

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**ABORDAGEM MULTIVARIADA NA SELEÇÃO DE
PROGÊNIES DE SOJA SUPERIORES E PORTADORAS DO
GENE RR**

Anderson Dallastra
Engenheiro Agrônomo

2013

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**ABORDAGEM MULTIVARIADA NA SELEÇÃO DE
PROGÊNIES DE SOJA SUPERIORES E PORTADORAS DO
GENE RR**

Anderson Dallastra

Orientadora: Prof^a. Dr^a. Sandra Helena Unêda-Trevisoli

Co-orientador: Prof. Dr. Antonio Orlando Di Mauro

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – UNESP, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Agronomia (Genética e Melhoramento de Plantas).

D145a Dallastra, Anderson
Abordagem multivariada na seleção de progênies de soja superiores e portadoras do gene RR / Anderson Dallastra. -- Jaboticabal, 2013
x, 77 p. : il. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2013
Orientadora: Sandra Helena Unêda-Trevisoli
Banca examinadora: Dilermando Perecin, João Antônio da Costa Andrade
Bibliografia

1. Análise multivariada. 2. Resistência ao glifosato. 3. *Glycine max*.
I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 633.34:631.52

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

DADOS CURRICULARES DO AUTOR

ANDERSON DALLASTRA – Nasceu em 23 de abril de 1987 no município de Chopinzinho – Paraná. Em dezembro de 2004 concluiu o curso Técnico em Agropecuária pelo Centro Estadual de Educação Profissional Assis Brasil – CEEPAB, no município de Clevelândia – Paraná. Desde 2005, quando deu início ao curso superior, atua na área de melhoramento genético das culturas de milho e soja. Em 2008 concluiu a graduação em Agronomia pela Faculdade Assis Gurgacz – FAG, no município de Cascavel – Paraná. Em 2010 concluiu a Pós Graduação em nível de especialização em Proteção de Plantas também pela Faculdade Assis Gurgacz – FAG. Obteve o título de Mestre em Agronomia (Genética e Melhoramento de Plantas) em 2013, pela Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal – São Paulo onde foi bolsista CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Atua como Coordenador de Pesquisa na GDM Genética do Brasil (DONMARIO Sementes/Brasmax Genética) desde agosto de 2012.

"Deus nos concede, a cada dia, uma página de vida nova no livro do tempo.
Aquilo que colocarmos nela, corre por nossa conta."

Emmanuel, psicografado por Chico Xavier

Aos meus amados pais

Volmar

Carmen

DEDICO

Aos meus queridos irmãos

Andrei

Emanoel

A minha grande amiga e amada esposa

Eloane

OFEREÇO

SUMÁRIO

	Página
RESUMO	ii
SUMMARY	iii
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1 A cultura da soja – Origem, expansão e importância econômica	22
2.2 Adaptação da cultura e manejo no controle de plantas daninhas	22
2.3 Melhoramento de plantas e especificidades da cultura soja	22
2.4 Cultivares convencionais x tecnologia Roundup Ready (RR) na cultura da soja	22
2.5 Análise multivariada de dados.....	13
2.6 Análise de Componentes Principais – ACP.....	14
2.7 Análise de Agrupamentos – AA.....	16
2.7.1 Análise de Agrupamento pelo método K-Means	18
2.8 Distância Euclidiana	20
2.9 Análise de Agrupamento pelo Método de Ward	21
3. MATERIAL E MÉTODOS	22
3.1 Geração F ₂	224
3.2 Geração F ₃	226
4. RESULTADOS E DISCUSSÃO	32
5. CONCLUSÕES	55
6. REFERÊNCIAS BIBLIOGRÁFICAS	56

ABORDAGEM MULTIVARIADA NA SELEÇÃO DE PROGÊNIES DE SOJA SUPERIORES E PORTADORAS DO GENE RR

RESUMO - O melhoramento genético de plantas é considerado um processo complexo que gera múltiplas informações e, em muitos casos, de difícil compreensão. O objetivo deste trabalho foi selecionar progênies com caracteres superiores provenientes de cruzamentos bi-parentais de soja com fonte de resistência ao glifosato (RR), além de identificar cruzamentos e genitores mais eficientes, por meio de abordagens multivariadas. Além disso, objetivou-se ainda testar a eficiência dos métodos no processo seletivo para múltiplos caracteres de interesse. O experimento foi conduzido no delineamento experimental do tipo famílias com testemunhas intercalares, no ano agrícola 2010/2011 e 2011/2012 em Jaboticabal-SP sendo que, nas populações F₃ foram selecionadas seis plantas fenotipicamente superiores e avaliadas para os caracteres: número de dias para o florescimento (NDF), número de dias para a maturidade (NDM), altura de inserção da primeira vagem (AIV), altura de planta na maturidade (APM), acamamento (Ac), valor agrônômico (VA), número de ramos (NR), número de vagens por planta (NV), peso de cem sementes (PCS), número de sementes por planta (NS) e produção de grãos (PG). Os dados foram analisados utilizando-se o software Statistica 7.0. Os resultados obtidos possibilitaram a seleção de 77 progênies superiores através da Análise de Componentes Principais. A análise de Agrupamentos, pelo do método de K-means, agregou as progênies em seis grupos de acordo com os caracteres de maior importância em cada um e, através do método de Ward, identificou por meio do dendrograma a estrutura de similaridade e divergência entre as progênies selecionadas. Por fim, comparou-se os métodos de agrupamentos e verificou-se que houve concordância entre ambos quanto aos resultados obtidos.

Palavras-Chave: Análise multivariada, resistência ao glifosato, *Glycine max*.

MULTIVARIATE APPROACH IN SELECTING SOYBEAN SUPERIOR PROGENIES CARRYING GENE RR

ABSTRACT - The plant breeding is considered to be a complex process that generates multiple information sources and in many cases, difficult to understand. The aim of this work was to select progenies with superior characters from bi-parental crosses with soy source of resistance to glyphosate (RR), and identify intersections and parents more efficient through multivariate approaches. Furthermore, aimed to further test the efficiency of the methods in the selection process for multiple traits of interest. The experiment was conducted in the experimental design of type families to witness progress in the agricultural year 2010/2011 and 2011/2012 in Jaboticabal being that in F3 populations selected six plants were phenotypically superior and evaluated for the traits: number of days to flowering (NDF), number of days to maturity (NDM), height of the first pod (AIV), plant height at maturity (APM), lodging (Ac), agronomic value (VA), number of branches (NR), number of pods per plant (NV), one hundred seed weight (PCS), number of seeds per plant (NS) and grain yield (GY). Data were analyzed using the software Statistica 7.0. The results allowed the selection of 77 superior progenies by Principal Component Analysis. Cluster analysis by the K-means method, all progenies added into six groups according to the characters in each of utmost importance, and by the method of Ward identified by the dendrogram structure similarity and divergence between the progenies. Finally, we compared the clustering methods and found that there was an agreement between them as to the results obtained.

Keywords: Multivariate analysis, glyphosate resistance, *Glycine max*.

1. INTRODUÇÃO

É indiscutível a importância econômica e social que a cultura da soja representa nos países onde é cultivada. Além de ser mundialmente uma das mais importantes *commodities* do setor agrícola, representa uma excelente fonte de renda aos produtores. Além disso, uma cadeia produtiva formada por indústrias, agroindústrias, comércios, prestação de serviços, transportes e inúmeros outros setores fazem parte do complexo soja.

A grande importância da soja na agricultura brasileira está relacionada em parte pelas condições climáticas favoráveis e pelo bom manejo do solo, e, particularmente pelo alto número de cultivares melhoradas. Este número vem aumentando a cada ano, com cultivares novas e mais produtivas, resistentes a patógenos e com grande expansão e consolidação em novas áreas de cultivo. (EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA, 2006).

De cultura secundária, a soja rapidamente estabeleceu-se como a principal alternativa econômica da agricultura brasileira, impulsionada por picos de preços originados pela falta de estabilidade da oferta de outras fontes de proteína para elaboração de rações para animais e para o consumo humano. O cultivo da soja se expandiu, na década de 1970, para o Sudoeste e o Centro-Oeste e mais recentemente para o norte do Brasil (CUNHA, 1999).

Para a obtenção de genótipos superiores, é necessária a reunião de uma série de atributos favoráveis que confirmem rendimento comparativamente mais elevado e satisfaçam às exigências do mercado. A presença de variabilidade genética no estudo de populações segregantes é fundamental, quando se objetiva sucesso na condução de programas de melhoramento genético. Sendo assim, são de grande importância o conhecimento da natureza do germoplasma, a variabilidade genética disponível e o potencial agrônomo dos genótipos. Nessas condições, pode-se impor um processo seletivo que venha resultar em ganhos genéticos, o que torna viável e exequível o melhoramento genético (ALLARD, 1971).

Nos programas de melhoramento, sabe-se que, dentre as características de maior importância, encontra-se a produtividade. Portanto, os conhecimentos da grandeza de associação entre as características que influenciam na produtividade são primordiais para o melhoramento, pois permitem ao melhorista saber como a seleção de uma característica pode causar alterações nas outras (JOHNSON; ROBINSON; COMSTOCK, 1955; VENCovsky; BARRIGA, 1992).

A maioria dos programas de melhoramento rotineiramente possuem um conjunto de genótipos que constitui a base para o desenvolvimento de combinações genéticas favoráveis (PEEL; RASMUSSEN, 2000). O germoplasma de uma espécie existe separadamente em populações, variedades ou genótipos. O grande desafio do melhorista é reunir em um só genótipo a maior frequência de alelos favoráveis possível, oriundos das diferentes fontes. É óbvio que esta não é uma tarefa fácil e a dificuldade é ainda maior em espécies de autofecundação (REDDY; COMSTOCK, 1976).

A estimativa de área cultivada com lavouras transgênicas mostra que a área total deve atingir 37,1 milhões de hectares no ano agrícola 2012/13, ou seja, um crescimento de 4,6 milhões de hectares (+14,0%) em relação à safra anterior. A soja lidera, ocupando uma área estimada, para 2012/13 de 24 milhões de hectares com eventos biotecnológicos ou 88,8% da estimativa atual de plantio para essa safra. A tecnologia de tolerância a herbicida é líder com aproximadamente 70% da área total com a biotecnologia (CELERES, 2012).

Desta forma, o objetivo do presente trabalho consistiu na seleção de cruzamentos e progênies superiores de soja, sendo estas portadoras de genes de resistência ao glifosato, a partir de análises multivariadas com base em caracteres de importância agrônômica. Além disso, objetivou-se testar a eficiência dos métodos no processo seletivo para múltiplos caracteres de interesse.

2. REVISÃO DE LITERATURA

2.1 A cultura da soja – Origem, expansão e importância econômica

A soja é uma espécie autógama, herbácea, anual, ereta e com porte variável, pertencente à classe das *Dicotiledôneas*, família *Fabaceae* (*Leguminosae*), gênero *Glycine*, subgênero Soja, sendo a espécie cultivada [*Glycine Max* (L). Merrill], se diferenciando de outras espécies semelhantes pela presença de pubescência nas hastes e vagens (CHUANG; SHING, 2008).

Esta espécie é nativa da Ásia sendo considerada uma das culturas mais antigas daquela área. Com base na distribuição de *Glycine ussuriensis*, provável progenitor da soja atualmente cultivada, a origem seria na China, nas regiões Norte e Central (COSTA, 1996).

A soja é originária de clima temperado, com ampla adaptação aos climas subtropicais e tropicais (SEDIYAMA; TEIXEIRA; REIS, 2005). É uma leguminosa anual cultivada comercialmente em mais de 35 países, nas mais diversas condições edafoclimáticas (BORÉM; ALMEIRA; KIIHL, 2009).

Esta cultura tornou-se conhecida na Europa no século XVII, quando pela primeira vez foi semeada no Jardim Botânico de Paris. Foi introduzida, da China, nos Estados Unidos da América do Norte em 1765, por Samuel Bowen, e era cultivada por ele para fabricação de molho de soja e outros derivados. Entretanto, a maior parte da área plantada, no início do cultivo no sul país, foi principalmente para feno, ou, em alguns casos, como forragem verde. Era frequentemente utilizada, também, para fornecimento de nitrogênio para o milho pela adubação verde (COSTA, 1996).

No Brasil, a soja foi introduzida por Gustavo D'Utra em 1882 no estado da Bahia, porém, sem muito sucesso. Em São Paulo, foi cultivada pela primeira vez por Daffert, em 1892, no Instituto Agrônomo de Campinas (CÂMARA, 1998).

O primeiro registro de cultivo de soja no Brasil data de 1914 no município de Santa Rosa - RS. Mas foi somente a partir dos anos 40 que ela adquiriu alguma importância econômica (EMBRAPA, 2004).

Segundo Câmara (1998), em 1976 o Brasil já havia atingido 16% da participação na produção mundial contra os 0,5%, em 1954/58. Isto somente foi possível graças ao elevado valor da soja no mercado internacional e ao intenso trabalho de experimentação sobre os mais variados aspectos da cultura. Entre meados dos anos de 70 e 80, foi significativo o crescimento da cultura da soja na região do Brasil-Central, envolvendo os estados de Mato Grosso do Sul, Mato Grosso e Goiás.

