira que altas temperaturas culminam na redução do teor de óleo e aumento no teor de proteína (Pípolo et al., 2002; Rangel et al., 2004). Harris et al., 1978 afirmam em seu trabalho que o estresse hídrico afeta a produção de óleo através de seus efeitos sobre o crescimento e o desenvolvimento do grão (Harris et al., 1978).

Quanto à influência dos nutrientes na qualidade dos grãos de soja, destaca-se o potássio, que é um dos macronutrientes de maior absorção e exportação pela cultura, apresenta grande importância para o aumento de óleo no grão, além de influenciar positivamente a formação de nódulos na raiz, germinação, vigor e qualidade da semente (Yamada e Borkert, 1992; Sfredo et al., 2008). O enxofre destaca-se como um dos macronutriente secundários que quando mais disponíveis, culminaram no aumento não apenas dos teores de óleo, mas também dos teores de proteína, bem como incremento de produtividade (Chadel et al., 1989). Tanaka e Cargnelutti Filho (2002) relatam que a elevação do pH do solo em decorrência de crescentes doses de calcário, proporcionaram maiores níveis de fixação de nitrogênio pelo sistema raiz-rizóbio e por fim, a redução do teor de óleo e aumento do teor de proteína nos grãos. Os autores ainda relatam não haver efeito do gesso agrícola sobre a qualidade dos grãos.

De acordo com Ávila et al. (2007), o teor de óleo e proteína no grão de soja, é fortemente influenciado pelo ambiente de cultivo, sobretudo, no período de enchimento de

grãos, sendo o teor de proteína, mais vulnerável às variações ambientais (Benzain e Lane, 1986).

O principal interesse sobre o teor de proteína nos grãos de soja está relacionado a alimentação humana e animal. Conforme citado anteriormente 31,9% da soja produzida em território brasileiro é transformado em farelo e utilizado para suplementar a alimentação de animais (Zambom et. al, 2008; Mendes et al., 2010). Além disto, segundo Moraes et al. (2006) o farelo de soja pode ser destinado a exportação e para isto o produto é classificado de acordo com o teor de proteína: HyPro (>48%), Normal (46%) e LowPro (<43,5). Os autores afirmam que para obtenção de farelo de soja Normal e HyPro, são necessárias cultivares que apresentem acima de 41,5% e 43% de proteína nos grãos, respectivamente, com base na matéria seca.

Tanto o alto teor de óleo e quanto o alto teor proteína nos grãos são importantes para suprir a demanda de mercado. De acordo com a literatura, estas duas características são negativamente correlacionadas, devido à concorrência destas duas vertentes pelos esqueletos de carbono (Hanson, 1992; Pípolo 2002). Diante disto, os ganhos para altos teores de óleo e proteína simultaneamente são dificultados.

1.3 Melhoramento genético da soja para produtividade e teor de óleo

Os estudos na área de melhoramento genético vegetal têm dado fortes indícios de que os caracteres de maior importância para o melhorista são, em sua grande maioria, controlados por vários genes, como a produtividade e teor de óleo em grãos de soja. Sendo assim, o processo de melhoramento de tais caracteres se dá pelo acúmulo de alelos favoráveis, o que pode ser alcançado por meio da seleção recorrente.

Neste contexto, o uso da seleção recorrente em programas de melhoramento para qualidade de grãos de soja é uma alternativa para alavancar os ganhos de caracteres como produtividade e teor de óleo nos grãos.

A seleção recorrente tem como objetivo o aumento gradativo da frequência de alelos favoráveis envolvidos no controle dos caracteres quantitativos sob seleção, sem exaurir a variabilidade genética (Geraldí, 2005). De acordo com Ramalho et al. (2012), seleção recorrente é definida por qualquer processo cíclico de melhoramento composto pela obtenção de progênies, avaliação e recombinação das de maior mérito.

Para a obtenção das primeiras progênes a comporem um programa de seleção recorrente, é necessária a escolha de genitores para que sejam realizadas as hibridações. Borém e Miranda (2013) citam vários métodos para seleção de genitores, dentre eles, os cruzamentos em esquemas dialélicos, os quais fornecem informações suficientes para seleção não só de genitores, mas também de populações de bom mérito por meio das estimativas de capacidade geral e específica de combinação. A obtenção e avaliação do potencial de populações segregantes é uma etapa crucial (Kotzamanidis et al., 2008), pois está relacionada ao sucesso do programa de melhoramento (Kurek et al., 2001). Neste contexto, os delineamentos dialélicos se apresentam como metodologia interessante para adoção em programas de melhoramento genético.

Além disto, os cruzamentos dialélicos merecem destaques pelas inúmeras pesquisas e adaptações de métodos para situações e objetivos específicos, tais como generalizações para análise de dialelos desbalanceados estatisticamente, dialelos circulantes (Kempthorne e Curnow, 1961), dialelos parciais (Miranda Filho e Geraldi, 1984) e cruzamentos fatoriais (Comstock e Robinson, 1948).

Dentre as metodologias citadas, os cruzamentos dialélicos parciais se baseiam em cruzamentos entre genitores de grupos distintos, tem como resultado um menor número de populações híbridas sem reduzir o número de genitores, permitem formar grupos de genitores de modo que estes sejam mais contrastantes entre grupos do que dentro de grupos e assim, explorar mais amplamente a variabilidade genética.

A busca por metodologias estatísticas que permitem a seleção mais acurada de populações segregantes se faz necessária. De acordo com Resende (2007), existem dois principais objetivos na seleção de genótipos: prever os valores genotípicos dos genótipos e ranquear os mesmos com base nos valores genotípicos preditos.

Conforme citado por Duarte e Vencovsky (2001), o melhoramento vegetal tem se baseado mais usualmente em modelos fixos para a estimação de médias de genótipos, mesmo quando estes são tomados como amostra de uma população, o que, neste caso, a modelagem baseada em componentes de efeito fixo (ex: bloco) e aleatório de genótipos, modelos denominados mistos, seria a mais correta.

A metodologia de modelos mistos foi proposta por Charles Roy Henderson em 1949 tendo sido adotada primeiramente no melhoramento animal (Henderson, 1975). Enquanto a análise e interpretação de modelos fixos resultam em componentes de variância por meio do método dos mínimos quadrados, além de pressupor distribuição normal dos dados, modelo de

aditividade, e independência e homogeneidade dos erros, a adoção de modelos mistos por meio das equações de modelos mistos, permite a estimação dos componentes de variância e das variâncias simultaneamente, sem a necessidade de atender às mesmas premissas (Carvalho et al. 2008).

Neste sentido, Resende (2007) recomenda o uso da técnica REML/BLUP para estimação e predição de componentes de variância e valores genéticos respectivamente. Para a estimação dos valores genotípicos o método denominado melhor preditor linear não viesado (BLUP) desenvolvido por Henderson em 1949 é indicado como um procedimento ótimo para predição de valores genéticos por Resende (2004).

Contudo, são necessárias metodologias multivariadas que conduzam a escolha entre os melhores genótipos, de acordo com Pedrozo et al., 2009 e Cruz et al. (2014), a seleção com base em apenas um caractere culmina na obtenção de genótipos superior apenas para o caractere sob seleção. O índice proposto por Smith (1936) e Hazel (1943), conhecido ainda como índice clássico, baseia-se na adoção de pesos econômicos para cada caractere estudado de acordo com sua importância ou de acordo com parâmetros genéticos. Desde então, vários índices foram propostos ou adaptados para suprir a demanda dos pesquisadores e programas de melhoramento para a seleção simultânea de caracteres. (Cruz et al., 2014).

Para a construção do índice de seleção clássico, uma tomada de decisão muito importante é a escolha dos pesos econômicos que atuarão como fatores de ponderação sobre cada variável. Discussões acerca da seleção de genótipos e estimativas de ganhos percentuais por meio da utilização dos índices de Smith e Hazel, índice de Pesek e Baker, índice de Mulamba e Mock e índice da distância do genótipo ao ideótipo (Cruz, 2013) ponderados por diferentes pesos econômicos foram realizadas por Silva e Viana (2002). É importante também a escolha das variáveis a serem utilizadas para compor o índice de seleção. Cruz et al. (2012) destacam que variáveis que apresentam alta correlação, porém, baixo efeito direto ou alto efeito indireto sobre a variável principal, são passíveis de compor um índice de seleção multivariado.

1.4 Uso da seleção indireta de caracteres e análise de trilha no melhoramento para teor de óleo e proteína nos grãos

Caracteres quantitativos são controlados por vários genes e, de forma geral, sofrem forte influência ambiental e apresentam na maioria dos casos, baixa herdabilidade (Cruz,

2005). Na cultura da soja, tem-se como exemplos de caracteres quantitativos a produtividade de grãos, o teor de óleo e a proteína nos grãos (Marquez, 2010). A baixa herdabilidade constitui um fator perturbador para a seleção, já que quanto menor a herdabilidade, maior a fração da variação ambiental contida na variação total. Neste sentido, a seleção indireta sobre um caractere de alta herdabilidade, de fácil mensuração e fortemente correlacionado, tem grandes chances de resultar em ganhos mais rápidos do que a seleção direta sobre tal caractere (Carvalho et al., 2004).

O estudo das correlações entre caracteres é indispensável em estudos genéticos, pois este mede o grau de associação entre os mesmos e elucida a viabilidade da seleção indireta (Cruz et al., 2012). A seleção indireta apenas acontece devido à alta correlação entre dois ou mais caracteres, devido ao desequilíbrio de ligação, ou à pleiotropia. O desequilíbrio de ligação é causa transitória das correlações genéticas, principalmente em populações derivadas de cruzamentos entre linhagens divergentes, enquanto a pleiotropia constitui uma causa permanente (Falconer, 1987).

Vencovsky e BARRIGA (1992) ressaltam que as correlações entre caracteres são apenas uma medida de associação e não contemplam as relações de causa e efeito. Para entender as relações de causa e efeito, Wright (1921) propôs o uso da análise de trilha (*path coefficient analysis*) que, em suma, é uma expansão de uma regressão linear múltipla e permite desdobrar os coeficientes de correlação em efeitos diretos e indiretos sobre uma variável básica (Alcantra Neto et al., 2011; Cruz et al., 2012).

Para uma boa análise de trilha, Cruz et al. (2014) aconselham que seja realizado o diagnóstico de multicolinearidade na matriz de correlação das variáveis explicativas, a qual, de acordo com Neter e Wasserman (1974) ocorre quando as variáveis apresentam correlações altas ou perfeitas. A análise de trilha pode ser realizada mesmo com a presença de multicolinearidade, porém, neste caso, as estimativas dos efeitos diretos e indiretos, o que dificulta a interpretação biológica (Toebe e Cargnelutti Filho, 2013; Cruz et al., 2014). Então, para obtenção de estimativas confiáveis, Cruz et al. (2014) recomendam a exclusão de alguns caracteres do estudo, sendo estes, os que apresentam altas correlações e que contribuem mais para a presença de multicolinearidade.

O método mais comumente citado na literatura e utilizado nos estudos de diagnóstico de multicolinearidade é o que se baseia no número de condição (NC). O NC é a razão entre o maior e o menor autovalor da matriz de correlação das variáveis explicativas (Montgomery e Peck, 1981). Valores de NC abaixo de 100 significam pouco ou nenhum problema de

multicolinearidade; valores entre 100 e 1000, indica multicolinearidade fraca a moderada; e valores acima de 1000, indicam presença de multicolinearidade severa.

Como alternativa à estratégia de excluir alguns caracteres do estudo, Cruz et al. (2014) recomendam o uso de regressão em crista, que tem como fundamento permitir obter soluções para o vetor de parâmetros, mesmo sob multicolinearidade, pela adição de pequenos valores constantes (k) na matriz de correlação, conforme proposto por Hoerl e Kennard (1970a, b) até a estabilidade das estimativas.

Cruz et al. (2014) afirmam que a estabilização de estimativas em análises de trilha com multicolinearidade ocorre com a adição de pequeno valores de k somados à diagonal da matriz de correlação, porém quanto maior o valor de k adotado, maior será o viés da análise de regressão e por isso, cabe ao pesquisador escolher o menor valor de k suficiente para estabilização das estimativas.

2 REFERÊNCIAS BIBLIOGRÁFICAS

ABIOVE. Estatística. 2016. Disponível em: <<http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=>>>. Acesso em: 23 ago. 2016.

ALCANTARA NETO, F. et al. Análise de trilha do rendimento de grãos de soja na microrregião do Alto Médio Gurguéia. **Comunicata Scientiae**, v. 2, n. 2, p. 107-112, 2011.

ÁVILA, M. R. et al. Componentes do rendimento, teores de isoflavonas, proteínas, óleo e qualidade de sementes de soja. **Revista Brasileira de Sementes**, v. 29, n. 3, p. 111-127, 2007.

BENZIAN, Blanche; LANE, Peter W. Protein concentration of grain in relation to some weather and soil factors during 17 years of english winter-wheat experiments. **Journal of the Science of Food and Agriculture**, v. 37, n. 5, p. 435-444, 1986.

BEZERRA, A., R., G. et al. Importância Econômica. In: SEDIYAMA, Tuneo; SILVA, Felipe; BORÉM, Aluizio. **Soja: do plantio à colheita**. Viçosa: Editora Ufv, 2015. Cap. 1, p. 13.

BONETTI, L. P. Distribuição da soja no mundo : origem, história e distribuição. In : MIYASAKA, S.; MEDINA, J.C. (Ed.). **A soja no Brasil**. Campinas : ITAL, p. 1-6, 1981.

- BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 523p. UFV, 2013.
- BORNETTI, L.P. Distribuição da soja no mundo, In: MIYASAKA, S.; MEDINA, J. C. (Ed) **A soja no Brasil**. Campinas: ITAL, p.1-6, 1981.
- CAMPELO, G.J.A.; KIIHL, R.A.S.; ALMEIDA, L.A. Características agronômicas e morfológicas das cultivares de soja desenvolvida para as regiões de baixas latitudes. In: QUEIROZ, M.A.; GOEDERT, C.O.; RAMOS, S.R.R (Ed.). **Recursos genéticos e melhoramento de plantas para o Nordeste brasileiro**. (on line). Versão 1.0. Petrolina – PE: EMBRAPA Semi-Árido/Brasília-DF: EMBRAPA Recursos Genéticos e Biotecnologia, nov. 1999. Disponível em [HTTP://www.cpatsa.EMBRAPA.br](http://www.cpatsa.EMBRAPA.br). Acesso em 4 set. 2016.
- CARVALHO, F.I.F.; LORENCETTI, C.; BENIN, G. **Estimativas e implicações da correlação no melhoramento vegetal**. Pelotas: Editora e Gráfica da UFPel, 2004. 142p.
- CHADEL, A.S.; RAO, G.P.; SAXENA, S.C. Effect of sulfur nutrition on soybean (*Glycine max* L. Merrill). In: **CONFERENCIA MUNDIAL DE INVESTIGACION EN SOJA**, 4. Buenos Aires, 1989. Actas...Buenos Aires, AAS, 1989. p.363-368.
- CHUNG, G.; SINGH, R.J. Broadening the genetic base of soybean: a multidisciplinary approach. **Critical Reviews in Plant Sciences**, Boca Raton, v.40, n.5, p.1470-1481, 2000.
- COMSTOCK, R. E.; ROBINSON, H. F. The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. **Biometrics**, v. 4, n. 4, p. 254-266, 1948.
- CONAB. Série Histórica – Soja. 2016. Disponível em: <http://www.conab.gov.br/conteudos.php?a=1252&t=>>. Acesso em: 27 set. 2016.
- CRUZ, C. D. **Princípios de genética quantitativa**. UFV, 2005.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4 ed. Viçosa: UFV, 2012. v. 1. 514 p.
- CRUZ, C. D.; CARNEIRO, P. C. S. ; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 3 ed. Viçosa: UFV, 2014. v. 2. 668 p.

CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**. v.35, n.3, p.271-276, 2013.

DUARTE, J. B.; VENCOVSKY, R. Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. **Scientia Agricola**, v. 58, n. 1, p. 109-117, 2001.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária – Centro Nacional de Pesquisa de Soja. **Tecnologia de produção de soja região central do Brasil 2004**: a soja no Brasil. 2004. Disponível em <<http://www.cnpso.EMBRAPA.br/producaosoja/SojanoBrasil.htm>>. Acesso em 27 set. 2016.

FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, 279p. 1987.

FUKUDA, Y. Cytogenetical studies on the wild and cultivated Manchurian soybeans (*Glycine* L.). **Jpn J Bot**, v. 6, p. 489-506, 1933.

GERALDI, I. O. **Por que realizar seleção recorrente**. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 9., 2005, Lavras. Anais.:. Lavras: UFLA, Programa de Pós-Graduação em Agronomia/Genética e Melhoramento de plantas, 2005. p. 1-8.

HANSON, W.D. Modified seed maturation and seed yield potentials in soybean. **Crop Science**, Madison, v.32, n.4, p.972-976, 1992.

HARRIS, H.C.; McWILLIAM, J.R.; MASON, W.K. Influence of Temperature on Oil Content and Composition of Sunflower Seed. **Australian Journal of Agricultural Research**, Collingngwood, v. 29, n.3, p.1203-1212, 1978.

HAZEL, L. N. The genetic basis for constructing selection indexes. **Genetics**, v. 28, n. 1, p. 476-490, 1943.

HENDERSON, C. R. Estimation of changes in herd environment. **Journal of Dairy Science**, v. 32, p. 709, 1949.

HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v.31, p.423-447, 1975.

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

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55-67, 1970a.

HOERL, Arthur E.; KENNARD, Robert W. Ridge regression: applications to nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 69-82, 1970b.

JORGE, Neuza et al. Alterações físico-químicas dos óleos de girassol, milho e soja em frituras. **Química Nova**, v. 28, n. 6, p. 947, 2005.

KEMPTHORNE, O; CURNOW, R. N. The partial diallel cross. **Biometrics**, v. 17, n. 2, p. 229-250, 1961.

KOTZAMANIDIS, S. T. et al. Prediction criteria of promising F₃ populations in durum wheat: a comparative study. **Field Crops Research**, v.107, p.257-264, 2008.

KUREK, A. J. et al. Capacidade combinatória como critério de eficiência na seleção de genitores em feijoeiro. **Pesquisa Agropecuária Brasileira**, v.36, p.645-651, 2001.

LAZZAROTTO, J. J.; HIRAKURI, M. H. Evolução e perspectivas de desempenho econômico associadas com a produção de soja nos contextos mundial brasileiro. Londrina: **Embrapa Soja**, p. 46, 2010. (Embrapa Soja. Documentos, 319).

LINNAEUS, C. **Genera Plantarum**, Fifth ed.; Lars Salvius: Stockholm. 1754.

MARQUEZ, M.C. **Adaptabilidade, estabilidade e diversidade genética de cultivares de soja em três épocas de semeadura em Uberlândia – MG**. 2010. 84p. Tese (Mestrado) – Universidade Federal de Viçosa, Viçosa, MG, Brasil.

MEDEIROS, L. Comparativo de custos de produção entre a soja convencional e a soja transgênica na safra 2002/2003-Rio Grande do Sul. **ConTexto**, v. 3, n. 5, 2004.

MENDES, C. Q. et al. Substituição parcial do farelo de soja por ureia ou amireia na alimentação de cabras em lactação. **Revista Brasileira de Zootecnia**, v. 39, n. 8, p. 1818-1824, 2010.

- MIRANDA FILHO, J. B.; GERALDI, I. O. An adapted model for the analysis of partial diallel crosses. **Revista Brasileira de Genética**, v. 7, n. 4, p. 677-688, 1984.
- MIYASAKA, S.; MEDINA, J. C. A **Soja no Brasil**. 1 ed. Campinas: Instituto de Tecnologia de Alimentos, p.65-104, 1981.
- MONTGOMERY, D. C.; PECK, E. A. **Introduction to linear regression analysis**. New York: Wiley, 1981. 504p.
- MORAES, R. M A. de et al. Biochemical characterization of high protein soybean lines. **Pesquisa Agropecuária Brasileira**, v. 41, n. 5, p. 715-729, 2006.
- NETER, J.; WASERMAN, W. **Applied linear statistical models**. Homewood: Richard D. Irwin, 1974. 842 p.
- PALUDZYSZYN FILHO, E.; KIIHL, R. S.; ALMEIDA, L.A. Desenvolvimento de cultivares de soja na região Norte e Nordeste do Brasil. In: ARANTES, N.E.; SOUZA, P.I. M. (Ed.). **Cultura da soja nos cerrados**. Piracicaba: Potafos, 1993. p. 255-266.
- PEDROZO, C. A et al. Efficiency of selection indexes using the REML/BLUP procedure in sugarcane breeding. **Scientia Agraria**, v. 10, n. 1, p. 31-36, 2009.
- PÍPOLO, A. E. **Influência da temperatura sobre as concentrações de proteína e óleo em sementes de soja (*Glycine max* (L.) Merrill)**. Universidade de São Paulo: USP, Escola Superior de Agricultura “Luiz de Queiroz”, 2002. 128p. (Tese Doutorado).
- RAMALHO, M. A. P. et al. Aplicações da genética quantitativa no melhoramento de plantas autógamas. **Lavras: UFLA**, 2012.
- RANGEL, M. A. S. et al. **Efeito do genótipo e do ambiente sobre os teores de óleo e proteína nos grãos de soja, em quatro ambientes da Região Sul de Mato Grosso do Sul, safra 2002/ 2003**. Dourados: Embrapa Agropecuária Oeste, 2004. (Boletim de pesquisa e desenvolvimento, 17).
- RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. Embrapa Florestas, 2004.

RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Embrapa Florestas, Colombo. 561 p. 2007a.

RESENDE, M. D. V. Software **SELEGEN-REML/BLUP**: sistema estatístico e seleção computadorizada via modelos lineares mistos. Colombo: Embrapa Florestas, 2007b.

SEDIYAMA, T.; TEIXEIRA, R de C.; REIS, M. S.; Melhoramento da Soja. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, p.553-604. 2005.

SFREDO, G. J. **Soja no Brasil**: calagem, adubação e nutrição mineral. Embrapa Soja, 2008.

SILVA, J. M. et al. Determinação da necessidade de calagem para a soja em dois sistemas de preparo do solo em área de cerrado. In: **VI Encontro Latino Americano de Pós Graduação**, 2006, São José dos Campos. Revista Univap, 2006. v. 13. p. 2877-2880.

SILVA, M. G. de M.; VIANA, A. P. Alternativas de seleção em população de maracujazeiro-azedo sob seleção recorrente intrapopulacional. **Revista Brasileira de Fruticultura, Jaboticabal**, v. 34, n. 2, p. 525-531, 2012.

SMITH, H. F. A. discriminant function for plant selection. **Annals of Eugenics**, v. 7, n. 1, p. 240-250, 1936.

TANAKA, R. T.; MASCARENHAS, HAA. Resposta da soja à aplicação de gesso agrícola. **Instituto Agrônomo, Campinas**, 2002.

TOEBE, M.; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, v. 48, n. 5, p. 466-477, 2013.

VENCOVSKY, R.; BARRIGA, Patrício. **Genética biométrica no fitomelhoramento**. 1992.

WRIGHT, S. Correlation and causation. **Journal of agricultural research**, V. 20, N. 7, P. 557-585, 1921.

WYSMIERSKI, P. T.; VELLO, N. A. The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. **Genetics and molecular biology**, v. 36, n. 4, p. 547-555, 2013.

WYSMIERSKI, P. T. **Constituição genética dos ancestrais da soja às cultivares brasileiras**. 2010. 99 p. Dissertação (Mestrado em Melhoramento de Plantas) – Universidade de São Paulo – Escola Superior de Agricultura “Luiz de Queiroz”, 2010.

XU, B. et al. Three new evidences of the original area of soybean. 1989. In: Conferencia mundial de investigacion en soja, 4, 1989, Buenos Aires. **Actas...** Buenos Aires, t.1, p.124-128, 1989.

YAMADA, T.; BORKERT, C.M. Nutrição e produtividade da soja. In: Simpósio sobre a cultura e produtividade da soja, Piracicaba, 1992. **Anais**. Piracicaba: USP, ESALQ; FEALQ, 1992.p.180-212.

ZAMBOM, Maximiliane Alavarse et al. Valor nutricional da casca do grão de soja, farelo de soja, milho moído e farelo de trigo para bovinos. **Acta Scientiarum. Animal Sciences**, v. 23, n. 1, p. 937-943, 2001.

3 CAPÍTULO I – ESTRATÉGIAS DE SELEÇÃO PARA COMPONENTES DE QUALIDADE DO GRÃO EM POPULAÇÕES F1 DE SOJA COM O USO DE MODELOS MISTOS

3.1 RESUMO

Poucas estratégias são relatadas na literatura para a escolha de populações F₁. Geralmente, esta seleção se baseia na capacidade específica de combinação das populações e na capacidade geral de combinação dos genitores para dado caractere. Como a seleção baseada em um caractere pode conduzir a obtenção de genótipos superiores apenas para o caractere em questão, o uso de índices de seleção é uma boa alternativa para estes fins. Neste sentido, o presente trabalho teve como objetivo estimar parâmetros genéticos e valores genotípicos via REML/BLUP e comparar as estratégias de seleção de populações F1 de soja, oriundas de um dialelo parcial, utilizando as estimativas das capacidades combinatórias ou o índice de seleção, visando aumento dos teores de óleo e proteína nos grãos. A correlação estimada entre o teor de óleo e o teor de proteína nos grãos foi de -0,69. Já a correlação estimada entre o teor de óleo nos grãos e a massa de grãos por planta (0,43) evidencia a possibilidade de ganhos simultâneos para ambos os caracteres. Não foi verificada variabilidade para o teor de óleo nos grãos na capacidade específica de combinação e, portanto a estratégia utilizada para escolha de populações foi o uso de índice de seleção. As cultivares TMG 132 RR, DM 5958 IPRO, M 8349 IPRO, P98Y30, M7739 IPRO e DM 6563 IPRO são recomendadas para serem utilizadas como genitores em um programa para ganho em teor de óleo nos grãos. O índice de coincidência apontou que 6 populações foram selecionadas para alto teor de proteína por meio das duas estratégias. As cultivares BMX Apolo RR, BMX Desafio RR, P98Y30, TMG 7062 IPRO, M 8221 RR e P98N71 são recomendadas para serem utilizadas como genitores em um programa de melhoramento para ganhos no teor de proteína nos grãos. A seleção baseada nas capacidades de combinação apontou a escolha de populações de mais alta CEC, enquanto o índice de seleção apontou populações de maior valor genotípico predito.

Palavras-chave: *Glycine max* (L.) Merr.; índice de seleção; capacidade de combinação; teor de óleo; teor de proteína.

Selection strategies for grain quality components in F₁ soybean populations using mixed models

3.2 ABSTRACT

A few of strategies are reported in the literature to choice F₁ populations. Usually, this kind of selection is based on general combining ability of populations and specific combining ability of genitors for a given trait. As the selection based in only one trait can lead to a gain of superior genotypes to such trait. Selection indexes are a good alternative to these purposes. In this context, this study aimed estimate genetic parameters and genotypic values by REML/BLUP that is considered the best to achieving these estimates and compare selection strategies in soybean F₁ populations originated from a partial diallel, using combining ability estimates and selection index, aiming oil and protein content gains. The estimated correlation between grain oil content and protein content was -0,69. However the estimated correlation between grain oil content and mass of grains per plant (0,43) shows the possibility of achieving simultaneous gains to both traits. Was not verified variability on specific combining ability to oil content, therefore, was used selection index to choice the best populations. TMG 132 RR, DM 5958 IPRO, M 8349 IPRO, P98Y30, M7739 IPRO e DM 6563 IPRO are recommended to be used as genitor in a breeding program to grain oil content gain. The coincidence index showed that 6 populations were selected by both strategies to high grain protein content. The soybean varieties BMX Apolo RR, BMX Desafio RR, P98Y30, TMG 7062 IPRO, M 8221 RR e P98N71 are recommended to be used as genitors in a breeding program to grains protein content gains. The selection based on combining ability selected higher specific combining ability populations, while o selection index selected the higher genotypic value predict populations.

Key-words: *Glycine max* (L.) Merr.; selection index; combining ability; oil content; protein content.

3.3 INTRODUÇÃO

A soja [*Glycine max* (L.) Merrill] é uma das principais culturas oleaginosas cultivadas no mundo. Devido ao seu elevado potencial produtivo nas diferentes regiões, a cultura da soja exerce um importante papel socioeconômico para o agronegócio brasileiro. Além disso, a cultura é indispensável para impulsionar diversos complexos agroindustriais, tais como as indústrias de sementes, fertilizantes, agrotóxicos e máquinas agrícolas (Mauad et al., 2010).

Os grãos de soja apresentam em sua composição por volta de 40% de proteínas e 20% de óleo. De toda a soja produzida em território brasileiro, 42,1% é destinada ao processamento e destes, 8,33% são transformados em óleo e 31,9% em farelo (Abiove, 2016). Devido a estas propriedades, o grão de soja tem grande importância na alimentação humana e animal, de modo que o óleo se destaca pelo preço e qualidade (Ávila et al., 2007) para uso doméstico e a proteína para suplementação animal e exportação (Moraes et al., 2006).

A obtenção de cultivares superiores para caracteres como produtividade, teor de proteína e óleo nos grãos se faz necessário para suprir a demanda do agronegócio sojícola. Neste contexto, se faz necessário a obtenção de populações segregantes de maior potencial visando a otimização da utilização de recursos humanos e financeiros do programa de melhoramento (Pimentel et al., 2013).

Comumente para a obtenção de populações de melhoramento, cruzamentos em esquema dialélico são realizados. Neste sentido, a escolha das populações obtidas por cruzamentos dialélicos se baseia na avaliação das capacidades geral e específica de combinação para cada caractere alvo do programa (Daronch et al., 2014; Karyawati et al., 2015). No entanto, a seleção baseada em um ou poucos caracteres pode ser inadequada, pois conduz a seleção de genótipos superiores apenas para os caracteres sob seleção.

A seleção simultânea de caracteres pode ser realizada com o uso de índices de seleção adequados (Cruz et al., 2014). Contudo, poucos são os relatos descritos na literatura do uso de índices de seleção na escolha de populações F_1 oriundas de cruzamentos dialélicos, tais como estudados por Mendes et al., (2009) e Ndoumbé et al., (2001).