Lazzarotto e Hirakuri (2010) apresentam dados da produção mundial das principais commodities agrícolas segundo o Departamento da Agricultura dos Estados Unidos (USDA) onde percebe-se que a exploração sojícola é a atividade que apresentou a maior expansão, com um incremento de produção de 517%.

O Departamento de Agricultura dos Estados Unidos (USDA, 2013) divulgou o relatório mensal de março de 2013 com o quadro de oferta e demanda mundial das principais commodities agrícolas para a safra 2012/13, sendo que, a produção da soja está estimada em 264,3 milhões de toneladas.

Embora o Brasil responda por aproximadamente 23,54% da área mundial de soja, o cultivo nacional deste produto ocupa apenas 2,84% do território brasileiro (CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO, 2011). Em termos mundiais, ressalta-se que, o Brasil participa com cerca de 28,6% e 32,8%, respectivamente, da produção e da exportação de soja em grão (USDA, 2011).

Considerando as principais atividades agrícolas, de acordo com dados do relatório do mês de março da Companhia Nacional de Abastecimento (CONAB, 2013), a área cultivada com soja representa 52,2% da área total de produção de grãos do País, ou seja, 27,65 milhões de hectares, sendo 10,4% superior ao ano agrícola anterior. Já a produção está estimada em 82,06 milhões de toneladas e é 23,6% superior ao ano agrícola anterior.

O espetacular crescimento da produção de soja no país determinou uma cadeia de mudanças sem precedentes na história da agricultura brasileira. Foi a grande responsável pelo estabelecimento da agricultura comercial no Brasil, acelerar a mecanização das lavouras brasileiras, modernizar o sistema de transportes, por expandir a fronteira agrícola, acelerar a urbanização do país, por interiorizar a

população brasileira (excessivamente concentrada no sul, sudeste e litoral do nordeste) (DALL´AGNOL, 2011).

Embora haja, a cada ano, uma oscilação no que se refere à produção da cultura, Vencato (2010), relata que o cenário otimista de um país que tem para onde e como crescer a sua produção, projeta um salto produtivo na cultura de mais de 40% até 2020, enquanto que nos Estados Unidos, atualmente o maior produtor mundial, o crescimento no mesmo período deverá ser no máximo de 15%. Com essa projeção, o Brasil atingirá a produção de mais de 105 milhões de toneladas, quando será isoladamente o maior produtor mundial dessa commodity.

2. 2 Adaptação da cultura e manejo no controle de plantas daninhas

A soja melhor se adapta a temperaturas entre 20°C e 30°C. O florescimento da soja somente é induzido quando ocorrem temperaturas acima de 13°C. As diferenças de data de florescimento, entre anos, apresentadas por uma cultivar semeada numa mesma época e na mesma latitude, são devido às variações de temperatura (EMBRAPA, 2011).

De acordo com Sedyama et al. (1989) e Yuyama (1991), fatores ambientais como fotoperíodo, temperatura, radiação solar, vento, umidade, fertilidade do solo, época de semeadura e densidade de plantas, podem afetar diretamente a altura da planta, o grau de acamamento, produtividade e conseqüentemente influenciar na manifestação do máximo potencial genético de uma cultivar.

A adaptação de diferentes cultivares a determinadas regiões depende, além das exigências hídricas e térmicas, de sua exigência de fotoperíodo. A sensibilidade ao fotoperíodo é característica variável entre cultivares, ou seja, cada cultivar possui seu fotoperíodo crítico, acima do qual o florescimento é atrasado (EMBRAPA, 2011).

Um grande problema enfrentado, principalmente no estágio inicial da implantação da cultura, diz respeito às plantas daninhas. Estas invasoras competem com a cultura principalmente por luminosidade, umidade e nutrientes, resultando no

desenvolvimento vegetativo inadequado, fato que conseqüentemente prejudica a produtividade final.

O controle de plantas daninhas na cultura da soja deve ter como objetivo não somente a busca de maior produtividade, mas também a conservação do solo e a preservação do seu potencial produtivo (COSTA, 1996).

Segundo Silva et al. (2007), além da redução da produtividade das culturas, as plantas daninhas causam outros prejuízos diretos como: redução da qualidade do produto comercial podendo causar a não certificação das sementes quando são colhidas junto com determinadas espécies, podem intoxicar animais domésticos em pastagens, hospedar alternativamente organismos nocivos a espécies vegetais cultivadas e ainda reduzir o valor da terra por causar um alto custo no controle e até tornar economicamente inviável.

Para exercer um controle racional das plantas daninhas, é importante que se conheça o período crítico de competição, ou seja, o período durante o qual sua presença resulta em maiores prejuízos para a cultura. Dados existentes na literatura mostram que o período mais importante engloba o terço inicial do ciclo vegetativo de culturas cujas partes aéreas permitam boa cobertura do solo. De um modo geral, os estudos visam determinar qual o período, a partir da emergência da cultura, em que esta deve permanecer livre da presença de plantas daninhas, para que a produtividade não sofra perdas significativas e para evitar a competição em fases reprodutivas posteriores (COSTA, 1996).

Como relata Costa (1996), as cultivares de soja variam em relação a sua habilidade competitiva, assim como a pressão de competição sobre a comunidade infestante está relacionada com o espaçamento entre linhas e à população de plantas. A composição específica, a densidade e a distribuição das plantas daninhas também são importantes na determinação do grau de competição. Estudos realizados em condições brasileiras mostram que o período durante o qual a lavoura de soja deve ficar livre de competição das plantas daninhas vai de 30 a 50 dias após a emergência, variável com as condições predominantes nos diferentes trabalhos conduzidos (COSTA, 1996).

Os métodos normalmente utilizados para controlar as invasoras são o preventivo, o mecânico, o químico e o cultural. Quando possível, é aconselhável

utilizar a combinação de dois ou mais métodos. O método mais utilizado para controlar as invasoras é o químico, isto é, o uso de herbicidas. Suas vantagens são a economia de mão de obra e a rapidez na aplicação. Para que a aplicação dos herbicidas seja segura, eficiente e econômica, exigem-se técnicas adequadas. O reconhecimento prévio das invasoras predominantes é condição básica para a escolha adequada do produto que resultará no controle mais eficiente das invasoras (EMBRAPA, 2011).

Como relata Silva et al. (2007), cultivares transgênicas de soja e de outras culturas foram lançados recentemente no mercado. Depois disso, milhares de hectares foram incorporados ao processo produtivo. A soja tolerante a herbicida é atualmente a cultura geneticamente modificada predominante.

2.3 Melhoramento de plantas e especificidades da cultura da soja

Melhoramento de plantas é arte e ciência (FEHR, 1987). Isto porque, além de requerer um aguçado instinto observativo por parte do melhorista, é necessário um envolvimento de diversos campos do conhecimento, tal como; genética, biologia molecular, bioquímica, fisiologia, estatística, botânica, fitopatologia, entomologia e demais outras áreas ligadas à agronomia.

O progresso no melhoramento de plantas é dependente da habilidade em selecionar genótipos superiores, dentro de progênies homozigotas e heterozigotas oriundas de cruzamentos entre genitores divergentes. As populações de melhoramento são de tamanho finito e diferentes processos seletivos são adotados para que o genótipo superior seja obtido (FEHR, 1993).

Mesmo com a adição de novas tecnologias como ferramentas para os trabalhos de melhoramento de plantas, esta tem sido conduzida com alguns objetivos específicos, porém sua missão é a elevação do valor econômico das espécies (BORÉM; MIRANDA, 2005).

Os programas de melhoramento são embasados em objetivos gerais e, ou, específicos estabelecidos para solução de problemas reais ou potenciais da cultura.

Hibridações são realizadas para desenvolver germoplasmas com variabilidade genética onde são realizadas as seleções para características desejadas. Métodos tradicionais de melhoramento de plantas autógamas são utilizados para desenvolver novas cultivares de soja – introdução de germoplasma melhorado, genealógico, população (*bulk*), genealógico modificado (SSD) e retrocruzamento. Variações e/ou combinações de métodos são bastante utilizados pelos melhoristas durante o processo de avanço de gerações das populações segregantes. Técnicas de biotecnologia e engenharia genética, como marcadores moleculares e transformação de plantas, são importantes ferramentas tecnológicas de grande aplicação nos processos de seleção e aumento de variabilidade genética (ALMEIDA; KIIHL, 1998).

De acordo com Borém e Miranda (2005), estima-se que metade do incremento da produtividade das principais espécies agrônômicas nos últimos cinquenta anos seja atribuída ao melhoramento genético de plantas.

O melhoramento genético de soja no Brasil iniciou em 1930, na antiga Estação Experimental Fitotécnica, no município de Veranópolis, RS, onde surgiu a primeira cultivar desenvolvida lançada somente em 1960 com o nome de Pioneira (MAGALHÃES, 1981).

Segundo Verneti et al., (1981), enquanto o programa de melhoramento realizava o trabalho de hibridação e seleção de novas cultivares específicas para o Brasil, os agricultores utilizavam as melhores introduções como Hill, Hood, Majos, Bragg, Davis, Jew 45, Hampton, Hardee e Bienville, para substituírem pelo menos em parte as cultivares já plantadas como a Amarela do Rio Grande, Santa Rosa e Industrial que chegaram a ocupar uma grande área no estado.

Para o melhoramento é interessante a obtenção de grande variabilidade genética nas plantas para a imposição de processos seletivos que efetivamente resultem em ganhos significativos (BERNARDO, 2002).

Cultivares melhoradas portadoras de genes capazes de expressar alta produtividade, ampla adaptação e boa resistência/tolerância a fatores bióticos ou abióticos adversos representam usualmente a contribuição mais significativa à eficiência do setor produtivo. O ganho genético proporcionado pelas novas cultivares

ao setor produtivo tem sido muito significativo – cerca de 1,38% ao ano (EMBRAPA, 2011).

O conhecimento da natureza e magnitude dos efeitos gênicos que controlam um caráter é primordial para o processo de seleção e a predição do comportamento das gerações segregantes (BACKES et al., 2002).

Uma boa cultivar de soja deve ter alta produtividade e estabilidade de produção em ambientes mais variados possíveis. A estabilidade é conferida pela introdução de resistência a doenças, nematóides e insetos e pela introdução de características agrônômicas especiais como tolerância a solos ácidos, penetração profunda das raízes e alta qualidade fisiológica das sementes, permitindo à planta tolerar os fatores adversos que podem comprometer a produção (ALMEIDA; KIIHL, 1998).

Contudo a seleção de progênies superiores é um processo complexo, uma vez que os caracteres de importância, em sua maioria quantitativos, são influenciados pelo ambiente e inter-relacionados, de tal forma que a seleção de um provoca uma série de mudanças em outros (CRUZ, 2001).

Para uma cultivar ser selecionada, deverá apresentar principalmente elevado rendimento de grãos, porém outros aspectos como número de dias para florescimento e maturidade, altura de planta no florescimento e maturidade, altura de inserção da primeira vagem, índice de acamamento, resistência a doenças e pragas, não devem ser desprezados (REZENDE; CARVALHO, 2007).

2.4 Cultivares convencionais x tecnologia Roundup Ready (RR) na cultura da soja

O cultivo de genótipos convencionais apresenta susceptibilidade das cultivares utilizadas a determinados herbicidas, sendo necessário muitas vezes, usar misturas de diferentes princípios ativos para controlar plantas daninhas, tanto em

aplicações de pré-emergência como em pós-emergência, diferenciando-se ainda quanto a invasoras de folhas largas e estreitas e estágio de desenvolvimento.

Segundo Marinis (1971), o conceito de planta daninha tem sido muito discutido, mas em geral a definição dada pelos autores se baseia no princípio da indesejabilidade em relação ao homem.

Um resultado comumente observado na aplicação de herbicidas em cultivos convencionais é com relação à sensibilidade da cultura aos herbicidas. Muitas vezes, as plantas sofrem injúrias demasiadamente altas, podendo comprometer o desenvolvimento e a produtividade, a depender do momento da aplicação, condições ambientais, susceptibilidade da cultivar e característica do herbicida utilizado.

A demanda cada vez maior de produção de alimentos, fibras e energia por parte de uma população sempre crescente de consumidores e decrescente de produtores, destaca a importância do controle de plantas daninhas, assim como de pragas e doenças nas lavouras (DEUBER, 1992).

De acordo com James (2012), diretor do ISAAA - do International Service for the Acquisition of Agri-Biotech Application, um recorde de 170,3 milhões de hectares de cultivares transgênicos foi plantado mundialmente em 2012, a uma taxa anual de crescimento de 6%, 10,3 milhões a mais do que os 160 milhões de hectares registrados em 2011. O Brasil, pelo quarto ano consecutivo, foi o propulsor do crescimento mundial, aumentando sua área cultivada de transgênicos mais do que qualquer outro país – um aumento recorde impressionante de 6,3 milhões de hectares, 21% a mais do que em 2011, alcançando 36,6 milhões de hectares.

Embora a adoção de biotecnologias atinja patamares cada vez maiores, o cultivo da soja convencional continuará sendo exercido, seja pela demanda de países que consomem apenas esse tipo de cultura, podendo se tornar futuramente em um nicho de mercado, ou pelo fato de que possam ocorrer restrições a determinadas biotecnologias nos países consumidores.

De acordo com Barros, Silva e Sedyama (2009), para evitar os prejuízos causados pelas plantas daninhas, gasta-se, em média, cerca de 20-30% do custo de produção de uma lavoura no controle dessas plantas.

Segundo Borém e Miranda (2005), após um extensivo investimento em pesquisas com biotecnologia, as cultivares transgênicas tornaram-se comercialmente disponíveis. A primeira geração de genótipos disponibilizados apresentou tolerância a herbicidas e resistência a pragas e doenças

Embora a produção nacional seja alta, em determinadas regiões a produtividade média ainda é baixa, fato este que pode estar associado à falta de tecnologia aplicada, utilização de cultivares não recomendadas, ultrapassadas ou que apresentam características agronômicas inferiores e, principalmente, ao manejo incorreto da cultura

Um grande problema enfrentado, principalmente no estágio inicial da implantação da cultura, diz respeito às plantas daninhas. Estas invasoras competem com a cultura principalmente por luminosidade, umidade e nutrientes, resultando no desenvolvimento vegetativo inadequado e conseqüentemente na produtividade final.

De acordo com Procópio et al. (2006), a aplicação de herbicidas em dessecação de manejo tornou-se prática obrigatória em cultivos realizados no sistema de plantio direto.

O principal herbicida utilizado nessa operação é o glyphosate (glifosato), pertencente ao grupo químico dos derivados da glicina (MENDELSON, 1998) atuando como um potente inibidor da atividade da 5-enolpiruvilshiquimato-3-fosfato sintase (EPSPS), que é catalisadora de uma das reações de síntese dos aminoácidos aromáticos fenilalanina, tirosina e triptofano, influenciam também outros processos, como a inibição da síntese de clorofila, estimula a produção de etileno, reduz a síntese de proteínas e eleva a concentração do IAA (COLE, 1985; RODRIGUES, 1994).

Culturas tolerantes a herbicidas oferecem aos produtores uma ferramenta para o combate das plantas daninhas e são compatíveis com os métodos de plantio direto, que ajudam na conservação do solo. Além disso, dão aos produtores a flexibilidade de aplicá-los somente quando necessário, de reduzir o seu uso e de utilizar herbicidas com características ambientais desejáveis. Dentre os produtos que poderiam se encaixar nessa prática está o herbicida não seletivo Roundup^{®1} que controla efetivamente uma grande variedade de gramíneas e plantas de folha larga.

¹ ®Roundup é marca registrada utilizada sob licença da Monsanto Company

Em outras palavras, o Roundup inibe o crescimento por um bloqueio nas vias metabólicas das plantas (JEZOVSEK, 1998).

Ainda segundo Jezovsek (1998), pesquisadores da Monsanto nos Estados Unidos descobriram em uma bactéria do solo do gênero *Agrobacterium*, uma enzima EPSPS que não é bloqueada pelo Roundup. Após essa descoberta, o gene foi identificado e clonado, para em seguida ser transferido para algumas espécies vegetais, dentre elas a soja. O gene inserido nessas culturas tolerantes ao herbicida – conhecidas como Roundup Ready^{®2} - faz com que as plantas produzam um tipo diferente de EPSP sintetase, que não é inibida pelo herbicida, fornecendo um desvio ao redor do bloqueio. Este desvio torna possível que as culturas Roundup Ready desenvolvam-se mesmo após o Roundup ser aplicado sobre elas para controlar plantas daninhas.

As principais culturas geneticamente modificadas cultivadas no Brasil que fazem com que estes dados sejam tão expressivos são principalmente, a soja, o milho e o algodão.

Dos 30,3 milhões de hectares de culturas transgênicas cultivadas no Brasil em 2011, 20,6 milhões de hectares foram plantados com soja RR, ou seja, 68% do total (JAMES, 2012).

Um dado muito interessante relatado por James (2010) é com relação à disponibilidade no mercado de novas cultivares. Sendo que, de um total de 95 cultivares registradas para venda em 2009, dos quais 71, equivalente a 75%, foram de soja RR e as 24 cultivares restantes (25%) foram de convencionais. Em 2010, de janeiro a novembro, outras 53 cultivares de soja foram aprovadas no Brasil, dos quais 44 foram de soja RR ou 83% do total. Desde que a soja RR foi aprovada e comercializada em 2003 (aprovada em 1998, mas comercializada só em 2003) um total de 467 novas cultivares foram aprovadas, das quais 290 ou 62% eram de biotecnológicas e 177 ou 38% eram convencionais.

A associação de melhores cultivares com inovações tecnológicas, principalmente relacionadas a biotecnologias, fornecem ao mercado cultivares agronomicamente superiores, com características incorporadas ao seu genoma,

² ®Roundup Ready é marca registrada utilizada sob licença da Monsanto Company

possibilitando expressar com melhor eficiência suas características e, diretamente ou indiretamente aumentar os índices de produtividade.

2.5 ANÁLISE MULTIVARIADA DE DADOS

Os métodos estatísticos, para analisar variáveis, estão dispostos em dois grupos: um que trata da estatística, que olha as variáveis de maneira isolada – a estatística univariada, e outro que olha as variáveis de forma conjunta – a estatística multivariada (VICINI, 2005).

Ainda segundo Vicini (2005), em relação à análise univariada, quando um fenômeno depende de muitas variáveis, geralmente esse tipo de análise falha, pois não basta conhecer informações estatísticas isoladas, mas é necessário, também, conhecer a totalidade dessas informações fornecidas pelo conjunto das variáveis e suas relações. Quando as relações existentes entre as variáveis não são percebidas, efeitos desconhecidos, entre variáveis, dificultam a interpretação do fenômeno a partir das variáveis consideradas.

As análises multivariadas se referem a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre os indivíduos ou objeto de investigação. Muitas técnicas multivariadas são extensões da análise univariada e da análise bivariada. Analogamente, a variável dependente única encontrada na análise de variância é estendida para incluir múltiplas variáveis dependentes em análise multivariada de variância. Algumas técnicas multivariadas fornecem um meio de executar em uma única análise aquilo que antes exigia múltiplas análises univariadas para ser realizado (HAIR et al., 2009).

As variáveis devem ser aleatórias e inter-relacionadas de maneira que seus diferentes efeitos não podem ser significativamente interpretados de forma separada. Os métodos multivariados tornam possível levantar questões específicas e precisas de considerável complexidade em conjunto de dados transformando informação *m*-dimensional em tri ou bidimensional uma vez que a capacidade

humana de identificação por reconhecimento visual só é possível até a terceira dimensão (FERRAUDO, 2010).

Com relação às variáveis, estas podem ser divididas e classificadas como dependentes e independentes segundo a descrição feita por Hair et al., (2005), sendo que, técnicas de dependência podem ser definidas como uma metodologia na qual uma variável ou conjunto de variáveis é identificada como a variável dependente a ser prevista ou explicada por outras variáveis conhecidas como variáveis independentes. As diferentes técnicas de dependência podem ser categorizadas por duas características: o número de variáveis dependentes e o tipo de escala de medida empregada pelas variáveis. Dentro dessas categorizações podem ainda ser: análise discriminante, análise conjunta, análise de correlação canônica, análise multivariada de variância, modelagem de equações estruturais e regressão múltipla. Por outro lado, as técnicas de independência podem ser definidas como uma metodologia em que nenhuma variável ou grupo de variáveis é definida como sendo independente ou dependente. Ao invés disso, o procedimento envolve a análise simultânea de todas as variáveis no conjunto. Ainda há uma categorização de técnicas de independência, sendo: análise de agrupamentos, análise de componentes principais, análises de correspondência, análise fatorial e escalonamento multidimensional.

2.6 Análise de Componentes Principais – ACP

Análise de Componentes Principais (ACP) é executada com o objetivo de simplificar a descrição de um conjunto de variáveis interrelacionadas. A técnica busca reduzir o espaço de variáveis criando eixos ortogonais que são combinações lineares das variáveis originais denominados componentes principais. É medido o poder de cada variável no seu respectivo componente, o que permite diminuir a estrutura de variáveis originais numa nova estrutura menor que a inicial. As unidades amostrais (ou objetos) são distribuídas em gráficos bidimensionais ou

tridimensionais onde os eixos ortogonais são os componentes principais (FERRAUDO, 2010).

Este tipo de análise nada mais é do que o cálculo dos autovalores e correspondentes autovetores de uma matriz de variâncias-covariâncias ou de uma matriz de coeficientes de correlação entre variáveis (LANDIM, 2000).

Segundo Landim (2000), a técnica consiste numa transformação linear de m variáveis originais em m novas variáveis, de tal modo que a primeira nova variável computada seja responsável pela maior variação possível existente no conjunto de dados, a segunda pela maior variação possível restante e assim por diante até que toda variação do conjunto tenha sido explicada. Portanto, trata-se de uma técnica de transformação de variáveis. Se cada variável medida pode ser considerada como um eixo de variabilidade, estando usualmente correlacionada com outras variáveis, esta análise transforma os dados de tal modo a descrever a mesma variabilidade existente, com o mesmo número de eixos originais, porém não mais correlacionados entre si. Graficamente pode ser descrita como a rotação de pontos existentes num espaço multidimensional originando eixos, ou componentes principais, que dispostos num espaço a duas dimensões representem variabilidade suficiente que possa indicar algum padrão a ser interpretado.

Os componentes principais são não correlacionados o que é interessante, pois um pesquisador estando com um problema envolvendo variáveis originais de complexo inter-relacionamento pode analisar um conjunto menor de variáveis não correlacionadas que são os componentes principais. Num conjunto grande de variáveis nem todas tem quantidade de informação relevante podendo através da ACP selecionar aquelas que mais possuem quantidade de informação relevante (FERRAUDO, 2010).

O primeiro autovalor a ser determinado corresponderá à maior porcentagem da variabilidade total presente e assim sucessivamente. Geralmente os dois ou três primeiros autovetores encontrados explicarão a maior parte da variabilidade presente. Os autovetores correspondem às componentes principais e é o resultado do carregamento das variáveis originais em cada um deles. Tais carregamentos podem ser considerados como uma medida da relativa importância de cada variável em relação às componentes principais e os respectivos sinais, se positivos ou

negativos, indicam relações diretamente e inversamente proporcionais (LANDIM, 2000).

Ainda segundo Landim (2000), a matriz de carregamentos de cada variável nas componentes principais ao ser multiplicada pela matriz original de dados fornecerá a matriz de contagens (*scores*) de cada caso em relação às componentes principais. Esses valores poderão então ser dispostos num diagrama de dispersão, em que os eixos são os dois componentes mais importantes, e mostrar o relacionamento entre os casos condicionados pelas variáveis medidas.

Análise de Componentes Principais é considerada uma técnica estatística exploratória utilizada na tentativa de compreender o interrelacionamento entre as variáveis originais (FERRAUDO, 2010).

Como relata Cruz e Carneiro (2003), a técnica de Componentes Principais tem sido de grande utilidade no melhoramento genético.

2.7 Análise de Agrupamentos – AA

A análise de agrupamentos quer seja por métodos hierárquicos ou não fornece aos pesquisadores um método empírico e objetivo para realizar classificação de objetos em grupos. É uma poderosa ferramenta analítica aplicada em todas as áreas do conhecimento. Quando utilizada com conhecimento tem o potencial de revelar estruturas em banco de dados que não poderiam ser descobertas por outros meios (FERRAUDO, 2010).

Análise de Agrupamento é o nome para um grupo de técnicas multivariadas cuja finalidade primária é agregar objetos com base nas características que eles possuem, ou seja, classifica objetos de modo que cada objeto é muito semelhante aos outros no agrupamento em relação a algum critério de seleção pré-determinado. Os agrupamentos resultantes devem exibir elevada homogeneidade interna (dentro dos grupos) e elevada heterogeneidade externa (entre os grupos) (HAIR et al., 2005).

A tarefa de clusterização (ou agrupamento) consiste em particionar os registros da base de dados em subconjuntos (*clusters*) de maneira que elementos presentes em um cluster compartilhem um conjunto de propriedades comuns e que os diferenciem dos elementos de outros clusters. A análise de um cluster está envolvida principalmente com a organização de um conjunto de padrões conforme a medida de similaridade definida. Em geral o conjunto de dados utilizados para efetuar a clusterização não possui uma classe pré-definida, a clusterização pode ser utilizada para definir classes para o conjunto em análise (HAN; KAMBER 2006).

Segundo Davis (1986), os diversos métodos para a análise de agrupamentos podem ser enquadrados em quatro tipos gerais:

- a) Métodos de partição: procura classificar regiões no espaço, definido em função de variáveis, que sejam densamente ocupados em termos de observações daqueles com ocupação mais esparsa.
- a) Métodos com origem arbitrária: procuram classificar as observações segundo “k” conjuntos previamente definidos; neste caso k pontos arbitrários servirão como centróides iniciais e as observações irão se agrupando, por similaridade, em torno desses centróides para formar agrupamentos.
- b) Métodos por similaridade mútua: procuram agrupar observações que tenham uma similaridade comum com outras observações; inicialmente uma matriz $n \times n$ de similaridades entre todos os pares da observação é calculada; em seguida, as similaridades entre colunas são repetidamente recalculadas; colunas representando membros de um único agrupamento tenderão apresentar intercorrelações próximas a 1 (um) e valores menores com não membros.
- c) Métodos por agrupamentos hierárquicos: a partir da matriz inicial de dados obtém-se uma matriz simétrica de similaridades e inicia-se a detecção de pares de casos com a mais alta similaridade, ou a mais baixa distância; para essa combinação, segundo níveis hierárquicos de similaridade, escolhe-se entre os diversos procedimentos aglomerativo de tal modo que cada ciclo de agrupamento obedeça a uma ordem sucessiva no sentido do decréscimo de similaridade.