Para a construção de índices de seleção, uma tomada de decisão muito importante é a escolha dos pesos econômicos que atuarão como fatores de ponderação sobre cada variável. Discussões acerca da seleção de genótipos e estimativas de ganhos percentuais por meio da utilização dos índices clássico de Smith e Hazel, índice de Pesek e Baker, índice de Mulamba e Mock e índice da distância do genótipo ao ideótipo (Cruz, 2013) ponderados por diferentes pesos econômicos foram realizadas por Silva e Viana (2002).

Diante dessas duas estratégias de seleção de populações de interesse no melhoramento, via análise das capacidades combinatórias e via índice de seleção associados à pesos ideais para cada caractere sob seleção, se faz necessário a obtenção de valores genéticos mais próximos dos verdadeiros. Nesse sentido, a metodologia de modelos mistos REML/BLUP, proposta por Henderson (1949), tem sido fundamental para a predição de valores genéticos aditivos e genotípicos, tanto intra quanto interpopulacionalmente (RESENDE, 2000). Mesmo em condições de experimentos desbalanceados, essa abordagem permite a predição acurada e não viesada dos valores genéticos, propicia informações específicas sobre o mérito da progênie, de indivíduos nas progênies e considera até mesmo o ambiente da repetição onde está localizado um indivíduo (Resende, 2007a). Assim, as propriedades do BLUP permitem maximização da acurácia seletiva, minimização do erro de predição, predição não viciada de valores genéticos, maximização do ganho genético – por ciclo de seleção – e maximização da probabilidade de seleção do melhor entre dois ou vários genótipos (RESENDE, 2004).

Vários estudos têm buscado o aprimoramento dos métodos de seleção em plantas anuais a partir da aplicação de metodologias de modelos mistos. Ramalho & Araújo (2011) apontam a utilização de modelos mistos como estratégia adequada para maior eficiência do melhoramento de plantas autógamas pela identificação de progênies ou linhas com maior mérito genotípico.

Neste sentido, o presente trabalho teve como objetivo estimar parâmetros genéticos e valores genotípicos via REML/BLUP e comparar as estratégias de seleção de populações F1 de soja, oriundas de um dialelo parcial, utilizando as estimativas das capacidades combinatórias ou o índice de seleção, visando aumento dos teores de óleo e proteína nos grãos.

3.4 MATERIAL E MÉTODOS

As populações F1 foram obtidas por meio de cruzamentos oriundos de um dialelo parcial, conduzido entre os meses de setembro de 2015 e janeiro de 2016. Os cruzamentos foram realizados em casa de vegetação, pertencente ao Programa Soja do Departamento de Fitotecnia da Universidade Federal de Viçosa (PGMS/DFT/UFV), originando 34 populações conforme apresentadas na Tabela 1.

O dialelo parcial foi composto por dois grupos de cultivares. O primeiro (G1) foi constituído de 11 cultivares comerciais com alto rendimento de grãos, sendo destas, cinco adaptadas às condições da região centro-oeste do Brasil e seis adaptadas à região sul. O segundo grupo (G2) foi constituído por quatro cultivares com alto teor de óleo previamente selecionadas do Banco Ativo de Germoplasma do PGMS/DFT/UFV. As cultivares utilizadas no dialelo pertencem a grupos de maturação relativa entre 5.5 a 9.0 (Alliprandini, 2009).

TABELA 1 – Populações F1 obtidas por meio de cruzamentos entre dois grupos de genitores em esquema de dialelo parcial para a cultura da soja.

Grupo 1/Grupo 2	M 9056 RR	M 8221 RR	P98N71	CONQUISTA
BMX Apolo RR	Pop 1	Pop 7	Pop 13	Pop 24
DM 5958 IPRO	-	-	Pop 14	Pop 25
NA 5909 RR	Pop 2	Pop 8	Pop 15	Pop 26
TMG 7062 IPRO	-	-	Pop 16	Pop 27
DM 6563 IPRO	-	-	Pop 17	Pop 28
BMX Potência RR	Pop 3	Pop 9	Pop 18	Pop 29
BMX Desafio RR	Pop 4	Pop 10	Pop 19	Pop 30
M7739 IPRO	-	-	Pop 20	Pop 31
P98Y30 RR	Pop 5	Pop 11	Pop 21	Pop 32
M8349 IPRO	-	-	Pop 22	Pop 33
TMG 132 RR	Pop 6	Pop 12	Pop 23	Pop 34

As sementes das populações F1 e dos genitores foram semeadas no dia 22 de fevereiro de 2016 e avaliadas ao longo do ciclo em casa de vegetação. Para tanto, a casa de vegetação foi adaptada com lâmpadas para ajuste do fotoperíodo quando necessário, favorecendo um ambiente adequado para o bom desenvolvimento da soja levando em conta a importância do fotoperíodo como indutor de florescimento e fator fundamental para o desenvolvimento da cultura (Câmara, 1998).

O experimento foi conduzido sob delineamento de blocos ao acaso com cinco repetições. Cada parcela foi constituída por dois vasos contendo duas plantas cada, totalizando quatro plantas por parcela. Os tratamentos culturais foram realizados conforme recomendados por Sedyama et al. (2015).

Foram avaliados 12 caracteres, sendo estes, componentes vegetativos, produtivos e de qualidade de grãos. À medida que as plantas alcançavam o estágio fenológico R1 definido por Fehr e Caviness (1977), foram tomadas medidas das plantas individuais para os seguintes caracteres: dias para florescimento (DPF): contados a partir da emergência das plântulas; e,

altura no florescimento (AF) em cm: medida da distância da base até o ápice da planta. Já na ocasião da maturação, conforme as plantas atingiam o estágio fenológico R8 definido por Fehr e Caviness (1977), foram tomadas medidas das plantas individuais para: dias para maturação (DPM): contados a partir de VE; diâmetro do hipocótilo (DHP) em mm: medido dois centímetros abaixo do nó cotiledonar; altura na maturação (AM) em cm: medida da distância da base até o ápice da planta; número de vagens por planta (NV); número de nós na haste principal por planta (NN); número de hastes laterais por planta (NH); número de grãos por planta (NG); massa de grãos por planta (PROD) em gramas corrigida para 13% de umidade; e, teores de óleo (OL) e de proteína nos grãos (PT) definidos em porcentagem: mensurados em grãos intactos com 13% de umidade, no Laboratório de Genética Molecular de Plantas/Bioagro/UFV, via espectrometria do infravermelho próximo utilizando o equipamento Antaris™ II FT-NIR Analyzer.

A análise dialélica e estimação dos parâmetros genéticos foram procedidas via o software SELEGEN (Resende, 2007b) por meio da metodologia de modelos mistos REML/BLUP. Para tanto um modelo híbrido foi adotado, contemplando as informações de populações e dos genitores. A parte do modelo referente ao ajuste de populações é dada por:

$$y_1 = Xc + Zm + Wf + Sd + Tb + e, \text{ onde:}$$

c é o efeito fixo da média geral dos cruzamentos; m é o vetor dos efeitos das capacidades gerais de combinação dos genitores com lato rendimento de grãos (aleatórios); f é o vetor dos efeitos das capacidades gerais de combinação dos genitores com alto teor de óleo (aleatórios); d é o vetor dos efeitos das capacidades específicas de combinação dos cruzamentos (aleatórios); b é o vetor dos efeitos de blocos (aleatórios); e é o vetor dos efeitos do erro (aleatórios). X, Z, W, S, e T são matrizes de incidência.

A parte do modelo referente ao ajuste para genitores é dada por:

$$y_2 = Xu + Hg + Tb + e, \text{ onde:}$$

u é o efeito fixo da média geral dos genitores; g é o vetor dos efeitos genéticos aditivos dos genitores (aleatórios); b é o vetor dos efeitos de blocos (aleatórios); e é o vetor dos efeitos do erro (aleatórios). X, H e T são matrizes de incidência.

O teste da razão de verossimilhança (LRT), denominado de análise de deviance (ANADEV), para os efeitos aleatórios, foi realizado considerando o modelo adotado, conforme relatado por Resende et al. (2015). A ANADEV é uma estatística derivada da razão entre as verossimilhanças do modelo completo e do modelo reduzido na ausência do efeito que se deseja testar. Para testar a significância dessa razão utilizou-se o teste qui-quadrado com 1 grau de liberdade.

Para a seleção de dez melhores populações foram utilizadas duas estratégias: a primeira baseada na capacidade específica de combinação (CEC) e na capacidade geral de combinação (CGC) dos genitores e a segunda baseada na seleção via índice de seleção.

Os valores genotípicos (VG) das populações F1 foram expressos em: $u + CGC_i + CGC_j + CEC_{ij}$, em que: u é a média geral da caractere; CGC_i é o valor de capacidade geral de combinação do genitor i ; CGC_j é o valor de capacidade geral de combinação do genitor j ; e, CEC_{ij} é o valor de capacidade específica de combinação da população ij . Assim, a construção do índice de seleção aditivo (IS) se deu conforme descrito por Resende (2007b):

$$IS_{ij} = \sum(W_k \times VG_{p_k}), \text{ em que:}$$

IS_{ij} = valor estimado do índice para a população ij

W_k = peso econômico estabelecido para o k -ésimo caractere;

VG_{p_k} = valor genotípico padronizado da população ij para o k -ésimo caractere.

Os pesos econômicos foram estabelecidos adaptando o proposto por Viana e Resende (2014) em função das correlações genéticas do caractere sobre o qual se deseja realizar a seleção com os demais. Foram utilizados para a construção do índice apenas aqueles caracteres que apresentaram correlação significativa com o caractere principal. As estimativas dos pesos econômicos foram obtidas por:

$$W_k = \frac{rxk}{|1 + \sum_{k=1}^n rxk|}$$

rxk = correlação genética entre o caractere principal x e o k -ésimo caractere, sendo $k \neq x$;

O peso econômico adotado para o caractere principal foi obtido por:

$$W_x = \frac{1}{|1 + \sum_{k=1}^n rxk|}$$

Após escolhidas as melhores populações para alto teor de óleo e para alto teor de proteína nos grãos, foi estimado, para cada caractere, o índice de coincidência entre as estratégias de seleção. Tais estimativas de coincidência permitiram inferir sobre a eficiência do uso de índices de seleção e identificar populações com dupla aptidão.

As estimativas de correlações genéticas obtidas a partir dos valores genotípicos preditos das populações e o coeficiente de coincidência da seleção das populações ao utilizar as duas estratégias visando detectar populações com duplo propósito e avaliar a eficiência da seleção baseada em índices de seleção foram obtidos por meio do uso do software GENES (Cruz, 2013).

3.5 RESULTADOS E DISCUSSÃO

As estimativas dos parâmetros genéticos e os resultados da análise de deviance (ANADEV) para os caracteres DPF, AF, DPM, DHP, AM e NV estão apresentadas na Tabela 2. Verificou-se variabilidade entre as capacidades geral de combinação dos genitores pertencentes ao grupo 1 (σ^2_{cg1}) e grupo 2 (σ^2_{cg2}) do dialelo parcial e variabilidade entre genitores (σ^2_{genit}) desconsiderando a estrutura de grupos para essas características.

De acordo com a Tabela 2, apenas o caractere AM apresentou variabilidade entre as estimativas da capacidade específica de combinação das populações F1. Embora a herdabilidade dos valores devidos aos desvios da dominância (h^2_d) tenha sido estimada em 6,8%, enquanto que a herdabilidade no sentido restrito (h^2_a) foi estimada em 53,22%, a variabilidade notada para a CEC, indica presença de efeito heterótico para a altura das plantas na maturação. Com base nestas informações, a seleção de populações ou famílias nas primeiras gerações filiais para altura de plantas em soja não é uma boa alternativa, já que, neste caso, os desvios da dominância compõe um fator perturbador à seleção. A não significância dos efeitos da CEC para demais características (Tabela 2) foi resultado da pequena contribuição dos efeitos de dominância para as mesmas, como observado pelos valores de h^2_d .

As magnitudes das h^2_a , oriunda do somatório das herdabilidades referentes às capacidades geral de combinação de ambos os grupos (h^2_{cgc1} e h^2_{cgc2}), em detrimento à h^2_d evidencia a maior contribuição dos efeitos aditivos para a expressão dos caracteres (Tabela 2).

TABELA 2 – Estimativa de parâmetros genéticos para os caracteres: dias para florescimento (DPF), altura no florescimento (AF), dias para maturação (DPM), diâmetro do hipocótilo (DHP), altura na maturação (AM) e número de vagens (NV) para um dialelo parcial em soja

Parâmetros	DPF	AF	DPM	DHP	AM	NV
σ^2_{cgc1}	12,171 **	50,626 **	38,149 **	0,068 **	90,443 **	12,749 **
σ^2_{cgc2}	0,227 '	24,880 **	9,483 **	0,082 **	87,831 **	11,979 **
σ^2_{cec}	0,055 ns	0,662 ns	0,187 ns	0,006 ns	22,892 *	1,883 ns
σ^2_{genit}	6,995 **	25,452 *	18,812 **	0,132 *	98,549 '	14,255 *
σ^2_b	0,605	17,498	7,380	0,134	26,591	4,740
σ^2_e	3,424	44,752	19,121	0,288	107,212	19,800
σ^2_{fpop}	16,481	138,418	74,319	0,578	334,970	51,151
σ^2_{fgenit}	11,024	87,702	45,312	0,554	323,393	38,795
$r_{\hat{a}a}$	0,867	0,739	0,801	0,509	0,730	0,695
h^2_{cgc1}	0,738	0,366	0,513	0,117	0,270	0,249
h^2_{cgc2}	0,014	0,180	0,128	0,142	0,262	0,234
h^2_a	0,752	0,546	0,642	0,259	0,532	0,483
h^2_d	0,003	0,005	0,000	0,010	0,068	0,037
h^2_g	0,756	0,550	0,642	0,269	0,601	0,520
Média	42,620	49,535	119,414	5,151	77,020	32,594

** P < 0,01, * P < 0,05, ' P < 0,1 e ns não-significativo a 1 g.l. pelo teste de χ^2 (1% = 6,63, 5% = 3,84 e 10% = 2,71%), via teste de LRT (Likelihood Ratio Test) dos efeitos aleatórios.

σ^2_{cgc1} = variância da CGC dos genitores do grupo 1; σ^2_{cgc2} = variância da CGC dos genitores do grupo 2; σ^2_{cec} = variância da CEC das populações. σ^2_{genit} = variância genotípica entre genitores; σ^2_b = variância entre blocos; σ^2_e = variância do erro; σ^2_{fpop} = variância fenotípica entre populações. σ^2_{fgenit} = variância fenotípica entre genitores; $r_{\hat{a}a}$ = acurácia; h^2_{cgc1} = herdabilidade dos genitores do grupo 1; h^2_{cgc2} = herdabilidade dos genitores do grupo 2; h^2_a = herdabilidade no sentido restrito; h^2_d = herdabilidade dos valores devido aos desvios da dominância; h^2_g = herdabilidade no sentido amplo.

As estimativas dos parâmetros genéticos e os resultados da análise de deviance (ANADEV) para as variáveis NN, NH, NG, PROD, OL e PT estão apresentadas na Tabela 3. Não foi detectada variabilidade entre as estimativas das CGC dos genitores do grupo 1 para o caractere PROD e entre as estimadas para os genitores do grupo 2 para os caracteres PROD e OL. Tendo em vista que os genitores pertencentes aos grupos 1 e 2 são cultivares comerciais com alta produtividade de grãos para as devidas zonas agrícolas e alto teor de óleo nos grãos,

respectivamente, justifica-se a ausência de variabilidade para as estimativas de CGC no grupo 1 para o caractere PROD e as no grupo 2 para OL.