2.7.1 Análise de Agrupamento pelo método K-Means

É um método de agrupamento não-hierárquico por repartição, consistindo de um procedimento onde, dado um número de *clusters* previamente determinado, calcula pontos que representam os “centros” destes *clusters* e que são espalhados homogeneamente no conjunto de respostas obtidas por um questionário e movidos, heurísticamente, até alcançar um equilíbrio estático (HAIR et al., 2005). Procede-se, em seguida, a uma divisão de todos os casos obtidos pelos k grupos preestabelecidos e a melhor partição dos n casos será aquela que otimize o critério escolhido. Em outras palavras, o procedimento inicia-se usando os valores dos primeiros k casos como estimativas temporárias das médias dos clusters (GUIDINI et al., 2008).

Como usualmente o K-means utiliza medidas diretas de distância para calcular a similaridade entre os objetos, torna-se muito sensível a dados fora do padrão, ou seja, objetos que tenham um valor extremamente alto ou baixo, o que pode causar relevantes distorções nos resultados, com formação de agrupamentos de baixa qualidade de uma maneira geral (ANDRADE, 2004).

Para dar início a análise, o parâmetro k , que representa o número de grupos que se deseja formar com os dados a serem analisados deve ser indicado pelo pesquisador.

Os centróides iniciais são formados em função da designação de cada caso ao cluster de centro mais próximo. Com a inclusão de cada caso, a média altera-se, alterando assim o centróide. Um processo iterativo é usado para achar os centróides finais de cada *cluster*. A cada passo, os casos são agrupados ao *cluster* de centro mais próximo, e novamente as médias são recalculadas. Este processo continua até que não haja mais alterações nas médias ou que um número pré-determinado de interações aconteça, encerrando-se o processo (HAIR et al., 2005).

Assim, o primeiro passo do processamento propriamente dito será selecionar k objetos do conjunto de dados, aleatoriamente. Cada um desses objetos será o representante de seu grupo ainda unitário. Em sequência, os demais objetos serão testados em cada grupo e serão alocados naquele ao qual tenham a maior

similaridade, ou seja, a menor distância. A cada novo objeto alocado será calculada a média dos objetos do grupo – o centróide, contra quem será medida a distância dos objetos que ainda não foram alocados. O modo de processamento do algoritmo K-means pode ser descrito em breves passos como relatado por Andrade (2004):

1. Primeiramente é esperada a entrada do parâmetro k , representando o número de grupos que se deseja formar;
2. São selecionados aleatoriamente, dentre os objetos da coleção a ser analisado, o mesmo número k de objetos que serão, inicialmente, grupos unitários e, por conseguinte, o próprio centro do grupo;
3. Em sequência, os objetos são comparados, um a um, com cada um dos centros dos grupos, sendo o objeto do momento alocado no grupo cujo centro lhe seja o mais próximo, utilizando para isso uma das medidas de distância;
4. Todo grupo que recebe um novo objeto tem seu centro recalculado, por meio da média dos objetos que o compõem;
5. O processo é repetido iterativamente até que uma determinada função objetivo convirja. Este critério procura formar grupos compactos e separados entre si.

A maior desvantagem do método K-means certamente é a exigência do valor de k como parâmetro inicial. Para uma aplicação real, em grandes bases de dados, esta informação normalmente é desconhecida, podendo ser necessária uma prévia análise de um especialista para que haja possibilidade de percepção da quantidade de grupos em que a coleção poderá dividir-se de forma satisfatória, o que por si só já representa uma grande carga de trabalho que, até mesmo, diante de uma base muito grande poderia tornar a análise inviável ou desinteressante. (ANDRADE, 2004).

Ao aplicar um critério de otimização que divida uma amostra em k grupos homogêneos, pretende-se que, dentro de cada grupo, os elementos sejam os mais semelhantes possíveis entre si, ao passo que as semelhanças entre os elementos de grupos distintos sejam as menores possíveis (LOURENÇO; MATIAS, 2000).

Como relata Ramos (2006), um conceito básico para se proceder à realização de um agrupamento é a noção de distância. A diferença ou semelhança entre registros traduz-se na medida da distância entre eles a partir do centro do cluster.

São utilizadas diversas medidas de distância, sendo as mais comumente usadas a distância Euclidiana, distância de Mahalanobis, distância de Manhattan entre outras.

Embora, do ponto de vista matemático, a Análise de Agrupamentos represente uma metodologia bastante simples e a Análise de Componentes Principais uma metodologia mais complexa, espera-se uma boa concordância entre os resultados de ambas, devendo uma ser utilizada como complemento da outra (FERRAUDO, 2010).

2.8 Distância Euclidiana

As medidas de dissimilaridade são de grande importância em estudos de diversidade genética em que se procura identificar genitores a serem utilizados em programas de hibridação. Em outras situações, estudos sobre a diversidade genética têm sido realizados com o intuito de identificar grupos de cultivares com maior similaridade, visando à formação de multilinhas (CRUZ e CARNEIRO, 2006).

A similaridade entre objetos é uma medida de correspondência ou semelhança entre os objetos a serem agrupados. As medidas de distância representam as similaridades como sendo a proximidade entre as observações ao longo dos atributos. Na realidade, essas medidas são de dissimilaridade na qual os valores maiores denotam uma menor similaridade. A distância é convertida em similaridade usando-se uma relação inversa ($1 - \text{distância}$) (SAAD, 2009).

Ainda segundo Saad (2009), dado n indivíduos, onde cada um deles possui valores para p variáveis, a distância euclidiana entre eles é determinada pelo teorema de Pitágoras, num espaço multidimensional.

Segundo Manly (1986) quando a distância euclidiana é determinada a partir das variáveis originais, pode ser influenciada pela escala de medida, pelo número de variáveis e pela correlação existente entre as mesmas. Assim, é feita a padronização das variáveis, para que possuam variância igual a um (1).

Note-se que a distância Euclidiana e a distância Euclidiana quadrática normalmente computam dados brutos e não dados normalizados, o que representa

certas vantagens. Por exemplo, a distância entre dois objetos quaisquer não é afetada pela adição de novos objetos, ainda que sejam dados fora do padrão. Contudo, as distâncias podem ser grandemente afetadas por diferenças de escala entre as dimensões dos objetos analisados (ANDRADE, 2004).

Embora os algoritmos possam definir automaticamente o número de grupos, o pesquisador pode também ele impor um número limite de *clusters*. O número de clusters encontrado está intimamente ligado com o grau de similaridade que é exigido a um algoritmo. Se, por exemplo, a um determinado algoritmo lhe for exigida uma fraca similaridade entre os objetos de um cluster, presumivelmente este algoritmo encontrará menos clusters do que outro algoritmo ao qual lhe seja exigido uma forte similaridade entre os registros pertencentes a um mesmo grupo (RAMOS, 2006).

2.9 Análise de Agrupamento pelo Método de Ward

O método de Ward é classificado como um método hierárquico, onde os genótipos são agrupados por um processo que se repete em vários níveis, até que seja estabelecido o dendrograma. As delimitações podem ser estabelecidas por um exame visual do dendrograma, em que se avaliam pontos de alta mudança de nível, tornando-os em geral como delimitadores do número de genótipos para determinado grupo (CRUZ e CARNEIRO, 2006).

Como relata Cruz e Carneiro (2006), neste método consideram-se, para a formação inicial do grupo, aqueles indivíduos que proporcionam a menor soma de quadrados dos desvios. Admite-se que, em qualquer estágio, há perda de informação em razão do agrupamento realizado. A soma de quadrados dos desvios dentro é calculada considerando apenas os acessos dentro do grupo em formação, e a soma de quadrados dos desvios total é calculada considerando todos os indivíduos disponíveis para a análise de agrupamento.

3. MATERIAL E MÉTODOS

Para a obtenção das populações segregantes, foram realizados cruzamentos biparentais sendo utilizados 13 genitores (linhagens) convencionais pertencentes ao programa de melhoramento genético da Universidade Estadual Paulista – UNESP/FCAV Câmpus de Jaboticabal-SP e 12 genitores (cultivares comerciais) portadores do gene RR.

A Tabela 1 apresenta a listagem dos genitores convencionais pertencentes a UNESP/FCAV utilizados neste trabalho e suas respectivas genealogias.

Tabela 1. Linhagens convencionais pertencentes a UNESP/FCAV e respectivas genealogias de cruzamento de origem

Linhagens convencionais UNESP/FCAV	Genealogia
JAB.01-21/4M1D	Tracy – M x Paraná
JAB.02-26/1K1B	FT-Cometa x Paraná
JAB.02-3/6A4D	FT-Cometa x Paraná
JAB.02-30/1G4A	FT-Cometa x Paraná
JAB.03-3/1C3D	FT-Cometa x Bossier
JAB.03-3/1H1D	FT-Cometa x Bossier
JAB.03-3/1H2D	FT-Cometa x Bossier
JAB.04-1/5A4D	BR-16 x Paraná
JAB.05-1/5C3B	FT-Cometa x IAC-8
JAB.05-5/4A2D	FT-Cometa x IAC-8
JAB.05-8/2D3C	FT-Cometa x IAC-8
JAB.06-2/2C1D	BR-16 x Ocepar-4
JAB.06-2/3I3D	BR-16 x Ocepar-4

A relação dos cruzamentos e genitores utilizados para formação das populações segregantes do presente estudo, encontra-se na Tabela 2.

Tabela 2. Genealogias de cruzamentos realizados entre genitores convencionais e portadores do gene de resistência ao glifosato (RR), com os respectivos números atribuídos a cada cruzamento. Jaboticabal, SP.

Número do Cruzamento	Genitor Convencional UNESP/FCAV		Genitor RR Monsoy
C1	JAB.04-1/5A4D	x	M 7639 RR
C2	JAB.04-1/5A4D	x	M 8230 RR
C3	JAB.04-1/5A4D	x	M 8360 RR
C4	JAB.04-1/5A4D	x	M 8921 RR
C5	JAB.03-3/1C3D	x	M 7211 RR
C6	JAB.03-3/1C3D	x	M 7578 RR
C7	JAB.03-3/1H1D	x	M 7578 RR
C8	JAB.03-3/1H1D	x	M 7908 RR
C9	JAB.03-3/1H1D	x	M 8336 RR
C10	JAB.03-3/1H2D	x	M 7211 RR
C11	JAB.03-3/1H2D	x	M 7639 RR
C12	JAB.03-3/1H2D	x	M 7908 RR
C13	JAB.03-3/1H2D	x	M 8336 RR
C14	JAB.03-3/1H2D	x	M 8360 RR
C15	JAB.01-21/4M1D	x	M 7639 RR
C16	JAB.01-21/4M1D	x	M 8230 RR
C17	JAB.01-21/4M1D	x	M 8336 RR
C18	JAB.05-5/4A2D	x	M 7578 RR
C19	JAB.05-5/4A2D	x	M 7908 RR
C20	JAB.05-5/4A2D	x	M 8221 RR
C21	JAB.05-1/5C3B	x	M 7211 RR
C22	JAB.05-1/5C3B	x	M 7639 RR
C23	JAB.05-1/5C3B	x	M 8211 RR
C24	JAB.05-1/5C3B	x	M 8336 RR
C25	JAB.05-1/5C3B	x	M 8360 RR
C26	JAB.05-8/2D3C	x	M 7211 RR
C27	JAB.05-8/2D3C	x	M 7636 RR
C28	JAB.05-8/2D3C	x	M 7908 RR
C29	JAB.05-8/2D3C	x	M 8336 RR
C30	JAB.06-2/3I3D	x	M 8230 RR
C31	JAB.06-2/3I3D	x	M 8236 RR
C32	JAB.06-2/2C1D	x	M 7211 RR
C33	JAB.02-26/1K1B	x	M 7211 RR
C34	JAB.02-30/1G4A	x	M 8360 RR
C35	JAB.02-3/6A4D	x	M 7211 RR
C36	JAB.02-3/6A4D	x	M 8360 RR
C37	JAB.02-3/6A4D-1	x	M 8360 RR

Os genitores portadores do gene RR foram utilizados como doadores de pólen (parental masculino) e os genitores convencionais como receptores de pólen (parental feminino), objetivando a introgressão do gene RR nas populações F_1 e subsequentes gerações segregantes.

Após realizados os cruzamentos, foram obtidas 37 populações F_1 que foram semeadas em vasos para obtenção das gerações F_2 .

3.1 Geração F_2

As sementes F_2 de cada cruzamento foram colhidas separadamente e semeadas na área experimental da Fazenda de Ensino, Pesquisa e Extensão - FEPE da Faculdade de Ciências Agrárias e Veterinárias/UNESP – Campus de Jaboticabal, no ano agrícola 2010/11. A semeadura ocorreu em 37 linhas individuais de 5,0 metros de comprimento e espaçamento de 0,50 metro entre linhas.

Dentro de cada linha no campo, que consistia de um diferente cruzamento, foram selecionadas oito plantas F_2 , visualmente superiores, obtendo-se um grupo de progênies que deram origem às progênies F_3 .

Estas avaliações consistiram na atribuição de notas e medidas do fenótipo de cada planta individual, sendo avaliadas as características agronômicas, a saber;

Número de dias para o florescimento (NDF): avaliado no estágio R2 de desenvolvimento da planta compreendendo o período entre a data da semeadura e a data em que as plantas encontravam-se em pleno florescimento, expresso em dias;

Número de dias para a maturidade (NDM): avaliado no estágio R8 de desenvolvimento da planta compreendendo o período entre a data da semeadura e a data em que 50% das plantas da área útil se encontravam com 95% das vagens maduras, expresso em dias;

Altura de inserção da primeira vagem (AIV): caráter avaliado no estágio R8 de desenvolvimento da planta compreendendo a distância desde a superfície do solo até a inserção da primeira vagem, expressa em centímetros;

Altura da planta na maturidade (APM): caráter avaliado no estágio R8 de desenvolvimento da planta compreendendo a distância da haste principal entre o colo e a inserção da vagem mais distal, expressa em centímetros;

Acamamento (Ac): caráter avaliado no estágio R8 de desenvolvimento da planta através de uma escala de notas visuais, variando de 1 (planta ereta) a 5 (planta acamada);

Número de ramos por planta (NR): caráter avaliado no estágio R8 de desenvolvimento da planta considerando-se o número total de ramos unidos à haste principal da planta;

Número de vagens por planta (NV): caráter avaliado no estágio R8 de desenvolvimento da planta considerando o número total de vagens por planta com sementes formadas;

Valor agronômico (VA): caráter avaliado no estágio R8 de desenvolvimento da planta através de uma escala de notas visuais, a qual varia de 1 (plantas com características agronômicas ruins) a 5 (plantas com ótimas características agronômicas), sendo a nota atribuída representativa de um conjunto de caracteres visuais adaptativos: arquitetura da planta, quantidade de vagens cheias, vigor e sanidade da planta, debulha prematura das vagens e retenção foliar na maturidade;

Peso de cem sementes (PCS): Obtido pela relação entre peso total de sementes e número total de sementes por planta, multiplicado por 100;

Produção de grãos (PG): caráter obtido através do peso dos grãos por planta (até a umidade de 13%). Sendo expressa em gramas por planta (g/planta).

3.2 Geração F₃

As progênes da geração F₃, anteriormente selecionadas, foram semeadas novamente em campo na área da FEPE/UNESP – Campus de Jaboticabal, no agrícola 2011/12 utilizando o delineamento experimental do tipo famílias com testemunhas intercalares, totalizando 296 genótipos (originadas da seleção das oito plantas de cada uma das 37 populações F₂). Além das populações segregantes semeadas em linhas individuais de 5,0 metros de comprimento por 0,5 metro entre linhas, foram utilizadas como testemunhas as linhagens JAB.06-2/2C1D e JAB.01-21/4M1D, ambas desenvolvidas e pertencentes ao programa de melhoramento de soja da UNESP/FCAV e utilizadas como genitores nos cruzamentos, além das cultivares Conquista, V-Max, CD-207, CD-216 e CD-219.

O uso do delineamento experimental do tipo famílias com testemunhas intercalares permite a estimação das variações ambientais na avaliação de famílias com número reduzido de sementes, principalmente nas gerações iniciais, onde cada uma das famílias é representada uma única vez, enquanto se repetem apenas as testemunhas em cada bloco.

No momento da colheita da geração F₂, as condições climáticas não foram favoráveis devido ao alto índice pluviométrico ocorrido nesta fase. Assim, as sementes da geração F₃ foram prejudicadas resultando em um baixo índice de germinação e vigor, sendo que, muitas plantas de algumas populações selecionadas foram perdidas em sua totalidade ou tiveram um baixo índice de germinação.

Outro problema enfrentado esteve relacionado à germinação de 22 dos 24 genitores envolvidos nos cruzamentos, justificando assim, a presença de apenas dois genitores convencionais na avaliação final. Além disso, a germinação das populações segregantes (F₃) no ano agrícola 2011/12 foi prejudicada pelo déficit

hídrico. Estas condições ambientais adversas descritas, somadas, resultaram em perdas de aproximadamente 23% da totalidade das populações segregantes.

A semeadura e o manejo cultural em campo foram realizados seguindo as orientações técnicas apresentadas pela EMBRAPA (2011).

Na medida em que os genótipos segregantes atingiram os estádios fenológicos da soja (FEHR; CAVINESS, 1977) ideais para as respectivas avaliações, iniciou-se uma nova avaliação e seleção de progênies individuais por meio de caracteres agrônômicos, em seis plantas por família de genótipo.

Após tomados os dados, procedeu-se as análises estatísticas exploratórias multivariadas com auxílio do software computacional STATISTICA versão 7. (STATSOFT, 2004).

Em relação à Análise de Componentes Principais – ACP, segundo Cruz e Carneiro (2006), para realização desta, geralmente feita com dados padronizados, considera-se que x_{ij} é a média padronizada do j-ésimo caráter ($j = 1, 2, \dots, v$) avaliado no i-ésimo genótipo ($i = 1, 2, \dots, g$) e R a matriz de covariâncias ou de correlação entre esses caracteres. A técnica de componentes principais consiste em transformar o conjunto de v variáveis ($x_{i1}, x_{i2}, \dots, x_{iv}$) em um novo conjunto ($y_{i1}, y_{i2}, \dots, y_{iv}$) que são funções lineares dos x_j 's e independentes entre si.

Para a seleção dos componentes seguiu-se como o proposto por Kaiser (1958), também citado por Hair et al., (2005) onde somente serão considerados autovalores acima de um (1,0), pois geram componentes com quantidade relevante de informação das variáveis originais. Abaixo de um a quantidade de informação retida no componente não é relevante.

A Distância Euclidiana é provavelmente a medida mais comumente utilizada para analisar a semelhança entre objetos. É simplesmente a distância geométrica dos objetos x e y em um espaço multidimensional, dada pela equação:

$$d(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Distância Euclidiana quadrática é uma forma de empregar o padrão de distância

Euclidiana dando, progressivamente, maior peso aos objetos mais distantes. É calculada pela equação:

$$d(x, y) = \sum_i (x_i - y_i)^2$$

A Análise de Agrupamentos pelo método K-means consiste de um procedimento onde, dado um número de *clusters* previamente determinado, calcula-se pontos que representam os “centros” destes *clusters*.

O critério de agrupamento pelo método de K-means pode ser descrito como sendo:

$$E = \sum_{k=1}^k \sum_{x_1 \in C_k} d(x_i, x_{0k})$$

onde: x_{0k} é o centróide do cluster C_k e $d(x_1, x_{0k})$ é a distância entre os pontos x_i e x_{0k} . O centróide pode ser a média ou a mediana de um grupo de pontos. Em outras palavras, o objetivo do K-means é minimizar a distância entre cada ponto e o seu respectivo centróide (HAIR et al., 2005).

Para o método de Ward, procedimento também chamado de “mínima variância”, fundamenta-se nos seguintes princípios:

Inicialmente cada elemento é considerado como um único conglomerado. Em cada passo do algoritmo de agrupamento calcula-se a soma de quadrados dentro de cada aglomerado. Esta soma é o quadrado da distância Euclidiana de cada elemento amostral pertencente ao conglomerado em relação ao correspondente vetor de médias do aglomerado, isto é:

$$SS_i = \sum_{k=1}^{n_i} (X_{ik} - \bar{X}_i)'(X_{ik} - \bar{X}_i)$$

onde n_i é o número de elementos no aglomerado A_i quando se está no passo s do processo de agrupamentos, X_{ik} é o vetor de observações do k – ésimo elemento amostral que pertence ao i – ésimo conglomerado, \bar{X}_i é o centróide do conglomerado

A_i , e SS_i representa a soma de quadrados correspondente ao aglomerado A_i . No passo s , a soma de quadrados total dentro dos grupos, SSR , é definido como:

$$SSR = \sum_{i=1}^{g_s} SS_i$$

onde g_s é o número de grupos existentes quando se está no passo s .

A distância entre os aglomerados A_i e B_l é , então, definida como:

$$d(A_i, B_l) = \left[\frac{n_i n_l}{n_i + n_l} \right] (\bar{X}_i - \bar{X}_l)' (\bar{X}_i - \bar{X}_l)$$

que é a soma de quadrados entre os *clusters* A_i e B_l . Em cada passo do algoritmo de agrupamento, os dois conglomerados que minimizam a distância, são combinados.

É possível demonstrar que a medida de distância, é a diferença entre o valor de SSR depois e antes de se combinar os conglomerados A_i e B_l num único conglomerado. Portanto, em cada passo do agrupamento, o método de Ward combina os dois conglomerados que resultam no menor valor SSR . Este processo gera um dendrograma com diferentes grupos.

Como a análise multivariada, especificamente as análises de Componentes Principais e Agrupamento usam dados quantitativos contínuos, as variáveis valor agrônômico (VA) e acamamento (Ac), a priori, não foram utilizadas nas análises pois são variáveis oriundas de notas atribuídas à escalas pré-definidas classificadas como variáveis qualitativas, sendo utilizadas em análises e comparativos posteriores. Assim, posteriormente e juntamente, as progênies selecionadas resultantes da ACP foram analisadas e comparadas uma a uma quanto aos dados atribuídos individualmente pela avaliação em campo quanto ao VA e Ac e descartadas aquelas que apresentavam inconsistência ou não condiziam com o propósito da seleção, visto que, por se tratar de análises que consideram todas as variáveis simultaneamente, nem sempre uma progênie que apresente uma variável

indesejável é descartada ou não selecionada, pois esta é analisada e selecionada pelo seu conjunto de dados de suas variáveis em análise e não de modo univariado.

Os dados foram então padronizados para as análises de acordo com a recomendação (STATSOFT, 2004), pois como descreve Ferraudo (2010), há duas razões para a padronização da matriz de dados. Primeiro, as unidades associadas aos atributos podem arbitrariamente influenciar o grau de semelhança entre os objetos. Com a padronização o efeito dessa arbitrariedade é eliminado. Segundo, a padronização faz com que os atributos contribuam com o mesmo peso no cálculo do coeficiente de semelhança entre objetos. Se uma variável possui um intervalo de valores superior a um outro de uma outra variável, certamente a primeira variável contará com um peso maior na determinação do grau de semelhança entre objetos. Este efeito pode ser compensado pela padronização.

Muitas são as funções de padronização e uma das mais utilizadas e adotada pelo software STATISTICA é:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

onde $j=1, 2, \dots, p$ atributos e $i=1, 2, \dots, n$ objetos, Z_{ij} o valor padronizado de X_{ij} e \bar{X}_j e S_j a média e o desvio padrão da coluna j . As variáveis assim padronizadas tem média nula e variância unitária.

Para selecionar as progênies na análise de Componentes Principais, foi utilizada uma metodologia onde pré-determinou-se escalas no plano do gráfico para que fosse possível identificar, mais facilmente, as progênies que possuíam características que às diferenciavam das demais dentro de cada cruzamento. Assim, a primeira escala, menos rigorosa, compreende em seu eixo X valores de 5 a -5 e no eixo Y valores de 2 a -2. Na segunda escala, sendo esta mais rigorosa, determinou-se valores no eixo X compreendendo de 6 a -6 e no eixo Y com valores de 2,5 a -2,5. Tais escalas foram pré-definidas esperando-se alcançar o objetivo de realizar uma seleção de até dez por cento do total da população em avaliação.

Este procedimento auxilia na identificação e seleção das progênies diferenciais, uma vez que, todos os indivíduos que alocarem-se próximas ao centro

não possuem diferenças importantes entre elas, ou seja, possuem certa homogeneidade e seguem um padrão de características dentro do cruzamento, não sendo interessantes para a seleção. Para os indivíduos que alocarem-se entre a escala de menor rigor e a escala de maior rigor, pode-se dizer que estes possuem certa diferenciação dos demais com determinadas características superiores, dependendo do seu posicionamento. Caso o pesquisador julgue que este rigor de seleção seja adequado, selecionando indivíduos que embora sejam diferentes entre si, esta margem é pequena e seguem um padrão dentro do cruzamento mesmo que em menor escala, estes podem ser selecionados.

Por fim, para os indivíduos que posicionarem-se externamente a escala de maior rigor, pode-se dizer que estes possuem uma diferença marcante, ou seja, são progênies que possuem uma característica, com uma ou mais variáveis de alta expressão que as faz diferirem-se dentro dos demais membros do cruzamento. Estes indivíduos, neste caso progênies, são as que devem receber atenção especial e são passíveis de serem selecionadas.

Essas análises foram primeiramente realizadas separadamente para cada grupo de progênies (famílias) de cada cruzamento, sendo possível identificar quais plantas selecionadas a campo dentro de cada cruzamento apresentaram características (variáveis) superiores e que condiziam com a seleção realizada pela análise de Componentes Principais. Esse processo foi repetido em todos os cruzamentos e também para as cultivares testemunhas utilizadas para o comparativo, ou seja, uma análise e seleção dentro de famílias. Cada cruzamento possuía diferentes números de progênies em suas famílias.