Contudo, as cultivares utilizadas como genitores apresentaram variabilidade para todos os caracteres apresentados na Tabela 3, conforme análise da variância entre genitores (σ^2_{genit}). Quanto à CEC, foi notado variabilidade apenas para o caractere PT. Como apresentado anteriormente, a não significância dos efeitos da CEC para demais características (Tabela 3) também foi resultado da pequena contribuição dos efeitos de dominância para as mesmas, como observado pelos valores de h^2_d . Da mesma forma ficou evidente a maior importância dos efeitos aditivos em detrimento aos de dominância para a expressão dos caracteres.

TABELA 3 – Estimativa de parâmetros genéticos para os caracteres: número de nós (NN), número de hastes (NH), número de grãos por planta (NG), massa de grãos por planta (PROD), teor de óleo (OL) e proteína nos grãos (PT) para um dialelo parcial em soja.

Parâmetros	NN	NH	NG	PROD	OL	PT
σ^2_{cgc1}	0,310 *	0,158 **	28,192 **	0,170 ns	0,100 '	0,594 **
σ^2_{cgc2}	0,278 **	0,160 **	54,576 **	0,329 ns	0,016 ns	0,129 '
σ^2_{cec}	0,105 ns	0,009 ns	7,899 ns	0,337 ns	0,052 ns	0,160 *
σ^2_{genit}	1,489 **	0,312 *	84,063 **	1,636 *	0,629 *	0,582 '
σ^2_b	0,091	0,216	25,539	1,093	0,082	0,177
σ^2_e	0,778	0,747	110,007	3,019	1,023	0,994
σ^2_{fpop}	1,563	1,291	226,213	4,947	1,273	2,053
σ^2_{fgenit}	2,358	1,276	219,609	5,747	1,734	1,752
r_{aa}	0,614	0,496	0,608	0,317	0,301	0,593
h^2_{cgc1}	0,199	0,122	0,125	0,034	0,079	0,289
h^2_{cgc2}	0,178	0,124	0,241	0,067	0,012	0,063
h^2_a	0,377	0,246	0,366	0,101	0,091	0,352
h^2_d	0,067	0,007	0,035	0,068	0,041	0,078
h^2_g	0,444	0,253	0,401	0,169	0,132	0,429
Média	12,868	4,670	70,529	12,210	19,269	39,304

** P < 0,01, * P < 0,05, ' P < 0,1 e ns não-significativo a 1 g.l. pelo teste de χ^2 (1% = 6,63, 5% = 3,84 e 10% = 2,71%), via teste de LRT (Likelihood Ratio Test) dos efeitos aleatórios.

σ^2_{cgc1} = variância da CGC dos genitores do grupo 1; σ^2_{cgc2} = variância da CGC dos genitores do grupo 2; σ^2_{cec} = variância da CEC das populações; σ^2_{genit} = variância genotípica entre genitores; σ^2_b = variância entre blocos; σ^2_e = variância do erro; σ^2_{fpop} = variância fenotípica entre populações; σ^2_{fgenit} = variância fenotípica entre genitores; r_{aa} = acurácia; h^2_{cgc1} = herdabilidade dos genitores do grupo 1; h^2_{cgc2} = herdabilidade dos genitores do grupo 2; h^2_a = herdabilidade no sentido restrito; h^2_d = herdabilidade dos valores devido aos desvios da dominância; h^2_g = herdabilidade no sentido amplo.

Uma boa medida para se julgar a precisão experimental é a acurária (r_{aa}) que traduz a confiabilidade dos resultados obtidos. Quanto maior o valor da acurária, mais próximos estão os valores genéticos preditos dos valores genéticos verdadeiros, conforme relatado por Resende (2002). Esse mesmo autor propõe uma classificação para os valores de acurária, sendo estes de magnitudes alta ($r_{aa} \geq 0,70$), moderada ($0,40 \leq r_{aa} \leq 0,70$) e baixa ($0,10 \leq r_{aa} \leq 0,40$). Os menores valores de acurária apresentados nas Tabelas 2 e 3, considerados baixos segundo Resende (2002), foram para os caracteres PROD e OL. Sendo assim, a seleção direta sobre a produtividade e o teor de óleo nos grãos não é recomendada. Neste caso, justifica-se o uso de índice de seleção com base em caracteres que auxiliam a seleção sobre estes caracteres. As demais magnitudes foram consideradas entre moderadas e altas.

De acordo com Cruz et al. (2012), em estudos genéticos, é indispensável a estimação das correlações entre os caracteres estudados, pois estas medem o grau de associação entre os mesmo e elucida a possibilidade ou não da seleção indireta. Segundo Falconer (1987), a pleiotropia é causa permanente de correlações genéticas, já as ligações gênicas, constituem uma causa transitória, principalmente em populações derivadas de cruzamento entre linhagens divergentes. No presente estudo, verificou-se correlação significativa e negativa entre os caracteres OL e PT (-0,69) (Tabela 4). Estes resultados estão em conformidade com a literatura (Liang et al., 2010; Hwang et al., 2014). Hanson (1991) e Pípolo (2002) citados por Albrecht et al., (2008), explicam que esta correlação negativa é devido a concorrência destas duas vertentes pelos esqueletos de carbono. Além de verificada também entre os genitores, a variabilidade entre os efeitos de CEC para o caractere altura de plantas na maturação deve ser explorada para seleção de populações com alto teor de óleo e proteína nos grãos, já que foram encontradas correlações positivas significativas entre os caracteres AM x OL (0,61) e AM x PT (-0,68).

Utilizando as estimativas das correlações significativas com os caracteres teor de óleo e de proteína nos grãos foram determinados os pesos econômicos para a confecção do índice de seleção (Tabela 4). A confecção do índice de seleção com base apenas em caracteres que apresentam correlação significativa com os caracteres principais tem como vantagens, garantir que os caracteres auxiliares sejam àqueles de correlações diferente de zero e evitar que o peso dado a estes caracteres seja diluído em outros caracteres de pouca importância para a seleção. A utilização dessa estratégia de seleção favorece o maior acúmulo de alelos controladores da característica alvo, favorecendo a maior variabilidade genética a ser explorada em futuras gerações.

Tabela 4 – Correlações genéticas entre os caracteres: dias para florescimento (DPF), altura no florescimento (AF), dias para maturação (DPF), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de vagens (NV), número de nós (NN), número de hastes (NH), número de grãos por planta (NG), massa de grãos por planta (PROD), teor de óleo (OL) e proteína nos grãos (PT) e pesos econômicos (W) dos caracteres incluídos nos índices de seleção para seleção de populações F₁ alto teor de óleo nos grãos (IS_{ol}) e alto teor de proteína nos grãos (IS_{pt}).

Correlações genéticas													
	DPF	AF	DPM	DHP	AM	NV	NH	NN	NG	PROD	OL	PT	
DPF	1,00												
AM	0,71 **	1,00											
DPM	0,88 **	0,55 **	1,00										
DHP	0,43 *	0,33	0,57 **	1,00									
AM	-0,27	0,28	-0,19	-0,04	1,00								
NV	0,69 **	0,60 **	0,84 **	0,52 **	-0,01	1,00							
NH	-0,09	0,03	0,14	0,03	0,58 **	0,19	1,00						
NN	0,51 **	0,47 **	0,66 **	0,45 **	-0,27	0,79 **	-0,09	1,00					
NG	0,57 **	0,60 **	0,73 **	0,48 **	0,17	0,94 **	0,36 *	0,74 **	1,00				
PROD	0,31	0,58 **	0,43 *	0,33	0,40 *	0,59 **	0,52 **	0,45 **	0,74 **	1,00			
OL	0,55 **	0,61 **	0,40 *	0,17	0,05	0,39 *	-0,08	0,18	0,33	0,43 *	1,00		
PT	-0,44 **	-0,68 **	-0,34 *	0,04	-0,37 *	-0,54 **	-0,22	-0,20	-0,57 **	-0,56 **	-0,69 **	1,00	
W													
	DPF	AF	DPM	DHP	AM	NV	NH	NN	NG	PROD	OL	PT	
IS _{ol}	0,20	0,23	0,15	-	-	0,15	-	-	-	0,16	0,37	-0,26	
IS _{pt}	-0,14	-0,21	-0,11	-	-0,12	-0,17	-	-	-0,18	-0,18	-0,22	0,31	

*; ** Significativo a 1% e 5% de probabilidade, respectivamente de acordo com o teste de Mantel.

A escolha entre cruzamentos para comporem um programa de melhoramento genético é uma fase importante e dispendiosa do ponto de vista financeiro e de utilização de recurso humano (Pimentel et al., 2013). Pesquisadores tem estudado diferentes maneiras de aperfeiçoar este processo no melhoramento genético da soja, sendo mais comumente por meio de estimações de variabilidade genética por métodos multivariados (Almeida et al., 2011; Rigon et al., 2012) e seleção de populações de acordo com as capacidades de combinação estimadas à partir da análise dialélica (Menezes et al., 2010).

No presente estudo, a análise dialélica não apontou variabilidade entre os efeitos da CEC para os caracteres PROD e OL. Contudo, notou-se variabilidade entre os genitores para todos os 12 caracteres estudados. Segundo Cruz et al., (2014), a seleção com base em uma ou poucas características não compõe uma boa estratégia de seleção, pois conduz a obtenção de um produto final superior para os caracteres selecionados, porém baixo desempenho para os demais caracteres. Sendo assim, tomou-se com estratégia para a escolha entre as populações com bom desempenho para PROD e OL, o uso de índices de seleção os quais permitam explorar a variabilidade existente entre genitores em caracteres auxiliares desde que estes sejam correlacionados com o caractere que se deseja exercer a seleção.

A confecção do índice de seleção clássico com pesos econômicos W , conforme proposto por Viana e Resende (2014), apresenta como vantagem utilizar como caracteres auxiliares aqueles que possuem correlação com o caractere principal, sendo que quanto maior a correlação, maior é o peso de tal caractere no índice. No entanto, a inclusão de caracteres cuja correlação com o caractere principal seja não significativa, ou estatisticamente não difere de zero, pode acabar por conduzir a classificação errônea dos indivíduos, sobretudo quando há alta variabilidade para o caractere em questão.

As estimativas dos índices de seleção para seleção de populações com alto teor de óleo (IS_{ol}) e alto teor de proteína nos grãos (IS_{pt}) são apresentados na Tabela 5.

Como as estimativas de CEC não apresentaram variabilidade para o caractere OL, a estratégia foi a seleção das populações via o uso do índice de seleção favorecendo a maior concentração de alelos favoráveis para esse caractere.

As populações selecionadas com base no índice, Tabela 5, podem ser recomendadas para um programa de seleção recorrente intrapopulacional, já que tais populações possuem alelos favoráveis para a expressão do caractere teor de óleo nos grãos e esse programa contempla a unificação desses alelos em futuras populações originadas dessas.

Tabela 5 – Valores genéticos (VG) das populações F₁ para os caracteres produção de grãos por planta (PROD), teores de óleo (OL) e proteína (PT) e índices de seleção para alto rendimento de grãos (IS_{prod}), alto teor de óleo nos grãos (IS_{ol}), alto teor de proteína nos grãos (IS_{pt}).

Populações	VG		IS _{ol} ²	IS _{pt} ³
	OL	PT		
Pop 1	18,87	40,83 ¹	-1,37	1,16
Pop 2	19,14	38,73	0,61 ³	-1,31
Pop 3	19,01	39,17	0,12	-0,74
Pop 4	19,08	39,90	-0,17	-0,11
Pop 5	19,33	39,00	1,60	-1,79
Pop 6	19,67	37,61	2,80	-3,09
Pop 7	18,77	41,16	-2,54	2,79
Pop 8	19,12	39,60	-0,92	0,93
Pop 9	19,10	39,78	-0,85	0,84
Pop 10	19,19	40,58	-1,12	1,42
Pop 11	19,47	39,87	0,63	-0,17
Pop 12	19,62	38,81	1,41	-1,15
Pop 13	18,58	41,00	-2,49	2,49
Pop 14	19,45	39,21	-0,38	0,50
Pop 15	18,96	39,91	-0,86	0,62
Pop 16	18,77	39,45	-1,28	1,13
Pop 17	19,28	38,64	0,17	-0,27
Pop 18	19,00	39,39	-0,30	-0,06
Pop 19	19,26	40,04	-0,33	0,37
Pop 20	19,42	39,28	0,27	-0,14
Pop 21	19,34	39,97	0,50	-0,15
Pop 22	19,44	38,95	1,14	-0,79
Pop 23	19,45	38,44	1,53	-1,48
Pop 24	19,00	40,53	-1,78	1,90
Pop 25	19,66	38,79	0,08	0,06
Pop 26	19,45	38,55	0,21	-0,41
Pop 27	18,89	39,45	-1,17	1,03
Pop 28	19,57	38,32	0,36	-0,27
Pop 29	19,42	38,68	0,18	-0,39
Pop 30	19,38	40,10	-0,25	0,40
Pop 31	19,46	39,41	0,10	0,16
Pop 32	19,54	39,62	1,14	-0,84
Pop 33	19,52	38,40	1,34	-1,03
Pop 34	19,51	38,36	1,63	-1,59

¹ Valores destacados em negrito indicam populações selecionadas com base na CEC com pelo menos um genitor com alta CGC; ² Valores em negrito indicam as populações selecionadas pelo índice de seleção calculado com base em pesos econômicos (W) dos caracteres DPF, AF, DPM, NV, PROD, OL e PT; ³ Valores em negrito indicam as populações selecionadas pelo índice de seleção calculado com base em pesos econômicos (W) dos caracteres DPF, AF, DPM, DHP, AM, NV, NH, NN, NG, PROD, OL e PT.

Contudo, para o caractere OL, a análise dialélica revelou variabilidade entre as estimativas das CGC dos genitores pertencentes ao grupo 1, o que mostra a existência de número diferente de alelos favoráveis entre os genitores. Conforme descrito por Castiglioni et al (1999), a CGC é função da concentração de genes predominantemente aditivos os quais são transmitidos às próximas gerações e refletem o desempenho das linhas puras de cada populações ao atingir a homozigose. Neste sentido, recomenda-se a utilização das cultivares TMG 132 RR, DM 5958 IPRO, M 8349 IPRO, P98Y30, M7739 IPRO e DM 6563 IPRO como genitores em programa de melhoramento para aumento do teor de óleo nos grãos (Tabela 6). Contudo, ao considerar as populações selecionadas via índice de seleção ficou evidenciado a contribuição dos genitores TMG 132 RR, P98Y30 e M 8349 IPRO para o aumento do caractere em questão.

Tabela 6 - Estimativas das capacidades geral (CGC) e específica (CEC) para teor de óleo nos grãos (OL) dos cruzamentos obtidos via sistema de dialelo parcial em soja.