4. RESULTADOS E DISCUSSÃO

A seleção fenotípica na geração F_3 de seis plantas de cada uma das 296 famílias de genótipos, no ano agrícola 2011/2012, baseado nos caracteres agronômicos (variáveis) já citados, considerando que houve perdas pela baixa germinação parcial ou total de algumas destas, resultou em 1382 plantas selecionadas, além de 110 plantas avaliadas oriundas das cultivares testemunhas utilizadas para o estudo. No conjunto de plantas selecionadas de cada cruzamento, aplicou-se primeiramente a técnica da análise de Componentes Principais pela metodologia já descrita.

É possível através do programa estatístico, visualizar o número de fatores e a representação de cada um no contexto da análise, como neste estudo se utiliza de nove variáveis, automaticamente irá conter nove fatores, resultando em uma qualidade representativa de 100% da amostra e seus respectivos autovalores.

Devido ao grande número de informações de cada um dos 37 cruzamentos gerado pela ACP, foi realizada a análise da média de cada um dos cruzamentos para se obter uma estimativa do comportamento dos mesmos. Tais médias são apresentados na Tabela 3.

Primeiramente, são analisados os resultados referentes aos autovalores da matriz de correlação para a média dos cruzamentos, disponível na Tabela 4. Estes mesmos dados são apresentados graficamente na Figura 1. Verifica-se que apenas dois autovalores devem ser considerados, pois possuem valores superiores a um (1,0). O autovalor para o primeiro fator foi 4,6482 o que é um valor expressivo, pois a soma de todos os autovalores é 9,0. A proporção de variância retida pelo primeiro fator é de 51,64%. O autovalor do segundo fator foi 1,9870. A proporção de variância retida pelo segundo fator é de 22,08%. Juntos estes dois fatores são responsáveis por 73,72% da variância total dos resultados.

Tabela 3. Médias originais das progênes selecionadas de cada um dos 37 cruzamentos na geração F3, baseado nas variáveis: número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da inserção da primeira vagem (AIV); altura da planta na maturidade (APM); número de ramos (NR); número de vagens (NV); peso de cem sementes (PCS); número de sementes (NS) e produção de grãos (PG).

Nº CRUZAMENTO	NDF	NDM	AIV	APM	NR	NV	PCS	NS	PG (g)
C1	46	117	14,0	91,5	6	139	14,0	291	39,5
C2	47	116	13,3	73,4	6	132	12,6	281	36,1
C3	51	117	15,3	77,5	5	134	12,8	265	32,5
C4	48	117	9,4	59,8	6	220	15,3	411	62,7
C5	45	116	6,2	56,0	9	258	18,5	493	93,3
C6	43	116	7,0	48,6	7	219	19,8	447	90,0
C7	48	110	8,6	57,2	6	160	18,9	349	52,3
C8	49	110	11,9	63,2	4	85	17,9	182	32,4
C9	44	119	11,3	61,4	4	84	16,1	188	30,4
C10	47	114	15,5	100,4	3	79	16,5	163	27,4
C11	47	118	15,0	86,6	5	141	14,4	313	43,8
C12	44	112	13,2	72,4	4	92	14,2	213	31,2
C13	40	111	15,2	81,0	3	77	14,3	237	36,1
C14	53	122	15,9	94,4	4	94	12,6	187	23,5
C15	43	107	14,8	88,9	4	91	16,5	191	31,1
C16	41	118	14,4	93,1	5	102	16,9	218	37,0
C17	39	111	16,4	109,7	5	86	17,3	174	30,7
C18	43	109	15,2	113,3	4	88	13,4	193	25,3
C19	47	115	11,6	77,9	4	95	15,2	199	30,5
C20	50	119	13,3	107,5	4	92	13,0	199	24,9
C21	44	114	10,6	109,8	4	86	15,0	148	24,9
C22	48	118	13,7	106,4	4	100	13,4	211	27,4
C23	52	120	10,4	97,2	4	97	13,6	177	23,6
C24	54	121	10,5	123,8	3	83	11,5	165	18,9
C25	53	120	11,2	107,5	3	85	12,5	149	18,3
C26	42	113	14,9	101,8	3	59	18,0	121	18,4
C27	45	112	22,9	108,9	4	94	21,0	168	20,3
C28	45	108	15,9	84,2	3	100	15,5	187	28,0
C29	42	112	19,9	89,1	4	89	12,7	190	23,8
C30	48	118	11,8	80,9	6	181	13,9	330	47,5
C31	50	122	10,9	82,8	4	125	14,0	250	36,0

Continua...

Continuação Tabela 3

Nº CRUZAMENTO	NDF	NDM	AIV	APM	NR	NV	PCS	NS	PG (g)
C32	43	116	12,2	112,0	4	109	17,9	173	30,1
C33	49	121	12,3	93,9	4	95	14,4	188	27,5
C34	40	114	13,2	102,3	5	105	17,2	320	54,7
C35	39	108	8,2	47,7	4	69	16,5	160	26,5
C36	49	120	17,6	114,6	3	75	12,5	161	20,6
C37	50	119	18,1	115,0	3	72	11,9	159	20,1
JAB.06-2/2C1D	45	110	10,2	62,7	6	135	17,2	313	54,2
JAB.01-21/4M1D	40	119	12,3	97,8	6	135	13,1	304	39,6
CONQUISTA	57	138	22,5	92,9	5	99	15,7	182	29,0
V-MAX	49	124	16,0	88,8	5	70	16,0	121	19,1
CD-219	57	132	24,2	102,9	3	93	13,9	198	27,6
CD-207	56	122	15,0	74,0	7	115	14,3	305	42,9
CD-216	42	101	13,3	72,0	5	72	16,8	177	29,0

Tabela 4. Autovalores da matriz de correlação e variâncias totais (%) e acumuladas para cada número de fatores a serem determinados na análise da média dos cruzamentos

Fatores	Autovalor	% Total de variância	Autovalor acumulado	% Acumulativa
1	4,648218	51,64687	4,648218	51,6469
2	1,987092	22,07880	6,635310	73,7257
3	0,844475	9,38305	7,479785	83,1087
4	0,572372	6,35969	8,052157	89,4684
5	0,422083	4,68981	8,474239	94,1582
6	0,248876	2,76529	8,723116	96,9235
7	0,184898	2,05442	8,908014	98,9779
8	0,074624	0,82915	8,982638	99,8071
9	0,017362	0,19292	9,000000	100,0000

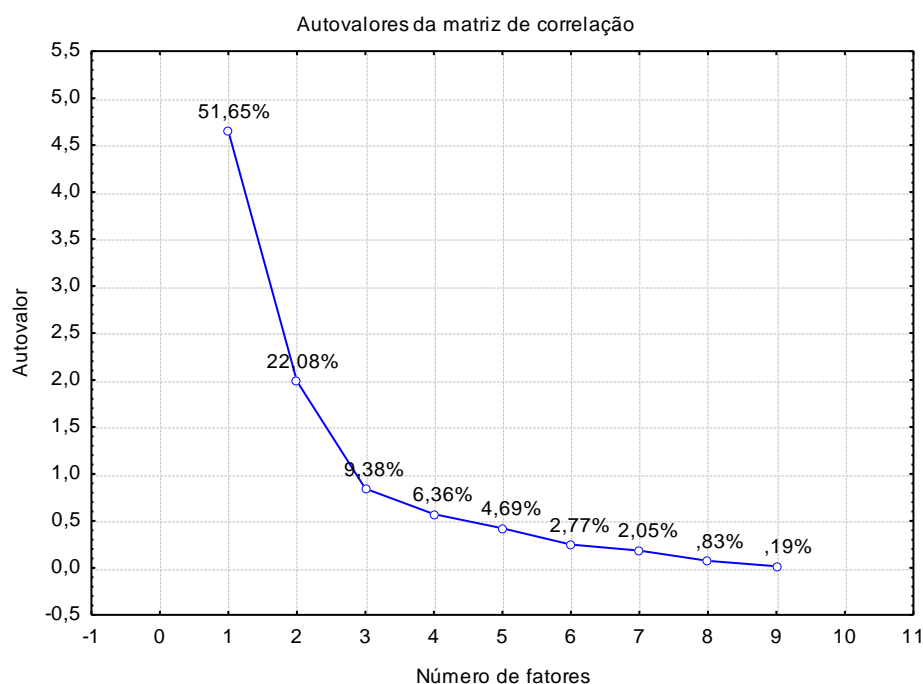


Figura 1. *Scree-plot* dos autovalores da matriz de correlação de acordo com o número de fatores a serem determinados na análise da média dos cruzamentos com a proporção de variância retida em cada autovalor.

Após definido o número de fatores a serem analisados, foi realizada a ACP para o conjunto de médias de cada cruzamento, representado na Figura 2, onde é possível verificar a dispersão destes de acordo com as variáveis. Verifica-se que os cruzamentos C5 e C6 apresentam uma forte influência das variáveis PG, NR, NS e NV, variáveis estas pertencentes aos componentes do grão que discriminam este grupo de plantas. Porém, estes cruzamentos não foram selecionados posteriormente pela ACP (Tabela 7). Este fato é atribuído ao baixo número de progênies nestes cruzamentos, sendo quatro e cinco progênies respectivamente. Para o C3 o número de progênies é maior permitindo a exploração da variabilidade gerada pela divergência genética entre os genitores. O que pode também estar ocorrendo para este cruzamento em específico, e que pode ser considerado, é a chamada segregação transgressiva, ou seja, o surgimento nas gerações iniciais de indivíduos

(progênies) com valores fora dos limites de variação dos genitores, principalmente por se tratar de variáveis quantitativas.

As variáveis APM e AIV se mostraram, de maneira geral, altamente relacionadas, sendo que, os cruzamentos C14, C24, C25, C36 e C37 apresentaram-se muito influenciados por estas variáveis.

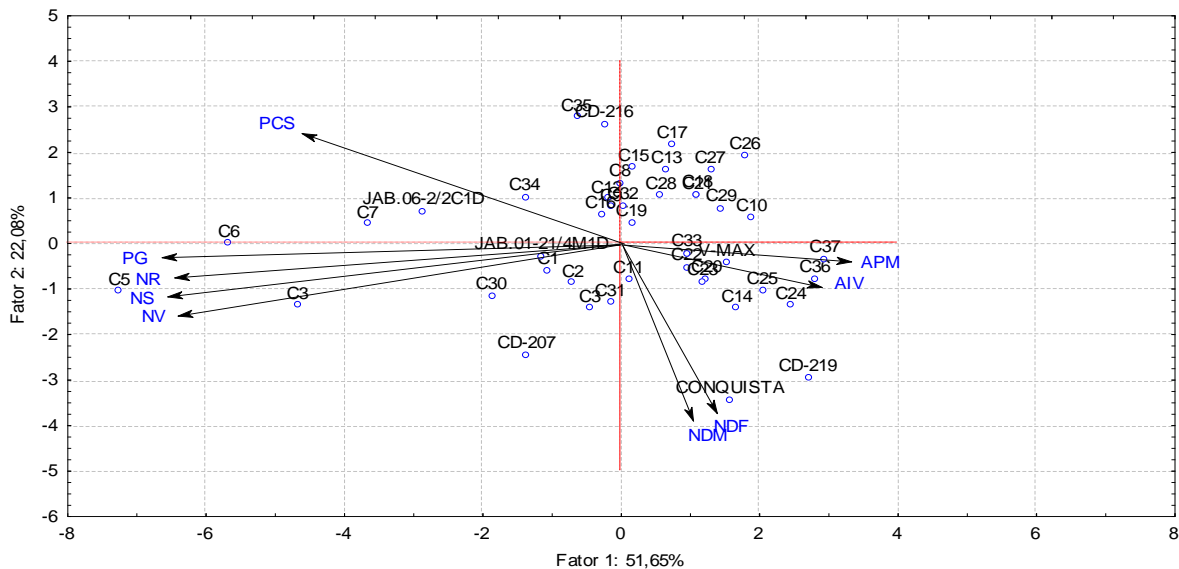


Figura 2. Gráfico biplot com dispersão dos 37 cruzamentos e testemunhas e projeção dos vetores das variáveis: número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da planta na maturidade (APM); altura da inserção da primeira vagem (AIV); número de ramos (NR); número de vagens (NV); número de sementes (NS); peso de cem sementes (PCS); produção de grãos por planta (PG).

As variáveis NDF e NDM mostraram-se estritamente relacionadas, ou seja, uma é dependente da outra. Este fato foi observado principalmente para as testemunhas Conquista e CD-219. Cabe lembrar que nem sempre isto é observado em campo para demais cruzamentos e cultivares, visto que, esta correlação também está ligada ao tipo de crescimento da soja (TAKAHASHI; FUKUYAMA, 1919; ETHERIDGE et al., 1929).

Os cruzamentos C7, C34, C35 e as testemunhas JAB.06-2/2C1D e CD-216 caracterizaram-se por apresentar maiores valores para a variável PCS.

Os demais cruzamentos permaneceram centralizados no plano do gráfico, significando que estes possuem valores médios para cada variável, não apresentando uma característica (variável) específica de alta expressão.

Para efeito explicativo, foram utilizados os dados do cruzamento C11 para o desdobramento da análise por apresentar dados que permitem explorar detalhadamente os resultados.

Assim, analisando primeiramente os resultados para os autovalores na matriz de correlação, número de fatores, autovalores, porcentagens da variância total, autovalor acumulado e representatividade respectivamente, apresentados na Tabela 5 e representados graficamente na Figura 3, é possível verificar que apenas três autovalores apresentaram valores superiores a um, ou seja, apenas estes devem ser considerados significativos. Assim, o próximo passo é definir a correlação entre os fatores significativos a serem utilizados. Esta correlação é feita sempre entre dois fatores, como neste caso, Fator 1 x Fator 2, Fator 1 x Fator 3 e Fator 2 x Fator 3.

Tabela 5. Autovalores da matriz de correlação e variâncias totais e acumuladas para cada número de fatores a ser determinado na análise do cruzamento C11.

Fatores	Autovalor	% Total de variância	Autovalor acumulado	% Acumulativa
1	4,062379	45,13754	4,062379	45,1375
2	1,578971	17,54412	5,641350	62,6817
3	1,067682	11,86314	6,709032	74,5448
4	0,900533	10,00592	7,609565	84,5507
5	0,615256	6,83618	8,224821	91,3869
6	0,462947	5,14385	8,687768	96,5308
7	0,227323	2,52581	8,915091	99,0566
8	0,079204	0,88005	8,994295	99,9366
9	0,005705	0,06339	9,000000	100,0000

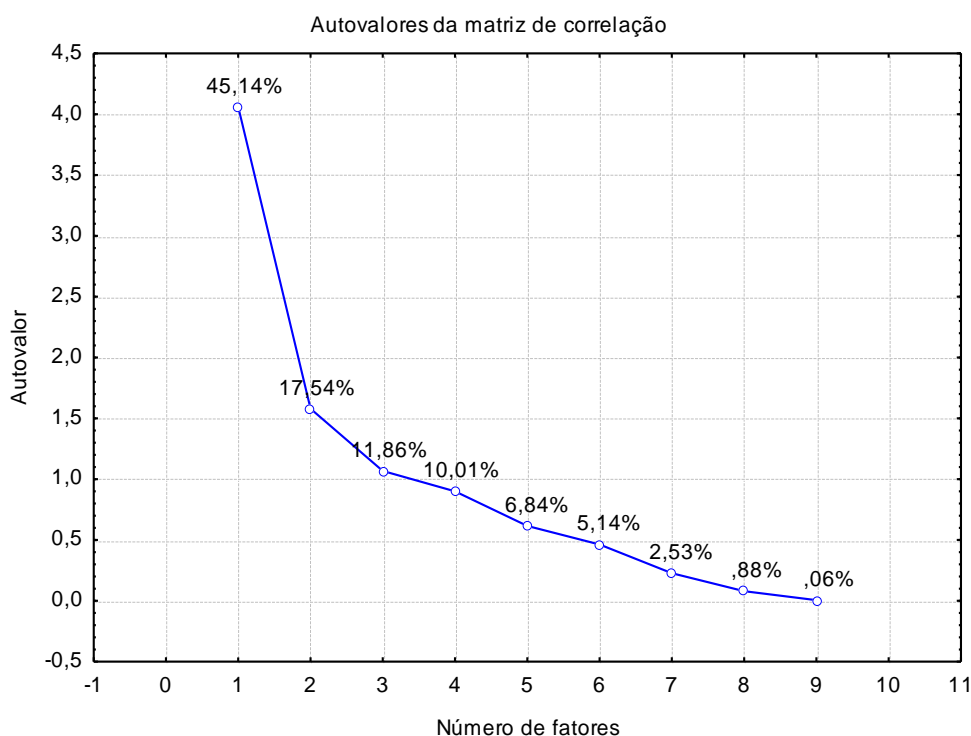


Figura 3. Scree-plot dos autovalores da matriz de correlação de acordo com o número de fatores a ser determinado para análise do cruzamento C11

Como se observa, o autovalor para o primeiro fator foi 4,062379 o que consiste em um valor expressivo, pois a soma de todos os autovalores é 9,0. A proporção de variância retida pelo primeiro fator é de 45,14%. Conforme observado na Tabela 6, as principais variáveis que explicaram essa retenção da variância foram: NR, NV, NS e PG com os valores de correlações com os componentes principais de -0,835181, -0,939138, -0,967741 e -0,954412 respectivamente. Foram consideradas importantes apenas as correlações acima de 0,6 (FERRAUDO, 2010).

O autovalor para o segundo fator foi 1,578971. A proporção de variância retida pelo segundo fator é de 17,54%. Os terceiro autovalor possui um valor muito próximo de um, sendo este passível de utilização para análise, porém, neste caso, será usada apenas a correlação entre os dois primeiros fatores por possuírem as maiores cargas.

O plano bidimensional formado pelos componentes principais do Fator 1 (45,13%) e componentes principais do Fator 2 (17,54%) reteve no total 62,68% da

variância original (Figura 4), sendo este o maior valor possível de se alcançar na correlação entre fatores significativos e ficaram caracterizados pelas variáveis NR, NV, NS, PG, APM e PCS.

Os demais autovalores possuem valores inexpressivos para serem usados neste estudo.

Tabela 6. Correlação entre as variáveis e os componentes principais (autovalores) das progênies de soja na geração F3 para o cruzamento C11. Número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da planta na maturidade (APM); altura da inserção da primeira vagem (AIV); número de ramos (NR); número de vagens (NV); número de sementes (NS); peso de cem sementes (PCS); produção de grãos (PG).

Variáveis	Fator 1	Fator 2	Fator 3
NDF	0,374743	-0,253412	-0,722188
NDM	0,529053	0,487764	0,317705
AIV	0,449863	0,299426	0,433096
APM	0,092804	0,815328	-0,183109
NR	-0,835181	0,001936	0,052274
NV	-0,939138	0,086353	0,028856
PCS	0,064254	0,667876	-0,463976
NS	-0,967741	0,109490	0,035934
PG	-0,954412	0,238572	-0,062924

Após realizado esse processo, a continuidade se dá pela análise e representação gráfica da correlação entre o Fator 1 x Fator 2, lembrando que, a escolha de quais correlações entre fatores que deve ser usada depende também da definição feita pelo pesquisador de acordo com seus critérios e objetivos.

Como descreve Hair et al., (2005), para cada fator, sinais iguais significam que as variáveis estão positivamente relacionadas e sinais opostos significam que as variáveis estão negativamente relacionadas

O resultado da ACP para o cruzamento C11, utilizado como exemplo, demonstra que NS, NV e NR estão positiva e fortemente relacionadas com PG, fato já esperado, pois tratam-se de variáveis componentes de produção. Assim, quanto mais alto o valor destas, mais alto é o valor de PG da progênie, exercendo assim forte influência de uma sobre as outras.

Para AIV e APM, estas também estão relacionadas, posicionando-se no mesmo quadrante, podendo ser observado que, geralmente, quanto maior a AIV, maior a APM.

Para as variáveis NDF e NDM, os resultados apresentaram uma relação negativa, ou seja, nem sempre uma progênie com menor tempo para iniciar o florescimento, também terá menor tempo para iniciar sua maturidade fisiológica, o que determina seu ciclo ou grupo de maturidade. O tipo de crescimento da soja classificado como determinado, semideterminado ou indeterminado (SEDIYAMA; TEIXEIRA; REIS, 2005) também pode influenciar na relação entre AIV/APM e NDF/NDM. Este pode ser um caso específico para este cruzamento, podendo variar de acordo com os genitores envolvidos na hibridação, local de cultivo e pelas influências climáticas.

Um fato interessante a se avaliar nesta análise trata-se da não correlação direta da variável peso de cem sementes (PCS) aos componentes de produção. Isto pode ser explicado pelo fato de que esta variável, isoladamente, não influencia na PG diretamente, mas sim quando associada às demais, uma vez que, geralmente, quanto maior o NR e NV maior é o NS e menor é o PCS.

O próximo passo é a análise e representação gráfica da dispersão das progênies do plano de fator de coordenadas. Neste ponto, seguem-se os mesmos princípios para análise anteriormente descrita. Assim, as progênies serão distribuídas de acordo com seus valores e a projeção de suas variáveis.

O resultado dessa análise, ainda para o cruzamento C11 é representada na Figura 4. Observa-se a distribuição de cada progênie de acordo com as coordenadas, considerando agora a relação com suas variáveis. Assim, pode se visualizar que um grande grupo de progênies localizaram-se no centro do plano, dentro da delimitação das escalas pré-definidas. Isto demonstra que nesse grupo não há diferenças significativas entre as progênies, ou seja, os valores de suas

variáveis estão dentro de uma conformidade que não as caracterizam como superiores umas das outras, podendo-se dizer que há certa homogeneidade dentro deste grupo, mesmo estas estando em geração F₃.

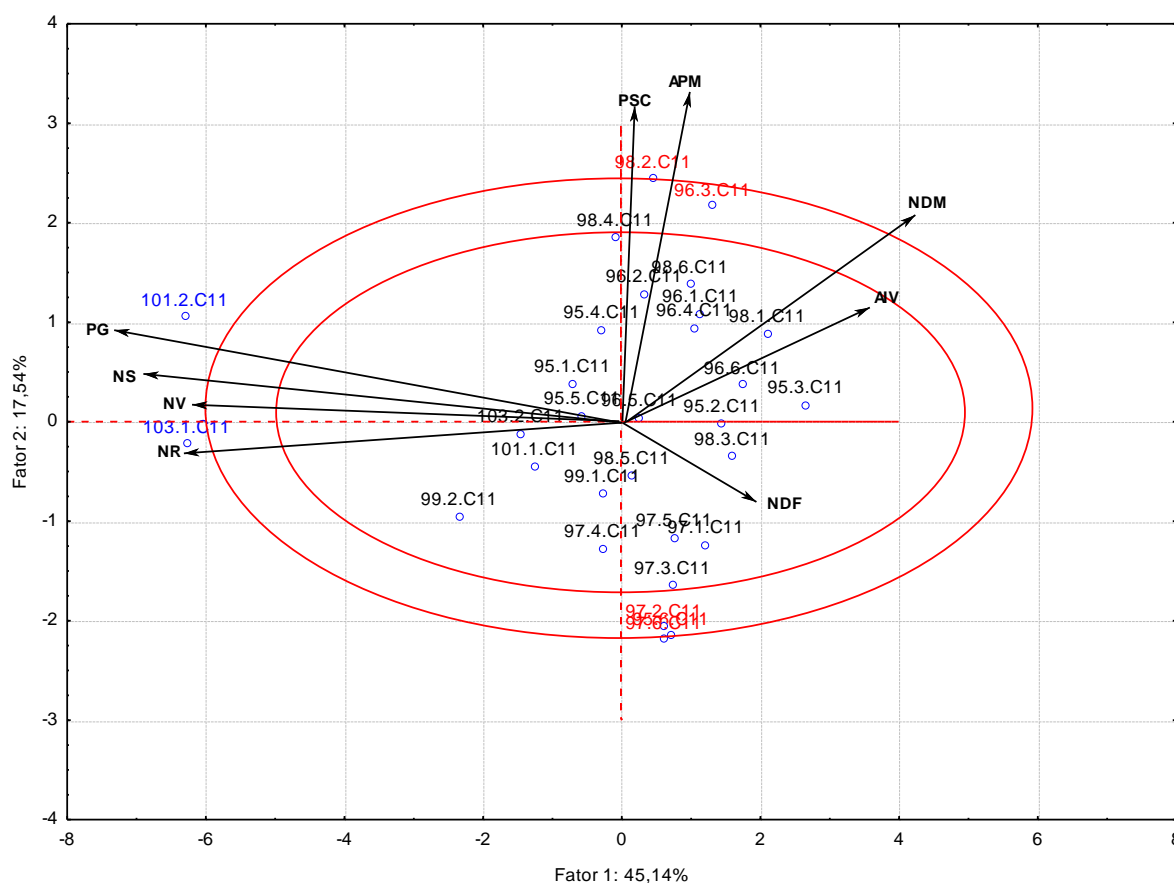


Figura 4. Gráfico biplot com dispersão das 30 progênies de soja do cruzamento C11 em função do Fator 1 x Fator 2 e projeção dos vetores das variáveis: número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da planta na maturidade (APM); altura da inserção da primeira vagem (AIV); número de ramos (NR); número de vagens (NV); número de sementes (NS); peso de cem sementes (PCS); produção de grãos (PG). Genótipos selecionados pela escala menos rigorosa estão em cor vermelha e genótipos selecionados pela escala de maior rigor estão em cor azul.

Nas progênies que se localizaram entre as delimitações das escalas, destacadas em cor vermelha, pode se dizer que estas possuem certa diferença das demais anteriormente descritas, podendo ser selecionadas como superiores, porém, utilizando um critério menos rigoroso.

Para as progênies que se localizaram nas extremidades, pode se concluir que apresentam diferenças significativas das demais, isto porque possuem caracteres específicos, com variáveis de alto valor que as torna superiores as demais. Como já dito anteriormente, esta análise deve ser interpretada em conjunto com o resultado da projeção das variáveis pelo fato de que após identificar quais progênies podem ser consideradas superiores, deve se esclarecer o motivo que as coloca nessa posição.