Grupo 1/ Grupo 2	OL				
	M 9056 RR	M 8221 RR	P98N71	CONQUISTA	CGC
BMX Apolo RR	0,0582	-0,0625	-0,1909	0,0609	-0,4317
DM 5958 IPRO			0,0273	0,0751	0,2207
NA 5909 RG	-0,0141	-0,0524	-0,1444	0,1734	-0,0934
TMG 7062 IPRO			-0,0716	-0,1217	-0,3603
DM 6563 IPRO			-0,0246	0,0954	0,1091
BMX Potência RR	-0,1062	-0,0282	-0,0660	0,1845	-0,1339
BMX Desafio RR	-0,1146	-0,0252	0,1169	0,0656	-0,0527
M7739 IPRO			0,0761	-0,0498	0,1411
P98Y30 RR	-0,0772	0,0505	-0,0130	0,0100	0,1580
M8349 IPRO			0,0526	-0,0317	0,1866
TMG 132 RR	0,1618	0,0973	-0,0016	-0,1096	0,2564
CGC	-0,0219	-0,0059	-0,0708	0,0987	

¹ Estimativas de CEC dos cruzamentos são apresentadas dentro da tabela e as estimativas de CGC nas marginais.

Quanto às estratégias de seleção de populações para alto teor de proteína nos grãos, a presença de variabilidade para PT verificada para a CEC das populações permitiu adotar como estratégia, a seleção de populações com base em altas magnitudes de CEC das populações oriundas de pelo menos um genitor com alta magnitude de CGC, além da seleção via índice de seleção (Tabela 7), sendo que o coeficiente de coincidência estimado entre as duas estratégias de seleção foi de 60%.

Em seleção baseada na CEC espera-se encontrar maior variabilidade dentro das populações selecionadas uma vez que esta é função da divergência genética (Hallauer e Miranda Filho, 1981; Rao et al., 1981; Mendes et al., 2009) e complementariedade dos genitores envolvidos no cruzamento (Cruz et al., 2012). Além disto, a magnitude dos efeitos da CEC é proporcional à existência de variabilidade em cada população e, portanto, espera-se que ocorram segregantes transgressivos nas próximas gerações (Barelli et al., 2000). Contudo, a seleção pelo índice utilizando o peso econômico para cada caractere conforme descrito no trabalho também favorece a seleção de populações com alta variabilidade genética, conforme comentado anteriormente. Por outro lado, o índice de seleção apontou a escolha das populações de maior valor genotípico predito.

Tabela 7 - Estimativas das capacidades geral (CGC) e específica (CEC) para teor de proteína nos grãos (PT) dos cruzamentos obtidos via sistema de dialelo parcial em soja.

Grupo 1/ Grupo 2	PT				
	M 9056 RR	M 8221 RR	P98N71	CONQUISTA	CGC
BMX Apolo RR	0,3273	-0,0299	0,0050	-0,1439	1,5370
DM 5958 IPRO	-	-	0,0316	-0,0593	-0,2848
NA 5909 RG	-0,1291	0,0625	0,5560	-0,4762	-0,1101
TMG 7062 IPRO	-	-	-0,1203	0,2019	0,1058
DM 6563 IPRO	-	-	-0,0756	-0,0665	-0,7485
BMX Potência RR	0,2144	0,1463	-0,0562	-0,4451	-0,0159
BMX Desafio RR	0,0994	0,0956	-0,2559	0,1297	0,8317
M7739 IPRO	-	-	-0,0947	0,3570	-0,0837
P98Y30 RR	-0,2388	-0,0526	0,2348	0,2055	0,2739
M8349 IPRO	-	-	0,0853	-0,1385	-0,5991
TMG 132 RR	-0,4459	0,0678	-0,1178	0,1277	-0,9059
CGC	-0,3373	0,3453	0,1572	-0,1651	

¹ Estimativas de CEC dos cruzamentos são apresentadas dentro da tabela e as estimativas de CGC nas marginais.

As populações Pop 1, Pop 8, Pop 9, Pop 10, Pop 15 e Pop 27 foram as selecionadas em ambas as estratégias de seleção. Os genitores que possuíram maiores valores de CGC e, portanto são promissores para os programas de melhoramento para aumento do teor de proteína nos grãos foram: BMX Apolo RR, BMX Desafio RR, P98Y30 e TMG 7062 IPRO, pertencentes ao grupo 1; e, M 8221 RR e P98N71, pertencentes ao grupo 2.

Não houve coincidência entre as populações selecionadas pelos índices IS_{ol} e IS_{pt} , e isso se deve principalmente pela magnitude e sentido da correlação entre os caracteres OL e PT. Contudo, vale ressaltar que neste trabalho o genitor P98Y30 apresentou-se como

potencial para programas de melhoramento que visam tanto para aumentar o teor de óleo nos grãos quanto para o de proteína. Trabalhos futuros poderão comprovar a presença de alelos favoráveis para ambos os caracteres nessa cultivar.

3.6 CONCLUSÕES

O uso de índices de seleção é uma boa alternativa para escolha de populações F_1 com boa qualidade de grãos.

A seleção baseada nas capacidades de combinação indicou a seleção de populações de mais alta CEC, enquanto o índice de seleção preconizou a escolha de populações de maior valor genotípico predito.

3.7 REFERÊNCIAS BIBLIOGRÁFICAS

ABIOVE. Estatística. 2016. Disponível em: <<http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=>>. Acesso em: 23 ago. 2016.

ALLIPRANDINI, L. F. et al. Understanding soybean maturity groups in Brazil: environment, cultivar classification, and stability. **Crop Science**, v. 49, n. 3, p. 801-808, 2009.

ALBRECHT, L. P. et al. Teores de óleo, proteínas e produtividade de soja em função da antecipação da semeadura na região oeste do Paraná. **Bragantia**, v. 67, n. 4, p. 865-873, 2008.

ALMEIDA, R. D. de; PELUZIO, J. M.; AFFÉRI, F S. Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no sul do Estado Tocantins. **Revista Ciência Agrônômica**, v. 42, n. 1, p. 108-115, 2011.

ÁVILA, M. R. et al. Componentes do rendimento, teores de isoflavonas, proteínas, óleo e qualidade de sementes de soja. **Revista Brasileira de Sementes**, v. 29, n. 3, p. 111-127, 2007.

BARELLI, M. A. A. et al. Diallel analysis for grain yield and yield components in *Phaseolus vulgaris* L. **Acta Scientiarum Agronomy**, v.22, p.883-887, 2000.

CÂMARA, G. M. S. Ecofisiologia da soja e rendimento. In: CÂMARA, G.M.S. Soja: tecnologias da produção. Piracicaba: ESALQ, 1998. p.256-77. CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4 ed. Viçosa: UFV, 2012. v. 1. 514 p.

CASTIGLIONI, V. B. R.; OLIVEIRA, M. F. de; ARIAS, C. A. A. Análise da capacidade combinatória entre linhagens de girassol. **Pesquisa Agropecuária Brasileira**, v. 34, n. 6, p. 981-988, 1999.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4 ed. Viçosa: UFV, 2012. v. 1. 514 p.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 3 ed. Viçosa: UFV, 2014. v. 2. 668 p.

CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**. v.35, n.3, p.271-276, 2013.

DARONCH, D. J. et al. Capacidade combinatória de cultivares de soja em f2, sob condições de cerrado tocantinense = Combining ability of soybean cultivars in f2, under conditions of cerrado tocantinense. **Bioscience Journal**, v. 30, n. 5, p. 688-695, 2014.

FALCONER, D. S.; MACKAY, T.F. **Introduction to quantitative genetics**. 4 ed. Londres: Longman Group, 464p. 1996.

FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. 1977.

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

HANSON, W. D. Modified seed maturation and seed yield potentials in soybean. **Crop Science**, Madison, v.32, n.4, p.972-976, 1992.

HENDERSON, C. R. Estimation of changes in herd environment. **Journal of Dairy Science**, v. 32, p. 709, 1949.

Hwang, E. Y. et al. A genome-wide association study of seed protein and oil content in soybean. **BMC genomics**, v. 15, n. 1, p. 1, 2014.

KARYAWATI, A. S., SITOMPUL, S. M., BASUKI, N. Combining Ability Analysis for Physiological Characters of Soybean (*Glycine max* L. Merrill). **International Journal of Plant Research**, v. 5, n. 5, p. 113-121, 2015.

LIANG, H. Z. et al QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). **Agricultural Sciences in China**, v. 9, n. 8, p. 1108-1116, 2010.

MAUAD, M. et al. Influência da densidade de semeadura sobre características agronômicas na cultura da soja. **Revista Agrarian**, Dourados-MS, v. 3, n. 9, p. 175-181, 2010.

MENDES, F. F.; RAMALHO, M. A. P.; ABREU, A. de F. B. Índice de seleção para escolha de populações segregantes de feijoeiro-comum. **Pesquisa Agropecuária Brasileira**, v. 44, n. 10, p. 1312-1318, 2009.

- MENEZES, M. et al. Aspectos químicos e estruturais da qualidade fisiológica de sementes de soja. **Pesquisa Agropecuária Brasileira**, v. 44, n. 12, p. 1716-1723, 2010.
- MORAES, R. M. A. de et al. Biochemical characterization of high protein soybean lines. **Pesquisa Agropecuária Brasileira**, v. 41, n. 5, p. 715-729, 2006.
- NDOUMBÉ, M.; BIEYSSE, D.; CILAS, C. Multi-trait selection in a diallel crossing scheme of cocoa. **Plant Breeding**, v. 120, n. 4, p. 365-367, 2001.
- PIMENTEL, A. J. B. et al. Comparação de métodos de seleção de genitores e populações segregantes aplicados ao melhoramento de trigo. **Bragantia**, v.72, n.2, p.113-121, 2013.
- PÍPOLO, A. E. **Influência da temperatura sobre as concentrações de proteína e óleo em sementes de soja (*Glycine max* (L.) Merrill)**. 128p. 2002. (Tese) Doutorado - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.
- RAMALHO, M. A. P.; ARAÚJO, L. C. de A. Breeding self-pollinated plants. **Crop Breeding and Applied Biotechnology**, v.11, p.1-7, 2011. DOI: 10.1590/S1984-70332011000500002.
- RAO, A. V. et al. Genetic divergence among some brown planthopper resistant rice varieties. **The Indian Journal of Genetic Plant Breeding**, v.41, n.2, p.179-185. 1981.
- RESENDE, M. D. V. et al. Selection index with parents, populations, progenies and generations effects in autogamous plant breeding. **Crop Science**, 2015.
- RESENDE, M.D.V. de. **Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes**. Colombo: Embrapa Florestas, 2000. 101p. (Embrapa Florestas. Documentos, 47).
- RESENDE, M.D.V. de. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004. 57 p. (Embrapa Florestas. Documentos, 100).
- RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Embrapa Florestas, Colombo. 561 p. 2007a.
- RESENDE, M. D. V. de. Software **SELEGEN–REML/BLUP**: sistema estatístico e seleção computadorizada via modelos lineares mistos. Colombo: Embrapa Florestas, 2007b.

RIGON, J. P. G. et al. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. **Revista Ceres**, v.59, n.2, p.233-240, 2012.

SEDIYAMA, T.; SILVA, F.; BORÉM, A. **Soja: do plantio a colheita**. Viçosa. Editora UFV, 2015. 333 p.

SILVA, M. G. de M.; VIANA, A. P. Alternativas de seleção em população de maracujazeiro-azedo sob seleção recorrente intrapopulacional. **Revista Brasileira de Fruticultura, Jaboticabal**, v. 34, n. 2, p. 525-531, 2012.

VIANA, A. P.; RESENDE, M. D. V.; **Genética quantitativa aplicada no melhoramento de fruteiras**. Interciência 2014, p. 282.

4 CAPÍTULO II – CORRELAÇÕES GENÉTICAS E ANÁLISE DE TRILHA PARA TEOR DE ÓLEO EM GRÃOS DE SOJA VIA VALORES GENOTÍPICOS PREDITOS (BLUP).

4.1 RESUMO

As correlações entre caracteres são apenas uma medida de associação e não contemplam as relações de causa e efeito. A análise de trilha proposta por Wright (1921) permite entender desdobrar as correlações em efeitos diretos e indiretos. A existência de multicolinearidade na matriz de correlação entre os caracteres explicativos traz problemas à condução das estimativas. Sendo assim, o presente estudo objetivou: a) estudar as correlações entre componentes vegetativos, de produção e de qualidade de grãos, bem como seus efeitos diretos e indiretos sobre o caractere teor de óleo em grãos de soja; b) verificar os efeitos da exclusão de variáveis redundantes e do emprego da análise de trilha com multicolinearidade por meio da regressão em crista sobre tais estimativas. Notou-se uma tendência de plantas de ciclo mais prolongado apresentarem maior produção de grãos por planta pelas altas correlações dos caracteres dias para maturação x número de nós, número de nós na haste principal x número de vagens e número de vagens x massa de grãos por planta. Verificou-se pela análise de trilha a relação de causa e efeito entre massa de grãos por planta e teor de óleo nos grãos. O caractere teor de proteína nos grãos apresentou alta correlação negativa e efeito direto baixo de mesmo sinal sobre teor de óleo nos grãos, logo devem ser empregados índices de seleção multivariados para obter ganhos genéticos sobre teor de óleo. A exclusão de caracteres da análise de trilha é a alternativa mais simples para contornar os problemas com a multicolinearidade e neste estudo conduziu as estimativas mais confiáveis.

Palavras-chave: *Glycine max* (L.) Merr.; teor de proteína; multicolinearidade; regressão em crista; exclusão de caracteres.

Genetic correlations and path analysis to soybean grain oil content by predicted genetic values (BLUP)

4.2 ABSTRACT

Correlations between traits are only an association measure and does not contemplate direct and indirect effects. The path analysis proposed by Wright (1921) allows to understand the relations of direct and indirect effects between traits. The existence multicollinearity on the correlation matrix between the explanatory traits may cause problems to the estimates process of direct and indirect effects by path analysis. Therefore, this study aimed: a) studying the correlation between growth, yield and grain quality components, as well as direct and indirect effects on the target trait grain oil content; b) verify the trait culling and ridge path analysis under multicollinearity effects on estimation of direct and indirect effects on the target trait. It was noted a tendency of long growth cycles that presented larger grain yield per plant by high correlation between grain ripening x number of main stem nodes, number of main stem nodes x number of pods per plant and number of pods per plant x mass of grains per plant. A very strong direct effects between mass of grains per plant and oil content was verified by the path analysis. The trait grain protein content presented high negative correlation and the low and negative direct effect on grain oil content, therefore must be used multivariate selection index to achieve genetics gains on grain oil content. Culling traits path analysis is the simpler alternative to work under multicollinearity problems and it was able to provide trustworthy estimates.

Key-word: *Glycine max* (L.) Merr.; protein content; multicollinearity; ridge regression; trait culling.

4.3 INTRODUÇÃO

A cultura da soja tem exercido papel importante no agronegócio brasileiro. Esta oleaginosa apresenta em sua composição cerca de 20% de óleo no conteúdo da semente (Lazzarotto e Hirakuri, 2010; Bezerra, 2015) e este produto tem sido empregado de várias formas, especialmente para produção de óleo de cozinha e biodiesel. Dados da Conab (2016) revelam a estagnação da produtividade de óleo de soja no Brasil.

De acordo com Cruz (2005), caracteres quantitativos são aqueles controlados por uma grande quantidade de genes, sofrem forte influência ambiental e apresentam menores estimativas de herdabilidade quando comparados àqueles que são controlados por um ou pouco genes (qualitativos). Na cultura da soja, o teor de óleo, o teor proteína e a produtividade de grãos em soja são caracteres classificados como quantitativos (Lee et al., 1996; Rodrigues et al., 2010; Marquez, 2010). A baixa herdabilidade geralmente estimada para caracteres de natureza quantitativa dificulta a obtenção de ganhos pela seleção direta. Carvalho et al. (2004) citam que a seleção indireta em um caractere de alta herdabilidade, de fácil mensuração e altamente correlacionado com outro de baixa herdabilidade, tem grandes chances de resultar em ganhos genéticos mais rápido do que a seleção direta sobre os caracteres de baixa herdabilidade.

O estudo das correlações entre caracteres é indispensável em programas de melhoramento, pois este mede o grau de associação entre os mesmo e elucida a viabilidade da seleção indireta (Cruz et al., 2012). Dentre as possíveis causas das correlações entre caracteres citam-se o desequilíbrio de ligação e a pleiotropia. Enquanto o desequilíbrio de ligação é causa transitória das correlações genéticas, principalmente em populações derivadas de cruzamentos entre linhagens divergentes, a pleiotropia constitui-se em uma causa permanente (Falconer, 1987).

Vencovsky e BARRIGA (1992) ressaltam que as correlações entre caracteres são apenas uma medida de associação e não contemplam as relações de causa e efeito. Para entender as relações de causa e efeito, Wright (1921) propôs o emprego da análise de trilha (*path coefficient analysis*) que, em suma, é uma expansão de uma regressão linear múltipla e permite desdobrar os coeficientes de correlação em efeitos diretos e indiretos sobre uma variável básica (Alcantara Neto et al., 2011; Cruz et al., 2012).

No entanto, Cruz et al. (2014) aconselham que seja realizado o diagnóstico de multicolinearidade na matriz de correlação das variáveis explicativas. Ainda que a análise de

trilha possa ser realizada com multicolinearidade severa, as estimativas dos efeitos diretos e indiretos tendem a ser viesadas e apresentam magnitude superior a 1, logo, sem interpretação biológica (Toebe e Cargnelutti Filho, 2013). Nestes casos, Cruz et al. (2014) recomendam o uso de regressão em crista ou em cumeeira, que tem como fundamento permitir obter soluções para o vetor de parâmetros, mesmo sob multicolinearidade, pela adição de pequenos valores constantes (k) na matriz de correlação, conforme proposto por Hoerl e Kennard (1970a, b) ou a exclusão de variáveis altamente associadas.

Para este trabalho foram propostos dois objetivos: a) entender as correlações entre componentes vegetativos, reprodutivos e de qualidade de semente, bem como estimar seus efeitos diretos e indiretos sobre o caractere teor de óleo em grãos de soja; b) verificar os efeitos da exclusão de variáveis redundantes e do emprego da análise de trilha com multicolinearidade por meio da regressão em crista.

4.4 MATERIAL E MÉTODOS

O experimento foi composto por 49 tratamentos sendo eles, 34 populações F_1 de soja oriundas de um dialelo parcial e mais 15 cultivares comerciais utilizadas como genitores: BMX Apolo RR, DM 5958 IPRO, NA 5909 RR, TMG 7062 IPRO, DM 6563 IPRO, BMX Potência RR, BMX Desafio RR, M7739 IPRO, P98Y30 RR, M8349 IPRO, TMG 132 RR, M 9056 RR, M 8221 RR, P98N71 e Conquista. As cultivares utilizadas como genitores para as hibridações pertencem a grupos de maturidade relativa compreendidos entre 5.5 a 9.0 (Alliprandini, 2009).

A implantação e condução do experimento se deu sob delineamento de blocos casualizado com 5 repetições em casa de vegetação do Programa de Melhoramento de Soja na Universidade Federal de Viçosa. A semeadura ocorreu no dia 22 de fevereiro de 2016. Para tanto, a casa de vegetação foi adaptada com lâmpadas para ajuste do fotoperíodo, favorecendo um ambiente adequado para o bom desenvolvimento da soja levando em conta a importância do fotoperíodo como indutor de florescimento e fator fundamental para o desenvolvimento da cultura (Câmara, 1998).

Os tratamentos culturais foram realizados conforme recomendados por Sediya et al. (2015). Foram avaliados 12 caracteres, sendo estes, componentes vegetativos, produtivos e de qualidade de grãos. À medida que as plantas alcançavam o estágio fenológico R1 definido por Fehr e Caviness (1977), foram tomadas medidas das plantas individuais para os seguintes

caracteres: dias para florescimento (DPF): contados a partir do estágio VE definido por Fehr e Caviness (1977); e, altura no florescimento (AF) em cm: medida da distância da base até o ápice da planta. Já na ocasião da maturação, conforme as plantas atingiam o estágio fenológico R8 definido por Fehr e Caviness (1977), foram tomadas medidas das plantas individuais para: dias para maturação (DPM): contados a partir de VE; diâmetro do hipocótilo (DHP) em mm: medido dois centímetros abaixo do nó cotiledonar; altura na maturação (AM) em cm: medida da distância da base até o ápice da planta; número de vagens por planta (NV); número de nós na haste principal por planta (NN); número de hastes laterais por planta (NH); número de grãos por planta (NG); massa de grãos por planta (PROD) em gramas corrigida para 13% de umidade; e, teores de óleo (OL) e de proteína nos grãos (PT) definidos em porcentagem: mensurados em grãos intactos com 13% de umidade, no Laboratório de Genética Molecular de Plantas/Bioagro/UFV, via espectrometria do infravermelho próximo utilizando o equipamento Antaris™ II FT-NIR Analyzer.

A estimação dos parâmetros genéticos e a predição dos valores genéticos dos genótipos foram realizadas utilizando apenas os dados dos genitores e foram procedidas via metodologia de modelos mistos REML/BLUP, descritos por Resende (2007a). Para tais análises, foi utilizado o software SELEGEN (Resende, 2007b) utilizando o modelo:

$$y = Xu + Hg + Tb + e, \text{ onde:}$$

u é o efeito fixo da média geral dos genitores; g é o vetor dos efeitos genéticos aditivos dos genitores (aleatórios); b é o vetor dos efeitos de blocos (aleatórios); e é o vetor dos efeitos do erro (aleatórios). X , H e T são matrizes de incidência.

O teste da razão de verossimilhança (LRT), denominado de análise de deviance (ANADEV), para os efeitos aleatórios, foi realizado considerando o modelo adotado, conforme relatado por Resende et al. (2015). A ANADEV é uma estatística derivada da razão entre as verossimilhanças do modelo completo e do modelo reduzido na ausência do efeito o qual se deseja testar. Para testar a significância dessa razão utilizou-se o teste qui-quadrado com 1 grau de liberdade.

De posse dos valores genotípicos preditos, foram realizadas as estimativas das correlações (r) pela seguinte expressão, originando em uma matriz R :

$$r = \frac{\sqrt{COV_{x,y}}}{\sqrt{V_x}\sqrt{V_y}}$$

Onde:

$COV_{x,y}$ = Covariância entre as variáveis x e y

V_n = Variância da enésima variável

Inicialmente, para a análise de trilha foi utilizada a matriz de correlação das variáveis explicativas R, a qual foi composta por todos os onze caracteres avaliados com exceção do caractere principal OL.

Foi realizado o diagnóstico de multicolinearidade sobre a matriz R, baseado no número de condição (NC) que é a razão entre o maior e o menor autovalor (Montgomery e Peck, 1981) da matriz de correlação das variáveis explicativas. Foram consideradas matrizes livres de multicolinearidade aquelas que apresentaram NC menor que 100. Quando foram observados valores de NC entre 100 e 1000, considerou-se multicolinearidade moderada a forte. Para valores acima de 1000, considerou-se multicolinearidade severa.

A fim de contornar os problemas ocasionados pela multicolinearidade, foram utilizados dois procedimentos: o primeiro foi realizar a exclusão de caracteres da matriz de correlação entre as variáveis explicativas (*trait culling*) e o segundo foi realizar a análise de trilha com multicolinearidade sobre a matriz R por meio da regressão em crista ou em cumeieira conforme descrito por Hoerl e Kennard (1970a, b). Como resultado do primeiro procedimento, foram originadas outras duas matrizes de correlações entre variáveis explicativas: R_1 e R_2 . A matriz R_1 foi composta pelos caracteres DPM, DHP, AM, NN, NG, PROD e PT, enquanto a matriz R_2 foi formada pelos caracteres DPM, DHP, AM, NV, NN, NH, PROD e PT. Após isto, foi realizado um novo diagnóstico de multicolinearidade sobre as matrizes R_1 e R_2 .

Para realização da análise de trilha sob multicolinearidade, conforme foi proposto no segundo procedimento, as estimativas dos coeficientes de trilha foram obtidos pela solução da equação:

$$(X'X + KI_p) \theta^* = X'Y \text{ em que:}$$

$X'X$ = matriz de correlações entre as variáveis independentes do modelo de regressão;

K = constante adicionada aos elementos da diagonal da matriz $X'X$;

I_p = matriz identidade;

θ^* = vetor dos estimadores dos coeficientes de trilha;

$X'Y$ = é a matriz de correlações entre a variável dependente com cada variável independente do modelo de regressão.

O valor do coeficiente (K) para a análise e trilha com multicolinearidade foi estabelecido por meio de gráficos, nos quais foram plotados os valores dos efeitos diretos das características em função dos valores de K priorizando o menor valor de K para todas as estimativas estáveis, conforme recomendado por Cruz et al. (2014).

Como alternativa à análise de trilha por meio da regressão em crista, foi realizada a exclusão de caracteres da matriz de correlação, conforme uma das recomendações feitas por Carvalho et al. (1999) e Cruz et al. (2014). Este segundo procedimento adotado para contornar os problemas com multicolinearidade deu origem a outras duas matrizes, sendo estas matrizes R_1 e R_2 . Ambas matrizes foram compostas por diferentes caracteres resultando um número menor de caracteres do que a matriz R.

As análises de trilha conforme Wright (1921) e a análise de trilha com multicolinearidade foi realizada pelo software GENES (Cruz, 2013).

4.5 RESULTADOS E DISCUSSÃO

Notou-se a existência de variabilidade para todos os 12 caracteres avaliados ao nível de significância de 1% ($P < 0,01$) (Tabela 1). Isto indica que este grupo de cultivares, consideradas elites, pode ser utilizado em programas de melhoramento como fonte de genitores. A existência de variabilidade entre genitores é fundamental para confecção de populações segregantes que apresentem variabilidade e proporcionem ganhos com a seleção (Falconer, 1987).

Os CVe% obtidos para os caracteres estudados foram classificados como baixo, na sua grande maioria, até alto, conforme Pimentel-Gomes e Garcia (2002) e variaram de 2,81353 para PT a 20,3364 para NH. Pimentel-Gomes e Garcia (2002) classificam como baixo, coeficientes abaixo de 10%, médios de 10% a 20%, altos de 20% a 30% e muito altos acima de 30%. No entanto, espera-se estimativas mais altas para CVe% para caracteres os quais são influenciados por maior número de genes tais como produção de grãos. Além disto, espera-se também maiores estimativas de CVe% sobre caracteres os quais são utilizadas medidas

indiretas ou sobre àqueles os quais as mensurações são mais trabalhosas, pois sofrem ação de fatores nem sempre controlados e não contemplados no modelo estatístico. Scapim et al. (1995) critica tal classificação e cita que estas estimativas estão associadas às particularidades da cultura, natureza do caractere e condições edafoclimáticas.

Resende (2002) propõe uma classificação para os valores de acurácia como medida de qualidade experimental, sendo estes altos quando $r_{\hat{a}a} \geq 0,70$, médio quando $0,40 \leq r_{\hat{a}a} \leq 0,70$ e baixo $0,10 \leq r_{\hat{a}a} \leq 0,40$. No presente estudo, foram encontrados valores para $r_{\hat{a}a}$ que variaram de 0,85372 a 0,99192 para os caracteres OL e DPF, respectivamente. Nestas condições os dados tornam-se mais confiáveis à inferência do pesquisador, pois quanto maior o valor da acurácia mais próximos estão os valores genéticos preditos dos valores genéticos verdadeiros (Resende 2002) e dessa maneira, permitem que o melhorista seja mais eficiente ao selecionar genótipos.

A variância do erro de predição dos valores genotípicos (PEV) constitui também um parâmetro para avaliar a qualidade experimental, pois esta é diretamente relacionada com precisão, acurácia experimental e variância genética aditiva (Resende e Perez, 1999; Resende e Duarte, 2007). No entanto, a maior importância dada a PEV se refere à possibilidade do cálculo do desvio padrão do valor genotípico predito e o intervalo de confiança dos valores genéticos preditos que servem como medida de risco ao selecionar um genótipo (Resende e Perez, 1999).

Para os componentes vegetativos, verificou-se a menor herdabilidade para o caractere NH (0,3685). Genótipos de soja apresentam diferentes capacidades de adequar o número de ramificações laterais de acordo com a competição por luminosidade devido aos diferentes espaçamentos e densidades de semeadura, o que é conhecido como plasticidade fenotípica (Tourino et al., 2002). Neste estudo, os vasos foram igualmente espaçados, no entanto, como os tratamentos foram casualizados dentro dos blocos, um mesmo genótipo foi submetido a diferentes níveis de competição entre um bloco e outro. Este efeito é somados aos efeitos ambientais, logo, reduzem a herdabilidade do caractere. Para mitigar a influencia dos efeitos ambientais, recomenda-se o uso de bordadura entre as parcelas a fim de promover a competição de todas as parcelas com um mesmo genótipo.

Tabela 1 – Estimativa de parâmetros genéticos e ambientais para os caracteres: dias para florescimento (DPM), altura no florescimento (AF), dias para maturação (DPF), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de vagens (NV), número de nós (NN), número de hastes (NH), número de grãos por planta (NG), massa de grãos por planta (PROD), teor de óleo (OL) e proteína nos grãos (PT).

	DPF	AF	DPM	DHP	AM	NV
σ^2_g	34,305 **	200,039 **	90,213 **	0,333 **	174,325 **	52,544 **
σ^2_e	2,808	24,409	28,493	0,228	71,898	10,629
σ^2_f	37,113	224,448	118,707	0,561	246,222	63,173
h^2_g	0,924	0,891	0,760	0,593	0,708	0,832
$r_{\hat{a}a}$	0,992	0,988	0,970	0,938	0,961	0,980
CVg%	14,089	32,092	8,0703	11,522	19,473	23,923
CVe%	4,031	11,210	4,535	9,537	12,506	10,760
CVr	3,495	2,863	1,779	1,208	1,557	2,223
PEV	0,552	4,7655	5,360	0,040	13,284	2,043
SEP	0,743	2,183	2,315	0,200	3,645	1,429
Média	41,511	44,071	117,691	5,008	67,803	30,300
	NN	NH	NG	PROD	OL	PT
σ^2_g	1,843 **	0,486 **	151,206 **	2,848 **	0,667 **	1,377 **
σ^2_e	0,544	0,832	65,408	1,795	1,241	1,209
σ^2_f	2,387	1,318	216,614	4,643	1,908	2,585
h^2_g	0,772	0,368	0,698	0,613	0,350	0,532
$r_{\hat{a}a}$	0,972	0,863	0,959	0,942	0,854	0,922
CVg%	11,206	15,536	18,414	14,768	4,279	3,003
CVe%	6,091	20,336	12,111	11,725	5,836	2,813
CVr	1,840	0,764	1,520	1,260	0,733	1,067
PEV	0,103	0,124	12,040	0,319	0,181	0,206
SEP	0,321	0,352	3,470	0,565	0,425	0,453
Média	12,113	4,486	66,777	11,427	19,085	39,074

** P < 0,01, * P < 0,05, ‘ P < 0,1 e ns não-significativo a 1 g.l. pelo teste de χ^2 (1% = 6,63, 5% = 3,84 e 10% = 2,71%), via teste de LRT (Likelihood Ratio Test) dos efeitos aleatórios.

σ^2_g = variância genotípica; σ^2_e = variância residual; σ^2_f = variância fenotípica individual; h^2_g = herdabilidade de parcelas individuais no sentido amplo; h^2_m = herdabilidade da média de genótipo, assumindo ausência de perda de parcelas; $r_{\hat{a}a}$ = acurácia da seleção de genótipos, assumindo ausência de perda de parcelas; CVg% = coeficiente de variação genotípica; CVe% = coeficiente de variação residual; CVr = CVg/CVe = coeficiente de variação relativa; PEV = variância do erro de predição dos valores genotípicos, assumindo ausência de perda de parcelas; SEP = desvio padrão do valor genotípico predito, assumindo ausência de perda de parcelas e Média = Média geral do experimento.

Quanto à massa de grãos por planta e os caracteres de qualidade de grãos OL e PT, apresentaram herdabilidade menores que a maioria dos componentes vegetativos. Isso se deve ao fato destes caracteres serem governados por um grande número de genes, logo, são mais fortemente influenciados pelo ambiente (Lee et al., 1996; Rodrigues et al., 2010; Marquez, 2010).

A razão entre as estimativas de $CV_g\%$ e $CV_e\%$ resultam no coeficiente de variação relativa (CV_r), que quando em magnitude superior a 1, permite inferências sobre a classificação dos genótipos com precisão e acurácia elevada (Resende e Duarte, 2007). No presente trabalho magnitudes superiores a 1 para CV_r só não foram detectadas para os caracteres NH e OL, que foram os caracteres de menores magnitudes de herdabilidades dentre os avaliados, evidenciando a necessidade de seleção indireta para esses caracteres.

Os valores genotípicos preditos para os genótipos quanto aos caracteres avaliados foram apresentados na Tabela 2. Entre os componentes produtivos, NV é o componente que mais influencia PROD (0,79), logo o número de vagens determina o potencial produtivo de um genótipo. As vagens da planta de soja são formadas a partir do desenvolvimento e diferenciação dos óvulos autofecundados das flores, as quais se desenvolvem das gemas axilares em conjuntos (rácemos) nos nós produtivos da planta, localizados principalmente na haste principal. Com isto, o número de vagens por planta (NV), é altamente correlacionado com o número de nós na haste principal (NN) da planta (0,88). Tendo em vista as altas correlações encontradas entre o caractere NN e os componentes vegetativos DPF (0,86) e DPM (0,86), depreende-se que o potencial produtivo de grãos em plantas de soja é influenciado pelo ciclo destas. Estes resultados corroboram com os que foram obtidos por Bizeti et al. (2004). As estimativas das correlações entre os valores genotípicos foram apresentadas na Tabela 3.

As altas e significativas correlações da produtividade de grãos (PROD) com dias para florescimento, dias para maturação, número de nós produtivos e número de vagens por planta concordam com os resultados obtidos por meio das estimativas de correlações genéticas por Machikowa e Laosuwan (2011). Os autores ainda encontraram alto efeito direto do número de vagens sobre a produtividade de grãos. Altas correlações negativas e significativas foram obtidas entre os caracteres PT x NH (-0,56), PT x NG (-0,53) e PT x OL (-0,60).

Tabela 2 – Valores genotípicos (u + g) dos genótipos avaliados quanto aos caracteres: dias para florescimento (DPM), altura no florescimento (AF), dias para maturação (DPF), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de vagens (NV), número de nós (NN), número de hastes (NH), número de grãos por planta (NG), massa de grãos por planta (PROD), teor de óleo (OL) e proteína nos grãos (PT).

Cultivar	Variáveis											
	DPF	AF	DPM	DHP	AM	NV	NN	NH	NG	PROD	OL	PT
	u + g											
BMX Apolo RR	35,96	21,43	106,04	4,45	45,39	20,92	10,06	3,45	54,15	8,97	17,85	41,22
DM 5958 IPRO	34,36	28,23	106,22	3,96	51,21	23,38	10,92	4,35	54,10	9,06	19,40	38,44
NA 5909 RR	35,23	42,74	107,42	4,57	72,35	26,05	10,51	4,92	59,98	10,65	18,79	38,59
TMG 7062 IPRO	36,67	35,59	103,45	4,65	59,53	22,07	10,87	4,09	53,02	10,88	18,87	38,75
DM 6563 IPRO	36,73	30,63	114,09	4,78	61,21	26,55	11,91	3,70	64,50	10,87	18,99	37,99
BMX Potência RR	37,06	29,65	111,26	4,73	59,88	23,96	11,30	4,16	58,59	9,70	18,31	39,60
BMX Desafio RR	34,29	30,80	117,26	5,61	62,71	23,81	11,67	3,94	58,19	10,81	18,10	40,18
M 7739 IPRO	45,40	50,45	119,40	5,54	66,54	33,13	12,29	5,01	76,41	12,49	19,47	37,57
P98Y30 RR	44,80	56,73	123,34	5,03	79,90	37,11	13,64	5,46	80,05	12,93	20,58	38,79
M8349 IPRO	45,35	45,95	124,27	5,06	63,74	35,54	12,58	4,95	82,55	13,92	19,79	37,83
TMG 132 RR	47,36	52,03	128,83	4,81	70,36	36,48	13,28	5,13	80,34	11,17	19,10	38,78
M 9056 RR	51,17	65,07	127,29	5,77	94,12	42,59	15,01	5,04	88,44	13,36	19,51	38,15
M 8221 RR	46,16	50,70	119,67	5,77	69,19	40,83	12,45	4,53	66,76	12,25	19,00	40,28
P98N71	45,82	57,09	131,15	4,85	76,30	31,12	13,09	4,50	60,95	10,92	19,25	39,81
Conquista	46,32	63,97	125,66	5,54	84,62	30,99	12,08	4,05	63,61	13,43	19,25	40,11

u + g = média geral somado aos efeitos genotípicos para cada genótipo.

Tabela 3 – Correlações entre as 12 variáveis estudadas.

Variáveis	DPF		AF		DPM		DHP		AM		NV		NN		NH		NG		PROD		OL		PT
DPF	1,00																						
AF	0,90	**	1,00																				
DPM	0,86	**	0,82	**	1,00																		
DHP	0,62	*	0,61	*	0,59	*	1,00																
AM	0,75	**	0,92	**	0,73	**	0,63	*	1,00														
NV	0,91	**	0,83	**	0,78	**	0,65	**	0,73	**	1,00												
NN	0,86	**	0,80	**	0,86	**	0,59	*	0,79	**	0,88	**	1,00										
NH	0,62	*	0,66	*	0,51	*	0,24		0,58	*	0,72	**	0,65	**	1,00								
NG	0,81	**	0,68	*	0,72	**	0,52	*	0,62	*	0,87	**	0,85	**	0,77	**	1,00						
PROD	0,78	**	0,81	**	0,69	**	0,73	**	0,73	**	0,79	**	0,70	**	0,57	*	0,77	**	1,00				
OL	0,60	*	0,66	*	0,51		0,16		0,54	*	0,65	**	0,64	**	0,75	**	0,67	**	0,67	**	1,00		
PT	-0,19		-0,20		-0,09		0,05		-0,19		-0,28		-0,33		-0,56	*	-0,53	*	-0,36		-0,60	*	1,00

*; ** Significativo a 1% e 5% de probabilidade, respectivamente de acordo com o teste de Mantel.

DPF = dias para florescimento, AF = altura no florescimento, DPM = dias para maturação, DHP = diâmetro do hipocótilo, AM = altura na maturação, NV = número de vagens, NN = número de nós, NH = número de hastes, NG = número de grãos por planta, PROD = massa de grãos por planta, OL = teor de óleo nos grãos e PT = proteína nos grãos.

De acordo com as estimativas de correlação o caractere que mais afetou o teor de óleo nos grãos foi o número de hastes laterais (0,75). O teor de óleo mostrou-se ainda fortemente correlacionado com os componentes produtivos os quais apresentaram alta correlação com a massa de grãos por planta (NV, NN e NG). Isto reforça a alta correlação entre o teor de óleo nos grãos e a massa de grãos por planta.

Utilizando a matriz de correlações R, em que todos os caracteres estavam presentes foi observado multicolinearidade severa (Tabela 4). As novas matrizes R_1 e R_2 foram estabelecidas descartando caracteres que fossem altamente correlacionados com outros a fim de reduzir o nível de severidade de multicolinearidade da matriz R. Esta técnica utilizada para contornar os problemas ocasionados pela multicolinearidade em matrizes de correlação é conhecida como “*trait culling*”. Para fins práticos, pares de caracteres tais como DPF e DPM ou AF e AM são entendidos como redundantes uma vez que apenas um caractere de cada um destes pares de caracteres é suficiente para interpretações biológicas. Os caracteres DPM e AM dão ao pesquisador noções de ciclo e altura da planta no momento da colheita. Como as correlações entre DPF e DPM ou AF e AM são de alta magnitude, estas, contribuem para o aumento da multicolinearidade. Sendo assim, os caracteres DPF e AF agregam poucas informações ao pesquisador e o descarte de tais caracteres reduz o nível de multicolinearidade da matriz. Por este motivo, os pares de caracteres DPF e DPM ou AF e AM não são encontrados nas matrizes R_1 e R_2 . Na Tabela 4 são apresentados os autovetores e os números de condição (NC) para as matrizes R, R_1 e R_2 .

De acordo com os resultados do diagnóstico de multicolinearidade, foi apontada presença de multicolinearidade severa na matriz R (2081,0076). Neste caso, para realizar a análise de trilha com multicolinearidade, foi adotado um valor de K de 0,4452, resultando em valores inferiores à unidade para os efeitos diretos e indiretos dos caracteres sobre OL (Tabela 5). Contudo, o efeito da variável residual (0,5853) se mostrou alto e o coeficiente de determinação (0,6574) não satisfatório. Cruz et al. (2014) afirmam que a estabilização de estimativas em análises de trilha com multicolinearidade ocorre com determinados valores de K, embora quanto maior o valor de K adotado, maior será o viés da análise de regressão e por isso, cabe ao pesquisador escolher o menor valor de K suficiente para estabilização das estimativas.

De forma contrastante, a análise de trilha tradicional sob multicolinearidade apresentada no trabalho de Bizeti et al (2004) com soja, utilizando a matriz de correlação com multicolinearidade severa, não resultou em uma análise satisfatória, pois conduziu a obtenção

de estimativas superiores a 1 para os efeitos diretos de dois caracteres sobre o principal. Em contrapartida, quando os autores adotaram $K = 0,05$, os efeitos diretos destes mesmos dois caracteres, nem sequer foram os mais importantes para explicar as variações do caractere principal e as estimativas se ativeram entre 0 e 1. Além disto, os autores ainda obtiveram coeficiente de determinação negativo e alto efeito residual. Estes resultados deixam clara a importância de se realizar o diagnóstico de multicolinearidade para que o pesquisador possa tomar medidas para contornar este problema quando encontrado. Os resultados evidenciam ainda que a adoção da metodologia proposta por Hoerl e Kennard (1970a, b) deve ser avaliada pelo pesquisador.

Tabela 4 – Autovalores e número de condição (NC) para a matriz de correlação original das variáveis explicativas e para as matrizes de correlação reduzidas.

Ordem	Autovalores		
	R	R ₁	R ₂
1	7,6664	5,1295	4,7195
2	1,4000	1,3583	1,1111
3	0,5902	0,5504	0,3877
4	0,4541	0,3382	0,3361
5	0,3138	0,2824	0,2400
6	0,2258	0,1825	0,1283
7	0,1472	0,1109	0,0774
8	0,1045	0,0477	
9	0,0577		
10	0,0366		
11	0,0037		
NC	2081,0076	105,2794	86,5201
Multicolinearidade	Severa	Moderada a forte	Fraca

R = matriz de correlação original entre as variáveis explicativas; R₁ = matriz de correlação com exclusão de três caracteres entre as variáveis explicativas e R₂ = matriz de correlação com exclusão de quatro caracteres entre as variáveis explicativas.

As estimativas de efeitos diretos e indiretos de todos os caracteres sobre OL realizada sobre a matriz R (Tabela 5) revelam alto efeito da variável residual (EVR) sobre o caractere principal. Nenhum caractere utilizado como variável explicativa apresentou efeito direto ou indireto superior ao EVR. Altos valores estimados de EVR são encontrados quando as variáveis explicativas não são suficientes para explicar o comportamento da variável principal, sendo necessário tomar medidas de um número maior de caracteres para compor o

conjunto de variáveis explicativas. Portanto, o alto EVR e o valor de K empregado na análise com multicolinearidade não permitem ao pesquisador tomar conclusões confiáveis sobre as estimativas de efeitos diretos e indiretos sobre os caracteres que compõe a matriz R.

Na análise de trilha realizada com a matriz R_1 , foram obtidos valores do coeficiente de determinação da regressão (0,8564) e do efeito da variável residual (0,3789) mais satisfatórios em relação aos resultados discutidos anteriormente (Tabela 6). Apesar do número reduzido de variáveis na matriz R_1 quando comparada com a matriz R, foram mantidos caracteres suficientes para interpretações biológicas, além do bom ajuste da regressão. Partindo do ponto de vista estatístico, torna-se viável a exclusão de caracteres redundantes, como por exemplo, a exclusão do caractere DPF em vista de DPM ou AF em vista de AM. Além disto, os caracteres mantidos na matriz R_1 contemplaram componentes vegetativos, reprodutivos e de qualidade de grãos. Apesar das pequenas diferenças entre as estimativas, Bizeti et al. (2004), obtiveram o melhor ajuste da regressão excluindo caracteres altamente correlacionados da análise em vista da análise sob multicolinearidade. Novamente, o resultado, ao se utilizar a matriz R_1 , evidenciou o alto efeito direto negativo de DHP sobre OL, no entanto com baixa correlação. Para a variável NG, foi encontrando, também, alto efeito direto negativo (-0,5556) para o caractere principal, contudo a magnitude da correlação foi de sentido contrário (0,67), logo, não existe uma relação de causa e efeito entre as variáveis, apesar da correlação significativa. Os efeitos indiretos dos caracteres via NN e PROD sobre o caractere principal foram os de maior magnitude. Somado a isto, o alto efeito direto de NN e PROD sobre OL, deixa clara a influência e a relação de causa e efeito entre o número de nós na haste principal e a massa de grãos por planta com o teor de óleo nos grãos. Como a massa de grãos exerce relação de causa e efeito com o teor de óleo nos grãos e o mesmo não pode ser afirmando para o número de grãos por planta, presume-se que o teor de óleo nos grãos esteja relacionado com o peso ou a densidade do grão. Apesar de negativos, altos efeitos indiretos via os caracteres NG e DHP foram notados. Dessa maneira, os caracteres DHP, NN, NG, PROD e PT comporiam um índice de seleção para ganho em OL.

Tabela 5 – Estimativas dos efeitos diretos e indiretos das variáveis dias para florescimento (DPF), altura no florescimento (AF), dias para maturação (DPM), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de vagens (NV), número de nós (NN), número de hastes (NH), número de grãos por planta (NG), massa de grãos por planta (PROD), e teor proteína nos grãos (PT), sobre teor de óleo nos grãos (OL), utilizando a matriz de correlação R sob multicolinearidade severa.

Efeito	Variável explicativa										
	DPF	AF	DPM	DHP	AM	NV	NN	NH	NG	PROD	PT
Direto sobre OL	0,0354	0,1413	0,0190	-0,2065	0,0196	0,0864	0,1046	0,1959	0,0223	0,2158	-0,2051
Indireto via DPF	-	0,0319	0,0306	0,0220	0,0267	0,0324	0,0305	0,0218	0,0287	0,0277	-0,0067
Indireto via AF	0,1271	-	0,1154	0,0860	0,1305	0,1167	0,1128	0,0937	0,0964	0,1138	-0,0285
Indireto via DPM	0,0165	0,0155	-	0,0113	0,0138	0,0149	0,0163	0,0098	0,0136	0,0130	-0,0017
Indireto via DHP	-0,1282	-0,1257	-0,1221	-	-0,1291	-0,1335	-0,1211	-0,0487	-0,1064	-0,1514	-0,0101
Indireto via AM	0,0147	0,0181	0,0142	0,0122	-	0,0143	0,0155	0,0113	0,0122	0,0143	-0,0038
Indireto via NV	0,0789	0,0714	0,0675	0,0559	0,0631	-	0,0757	0,0626	0,0751	0,0679	-0,0241
Indireto via NN	0,0902	0,0835	0,0899	0,0614	0,0827	0,0917	-	0,0675	0,0888	0,0734	-0,0342
Indireto via NH	0,1206	0,1299	0,1008	0,0462	0,1129	0,1420	0,1264	-	0,1507	0,1111	-0,1092
Indireto via NG	0,0181	0,0152	0,0160	0,0115	0,0139	0,0194	0,0189	0,0172	-	0,0172	-0,0118
Indireto via PROD	0,1690	0,1739	0,1480	0,1583	0,1583	0,1696	0,1515	0,1223	0,1668	-	-0,0767
Indireto via PT	0,0389	0,0414	0,0183	-0,0100	0,0395	0,0572	0,0671	0,1143	0,1084	0,0729	-
Total	0,5970	0,6593	0,5060	0,1563	0,5406	0,6495	0,6449	0,7550	0,6665	0,6702	-0,6033
R ²						0,6574					
K						0,4452					
EVR						0,5853					

Total = correlação entre a variável da coluna com o caractere OL; R² = Coeficiente de determinação; K = valor da constante utilizada na análise e; EVR = Efeito da variável residual.

Tabela 6 – Estimativas dos efeitos diretos e indiretos das variáveis: dias para maturação (DPM), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de vagens (NV), número de nós (NN), número de hastes (NH), massa de grãos por planta (PROD), e teor proteína nos grãos (PT), sobre teor de óleo nos grãos (OL), utilizando a matriz de correlação R_1 sob multicolinearidade moderada a forte.

Efeito	Variável explicativa							
	DPM	DHP	AM	NV	NN	NG	PROD	PT
Direto sobre OL	-0,1264	-0,7583	-0,1104	0,4266	0,6317	-0,5556	0,9671	-0,2231
Indireto via DPM	-	-0,0748	-0,0918	-0,0988	-0,1086	-0,0905	-0,0867	0,0113
Indireto via DHP	-0,4486	-	-0,4742	-0,4905	-0,4450	-0,3908	-0,5560	-0,0371
Indireto via AM	-0,0802	-0,0691	-	-0,0807	-0,0873	-0,0689	-0,0810	0,0213
Indireto via NV	0,3333	0,2759	0,3118	-	0,3740	0,3709	0,3352	-0,1189
Indireto via NN	0,5427	0,3707	0,4996	0,5539	-	0,5363	0,4434	-0,2067
Indireto via NG	-0,3978	-0,2863	-0,3466	-0,4830	-0,4717	-	-0,4293	0,2937
Indireto via PROD	0,6630	0,7091	0,7093	0,7599	0,6788	0,7472	-	-0,3437
Indireto via PT	0,1992	-0,0109	0,0429	0,0622	0,0730	0,1179	0,0793	-
Total	0,5060	0,1563	0,5406	0,6495	0,6449	0,6665	0,6702	-0,6033
R ²				0,8564				
EVR				0,3789				

Total = correlação entre a variável da coluna com o caractere OL; R² = Coeficiente de determinação e; EVR = Efeito da variável residual.

Os resultados obtidos dos efeitos diretos e indiretos e as estimativas dos parâmetros da regressão para análise de trilha realizada com a matriz R_2 estão apresentados na Tabela 7. Notou-se um menor efeito da variável residual (0,4096) em relação à análise da matriz R com multicolinearidade severa.

Apesar da aparente melhora das estimativas utilizando uma matriz com menor número de caracteres, verificou-se que o efeito direto da massa de grãos por planta (PROD) sobre o teor de óleo nos grãos (OL) apresentou magnitude superior a 1. Resultados como esse são esperados quando se utiliza matrizes de correlações com multicolinearidade, contudo, foi diagnosticado que a matriz R_2 apresentou multicolinearidade fraca.

Além do caractere PROD, os caracteres NN e PT apresentaram efeito direto sobre OL, já suas estimativas têm sinais iguais aos das respectivas correlações. Quanto aos efeitos indiretos, percebeu-se mais uma vez a influência positiva de PROD sobre OL, já que os efeitos indiretos dos demais caracteres, com exceção de PT, via PROD sobre o caractere principal variaram de 0,7180 a 0,8091 e estão entre os efeitos indiretos de maior magnitude. Fato semelhante, no entanto, em menor magnitude foi notado para os efeitos indiretos dos caracteres via NN. As mesmas conclusões quanto aos efeitos diretos de PROD e NN e efeitos indiretos via estes dois mesmos caracteres na análise de trilha com a matriz R_1 são tomadas na matriz R_2 .

O teor de proteína é relatado na literatura como negativamente correlacionado com teor de óleo (Liang et al., 2010; Hwang et al., 2014), no entanto, verificou-se neste trabalho a existência de um baixo efeito direto entre as variáveis, de maneira que PT influenciou OL principalmente via NN e PROD. Esse resultado evidencia a inclusão do caractere PT na composição de um índice de seleção multivariado para a seleção de genótipos com alto teor de óleo nos grãos.

Notou-se que caracteres que apresentam efeitos diretos de alta magnitude, de modo geral, também têm forte influencia indireta sobre o caractere principal. Quando estes resultados são notados, somados a uma alta correlação entre tal caractere o caractere principal, elucidam a relação de causa e efeito existente.

Tabela 7 – Estimativas dos efeitos diretos e indiretos das variáveis: dias para maturação (DPM), diâmetro do hipocótilo (DHP), altura na maturação (AM), número de nós (NN), número de grãos por planta (NG), massa de grãos por planta (PROD), e teor proteína nos grãos (PT), sobre teor de óleo nos grãos (OL), utilizando a matriz de correlação R_2 sob multicolinearidade fraca.

Efeito	Variável explicativa						
	DPM	DHP	AM	NN	NG	PROD	PT
Direto sobre OL	-0,1580	-0,7444	-0,0948	0,8037	-0,3471	1,0472	-0,1475
Indireto via DPM	-	-0,0935	-0,1147	-0,1357	-0,1131	-0,0183	0,0141
Indireto via DHP	-0,4403	-	-0,4655	-0,4368	-0,3836	-0,5458	-0,0364
Indireto via AM	-0,0688	-0,0593	-	-0,0749	-0,0591	-0,0695	0,0182
Indireto via NN	0,6904	0,4716	0,6356	-	0,6824	0,5641	-0,2602
Indireto via NG	-0,2485	-0,1789	-0,2165	-0,2947	-	-0,2682	0,1835
Indireto via PROD	0,7180	0,7679	0,7681	0,7351	0,8091	-	-0,3722
Indireto via PT	0,0132	-0,0072	0,0290	0,0483	0,0794	0,0541	-
Total	0,5060	0,1563	0,5406	0,6449	0,6665	0,6702	-0,6033
R ²				0,8322			
EVR				0,4096			

Total = correlação entre a variável da coluna com o caractere OL; R² = Coeficiente de determinação e; EVR = Efeito da variável residual.

4.6 CONCLUSÕES

Há uma relação de causa e efeito entre a massa de grãos por planta e o teor de óleo nos grãos. Existe ainda, uma relação de causa e efeito entre o número de nós da haste principal e o teor de óleo nos grãos.

Há uma correlação negativa de alta magnitude, no entanto, de baixo efeito direto do teor de proteína sobre o teor de óleo nos grãos, logo para ganhos em óleo, se faz necessário, a adoção de índices de seleção multivariados.

A exclusão de caracteres redundantes de uma matriz de correlação é mais simples, evita a obtenção resultados viesados e compõe uma alternativa viável para contornar os problemas de multicolinearidade, desde que proporcione bom ajuste da regressão e baixo efeito da variável residual.

4.7 REFERENCIAL BIBLIOGRÁFICO

ALLIPRANDINI, L. F. et al. Understanding soybean maturity groups in Brazil: environment, cultivar classification, and stability. **Crop Science**, v. 49, n. 3, p. 801-808, 2009.

ALCANTARA NETO, F., et al. Análise de trilha do rendimento de grãos de soja na microrregião do Alto Médio Gurguéia. **Comunicata Scientiae**, v. 2, n. 2, p. 107-112, 2011.

BEZERRA, André et al. Importância Econômica. In: SEDIYAMA, Tuneo; SILVA, Felipe; BORÉM, Aluízio. **Soja: do plantio à colheita**. Viçosa: Editora Ufv, 2015. Cap. 1, p. 13.

BIZETI, H. S. et al. Path analysis under multicollinearity in soybean. **Brazilian archives of biology and technology**, v. 47, n. 5, p. 669-676, 2004.

CÂMARA, G.M.S. Ecofisiologia da soja e rendimento. In: CÂMARA, G.M.S. **Soja: tecnologias da produção**. Piracicaba: ESALQ, 1998. p.256-77.

CARVALHO, C. et al. V. Análise de trilha sob multicolinearidade em pimentão. **Pesq. agropec. bras**, v. 34, n. 4, p. 603-613, 1999.

CARVALHO, F.I.F.; LORENCETTI, C.; BENIN, G. **Estimativas e implicações da correlação no melhoramento vegetal**. Pelotas: Editora e Gráfica da UFPel, 2004. 142p.

CONAB. Série Histórica – Soja. 2016. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1252&t=>>. Acesso em: 27 set. 2016.

CRUZ, C. D. **Princípios de genética quantitativa**. UFV, 2005.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4 ed. Viçosa: UFV, 2012. v. 1. 514 p.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 3 ed. Viçosa: UFV, 2014. v. 2. 668 p.

CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**. v.35, n.3, p.271-276, 2013.

FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, 279p. 1987.

FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. 1977.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55-67, 1970a.

HOERL, A. E.; KENNARD, R. W. Ridge regression: applications to nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 69-82, 1970b.

HWANG, E. Y. et al. A genome-wide association study of seed protein and oil content in soybean. **BMC genomics**, v. 15, n. 1, p. 1, 2014.

LAZZAROTTO, J. J.; HIRAKURI, M. H. Evolução e perspectivas de desempenho econômico associadas com a produção de soja nos contextos mundial brasileiro. Londrina: **Embrapa Soja**, p. 46, 2010. (Embrapa Soja. Documentos, 319).

LEE, S. H et al. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. **Theoretical and applied genetics**, v. 92, n. 5, p. 516-523, 1996.

LIANG, Hui-Zhen et al. QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). **Agricultural Sciences in China**, v. 9, n. 8, p. 1108-1116, 2010.

MACHIKOWA, T.; LAOSUWAN, P. Path coefficient analysis for yield of early maturing soybean. **Sonklanakarín Journal of Science and Technology**, v. 33, n. 4, p. 365, 2011.

MARQUEZ, M.C. **Adaptabilidade, estabilidade e diversidade genética de cultivares de soja em três épocas de semeadura em Uberlândia – MG**. 2010. 84p. Tese (Mestrado) – Universidade Federal de Viçosa, Viçosa, MG, Brasil.

MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. New York: Wiley, 1981. 504p.

PIMENTEL-GOMES, F.; GARCIA, C.H. **Estatística aplicada a experimentos agrônômicos e florestais**: exposição com exemplos e orientações pra uso de aplicativos. Piracicaba: Fealq, 2002. 309p.

RESENDE, M. D. V. et al. Selection index with parents, populations, progenies and generations effects in autogamous plant breeding. **Crop Science**, 2015.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182-194, 2007.

RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Embrapa Florestas, Colombo. 561 p. 2007a.

RESENDE, MDV de. Software **SELEGEN-REML/BLUP**: sistema estatístico e seleção computadorizada via modelos lineares mistos. Colombo: Embrapa Florestas, 2007b.

RESENDE, MDV de; PEREZ, J. R. H. R. Melhoramento animal: predição de valores genéticos pelo modelo animal (BLUP) em bovinos de leite, bovinos de corte, ovinos e suínos. **Archives of Veterinary Science**, v. 4, n. 1, p. 17-30, 1999.

RODRIGUES, J. I. Mapeamento de QTL para conteúdos de proteína e óleo em soja. **Pesq. agropec. bras., Brasília**, v. 45, n. 5, p. 472-480, 2010.

SCAPIM, C. A. ; CARVALHO, C. G. P.; CRUZ, C. D. Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, p.683-686, 1995.

SEDIYAMA, T.; SILVA, F.; BORÉM, A. **Soja**: do plantio a colheita. Viçosa: Editora UFRV, 2015.

TOEBE, M; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, v. 48, n. 5, p. 466-477, 2013.

TOURINO, M. C. C.; REZENDE, P. M. de; SALVADOR, N. Espaçamento, densidade e uniformidade de semeadura na produtividade e características agrônômicas da soja. **Pesquisa Agropecuária Brasileira**, Brasília-DF, v. 37, n. 8, p. 1071- 1077, 2002.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.

WRIGHT, S. 1921. Correlation and causation. **Journal of Agricultural Research.** 20: 557-585.

5 CONCLUSÕES GERAIS

O uso de índices de seleção para escolha entre populações F₁ é uma boa alternativa à seleção realizada com base nas capacidades de combinação, quando a CEC não apresenta variabilidade.

O uso do índice de seleção permite a identificação de bons genitores para comporem programas de melhoramento pelo número de combinações selecionadas oriundas de tal genitor.

A escolha dos pesos econômicos e das variáveis a serem inseridas em um índice de seleção são fatores determinantes para o sucesso da escolha entre populações F₁ em soja.

Os estudos de correlação, apesar de importantes, não são suficientes para inferir sobre relações de causa e efeito entre caracteres. No entanto, a análise de trilha fornece informações para determinar as estratégias a serem adotadas na seleção, tais como seleção indireta ou uso de índices de seleção.

Os resultados encontrados para as correlações entre os caracteres de qualidade de grãos e rendimento de grãos são condizentes com a literatura.

A produtividade de grãos em soja apresenta alta correlação positiva e alto efeito direto com o teor de óleo nos grãos.

O teor de proteína apresenta alta correlação negativa com o teor de óleo nos grãos de soja, porém com baixo efeito direto.