Sendo assim, quanto mais uma progênie se distanciar do centro do sistema de coordenadas, mais padrão específico ela apresenta. Assim, as progênies 101.2.C11 e 103.1.C11 possuem padrão específico por terem maiores valores de PG, NS, NV e NR; as progênies 98.2.C11 e 96.3.C11 possuem padrão específico por terem maiores valores de APM, PCS, NDM e AIV enquanto que as progênies 97.6.C11, 95.6.C11 e 97.2.C11 possuem baixos valores de APM, PCS, NDM e AIV.

Assim como na maioria dos programas de melhoramento, entre as características buscadas em uma nova cultivar, a produtividade é o fator de maior relevância, não sendo diferente neste caso. Portanto, as progênies 101.2.C11 e 103.1.C11 foram selecionadas para representar o respectivo cruzamento na avaliação final.

Este mesmo processo, exemplificado apenas para o cruzamento C11 é aplicado a todos os demais cruzamentos em estudo, de forma independente, permitindo assim, selecionar apenas as progênies superiores dentro de cada cruzamento e identificar posteriormente, qual destes obteve maior número de representantes na seleção final. Isto é importante para se definir se os genitores utilizados em cada cruzamento foram eficientes em transmitir para a sua descendência, seus caracteres superiores marcantes.

A análise de dados relacionados às variáveis componentes de produção (NR, NV, NS e PG) é de grande importância para a seleção de genótipos produtivos. Costa et al. (2004), Alcantara Neto et al. (2011), Navarro Junior e Costa (2002) e

Vianna (2013) mostraram resultados satisfatórios nos ganhos de seleção para estes caracteres e na contribuição relativa dos mesmos para a produção de grãos em soja.

Ainda na Figura 4, é possível observar que dentro do total de 30 progênies na família do cruzamento C11, 23 delas permaneceram centralizadas, cinco entre as escalas de maior e menor rigor (destacadas na cor vermelha) e apenas duas alocaram-se nas extremidades (destacadas na cor azul). Isso indica que grande parte das progênies estão dentro da média ou padrão do cruzamento, mas que algumas possuem certa superioridade e destacam-se.

A quantidade de progênies selecionadas em cada cruzamento é variável, pois nem todos os genitores se recombinam geneticamente de forma a gerar uma família de progênies com características superiores. Desta forma, nem todos os cruzamentos em estudo contaram com representantes na seleção final.

De fato, podemos citar o exemplo do cruzamento C22 que possui uma família de 48 progênies, das quais, seis foram selecionadas pela análise de Componentes Principais e avaliação quanto a V_a e A_c , ou seja, um índice de seleção em torno de 12,5%, sendo o cruzamento que apresentou maior número de representantes na seleção final. Outros cruzamentos porém, como o caso do C5, C6, C33 e C35 não apresentaram nenhuma progênie que pudesse ser selecionada. Isto pode ser explicado por dois motivos. Primeiro, o número de progênies na família era extremamente pequena, em torno de 4 a 6 progênies, o que limitou explorar uma maior variabilidade que poderia existir resultante da recombinação genética e segregação oriunda do cruzamento (Tabela 7). No entanto, este fato não pode ser considerado para o C33, pois este apresenta um grupo de progênies relativamente grande dentro da família, contando com 35 progênies. O segundo motivo pode estar relacionado ao fato destes cruzamentos realmente não apresentarem bons resultados, não sendo eficiente quanto à recombinação genética entre os genitores utilizados, podendo dizer que se trata de um cruzamento não indicado, tanto pelo fato de não haver eficiência em gerar progênies superiores, bem como nem mesmo gerar um número suficientemente viável de progênies para que houvesse a seleção, sendo que, esta ocorrência pode ser resultado de uma baixa divergência genética entre os genitores utilizados para este cruzamento.

O índice de seleção é variável, dependendo da combinação específica dos genitores. Bertini et al. (2010), obtiveram uma seleção de 27% dos genótipos, utilizando como critério aqueles que apresentaram as melhores combinações para porte, produção de grãos, menor tempo para maturidade e qualidade do grão em feijão-caupi. Morceli (2009) verificou diferentes resultados para ganhos com seleção, herdabilidade e ganhos genéticos para caracteres primários e secundários de acordo com os genitores utilizados nos cruzamentos na cultura da soja avaliando genitores resistentes ao nematóide do cisto raça 3.

É preciso identificar quais genitores foram utilizados nestes cruzamentos ineficientes e verificar qual foi o resultado quando estes foram utilizados com outros genitores, pois não significa que em todas as recombinações um parental ineficiente em um cruzamento vai apresentar os mesmos resultados. Isto porque a genética é uma ciência complexa em que podem ocorrer diferentes resultados quando no cruzamento com diferentes genitores. Fato este que pode ser explicado pela capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC) de cada um dos genitores.

Outro ponto a ser discutido é em relação ao rigor da análise estatística utilizada no auxílio da seleção de populações segregantes. Neste caso, utilizando a mesma ACP, com um menor rigor quanto à delimitação das escalas de seleção, com certeza teríamos um valor mais expressivo de progênies selecionadas. Assim, teríamos um maior número de progênies que poderiam receber maior atenção no processo de melhoramento, podendo resultar na maior probabilidade de se conseguir o objetivo de obter algumas cultivares superiores. Por outro lado, a diminuição do rigor e aumento no número de progênies selecionadas consequentemente acarretaria no maior dispêndio de tempo, mão de obra, trabalho e custos no processo de melhoramento, podendo ainda ser ineficiente pelo fato de se utilizar de progênies com menor potencial agronômico.

O melhorista ou pesquisador é o responsável por definir qual será o rigor utilizado na seleção, baseado no estudo da divergência genética existente nos genitores, no número de famílias e do tamanho das populações segregantes em cada. Além disso, dependerá da estrutura física e financeira disponível e principalmente, no seu objetivo.

É importante ressaltar que estes resultados são baseados em uma população segregante, em geração F_3 que apresenta ainda uma alta heterozigose, em torno de 12,5%, portanto, para cada geração avançada é interessante realizar estas mesmas análises para verificar e comprovar o resultado da seleção.

A seleção final de progênies resultante da ACP e posteriormente com a comparação com os dados das variáveis VA e Ac, resultou na seleção de 77 progênies entre todos os cruzamentos utilizados no estudo, ou seja, uma seleção de 5,6% do total, listados na Tabela 8. Embora a porcentagem de seleção de progênies seja relativamente baixa devido ao rigor da seleção, está de acordo com o objetivo estipulado de selecionar até dez por cento do montante de progênies existente. Além disso, isto não significa que apenas as progênies selecionadas pela ACP seguirão para as próximas gerações, mas sim, que estas, deverão ser observadas com maior atenção no decorrer das etapas de melhoramento devido as suas características específicas que as diferenciaram das demais.

Com a ACP foi possível identificar dentro de uma grande quantidade de progênies inicialmente analisadas, aquelas que possuem certo potencial agronomicamente superior ou alguma característica diferencial e/ou específica, auxiliando o melhorista no processo de obtenção de uma nova linhagem para futuras hibridações e geração de novas populações com finalidade de criar variabilidade genética ou hibridações direcionadas buscando inserir, alterar ou excluir alguma característica específica ou ainda, selecionar uma nova cultivar com fins comerciais, de acordo com o objetivo pretendido.

Com relação aos genitores envolvidos nos cruzamentos, a análise identificou que os genitores femininos JAB.04-1/5A4D e JAB.05-1/5C3B, foram os mais eficientes, com 13 progênies selecionadas, respectivamente. Por sua vez, o genitor masculino M 8360 RR foi o mais eficiente, com 14 progênies selecionadas (Tabela 7). Ao analisar os cruzamentos que envolvem tais genitores mais eficientes combinados, verifica-se que o C3 apresenta cinco progênies selecionadas enquanto o C25 apenas duas progênies, resultados estes inferiores quando comparando com o C22 que apresentou a melhor combinação com seis progênies selecionadas. Este resultado bem como, o número de progênies por cruzamento e selecionados de todos os cruzamentos é apresentado na Tabela 7. Isso pode ocorrer pelo fato dos

genitores JAB.04-1/5A4D, JAB.05-1/5C3B e M 8360 RR apresentarem alta capacidade geral de combinação (CGC) com alta frequência de alelos favoráveis o que não significa que, conseqüentemente, também terão alta capacidade específica de combinação (CEC) e assim apresentar as melhores combinações, pois podem não possuir a complementariedade necessária entre si ou não apresentam divergência genética que permita maior exploração da variabilidade genética de tais combinações.

Tabela 7. Cruzamentos e respectivos números de progênies por cruzamento, progênies selecionadas pela ACP e progênies selecionadas após a ACP e comparativo com as variáveis VA e Ac.

Nº do Cruzamento	Nº de progênies por cruzamento	Nº progênies selecionadas pela ACP	Nº progênies selecionadas pela ACP + VA e Ac
C1	63	8	3
C2	36	3	3
C3	44	6	5
C4	24	2	2
C5	4	0	0
C6	5	0	0
C7	24	2	2
C8	39	5	1
C9	29	3	3
C10	40	3	1
C11	30	2	2
C12	28	3	2
C13	36	4	1
C14	41	3	1
C15	20	3	2
C16	44	9	2
C17	37	5	3
C18	64	7	4
C19	47	5	4
C20	48	5	0
C21	40	3	1
C22	48	12	6

Continua...

Continuação Tabela 7

Nº do Cruzamento	Nº de progênies por cruzamento	Nº progênies selecionadas pela ACP	Nº progênies selecionadas pela ACP + VA e Ac
C23	48	8	3
C24	36	2	1
C25	48	2	2
C26	26	2	1
C27	30	4	3
C28	52	2	1
C29	48	4	0
C30	38	3	3
C31	55	8	5
C32	35	4	2
C33	35	0	0
C34	40	2	1
C35	6	0	0
C36	53	7	4
C37	41	5	3
TOTAL	1382	146	77

Tabela 8. Relação das progênies superiores selecionadas pela Análise de Componentes Principais e submetidas as Análises de Agrupamentos. A nomenclatura inicia-se com a identificação da parcela, identificação sequencial da planta selecionada e o número do cruzamento, separadas por ponto, respectivamente.

1.2.C1	81.2.C9	153.3.C17	212.5.C23	294.1.C31
10.1.C1	81.3.C9	161.1.C18	216.1.C23	294.2.C31
10.2.C1	88.5.C10	161.4.C17	216.5.C23	294.3.C31
16.1.C2	91.3.C10	161.6.C18	223.6.C24	297.4.C32
20.5.C2	101.2.C11	162.1.C18	228.6.C25	300.1.C32
20.6.C2	103.1.C11	172.1.C19	229.1.C25	317.2.C34
28.1.C3	104.6.C12	174.2.C19	242.2.C26	330.3.C37
30.1.C3	106.5.C12	176.2.C19	253.3.C27	330.5.C37
30.2.C3	119.3.C13	176.5.C19	253.4.C27	336.1.C36
30.4.C3	123.1.C14	197.2.C21	253.5.C27	336.3.C36

Continua...

Continuação Tabela 8

31.2.C3	133.1.C15	199.2.C22	254.3.C28	336.4.C36
34.6.C4	133.5.C15	203.4.C22	272.6.C30	336.5.C36
37.2.C4	143.1.C16	203.5.C22	276.1.C30	344.5.C37
66.1.C7	144.5.C16	203.6.C22	276.2.C30	
66.3.C7	153.1.C17	204.4.C22	288.3.C31	
72.5.C8	153.2.C17	204.6.C22	290.1.C31	

Na continuidade, procedeu-se então a AA pelo método de K-means nas progênes selecionadas pela ACP. A Figura 5 ilustra o gráfico biplot gerado por esta análise contendo as médias e distribuições de cada grupo (*clusters*), de acordo com os dados de suas variáveis baseadas na distância Euclidiana entre os grupos.

Neste estudo, determinou-se que a divisão em seis grupos proporcionou uma resposta adequada quanto à distribuição dos resultados. A interpretação desta análise é melhor se feita independente, ou seja, observando grupo a grupo e depois, de forma geral.

Analisando o grupo um, observa-se que este possui o maior valor para a variável NDF, porém, esse valor diminui ao atingir a maturidade ficando próximo à média (NDM). Apresenta ainda progênes com porte abaixo da média de acordo com o resultado de AIV e APM. Para NR e NV os valores são relativamente altos. O PCS ficou abaixo da média e, aliado ao fato de NS não ser uma variável tão expressiva para este grupo, a variável PG foi influenciada, mesmo permanecendo acima da média. Isso indica que esse grupo tem um bom potencial produtivo, com os maiores valores para NR e NV e que possivelmente, devido aos fatores ambientais como estresse hídrico ocorrido durante parte do ciclo, fez com que ocorressem baixos valores de PCS. Porém, se as condições ambientais favorecerem a cultura, este resultado poderá ser positivo.

Analisando o grupo dois, observa-se que este possui um ciclo pouco acima da média, mas comparado com os demais grupos, apresenta o maior valor, caracterizando-se por ser formado por progênes com ciclo mais longo da germinação até a maturidade fisiológica. Apresenta ainda baixa AIV mesmo com APM próximo à média. Nesse caso, os valores inferiores de AIV não são tão expressivos, logo, as progênes deste grupo não apresentam maiores problemas

quanto à colheita, uma vez que, materiais que apresentam esse valor muito abaixo da média podem resultar em perdas quando da colheita mecanizada. As variáveis NR e NV encontram-se abaixo da média, porém PCS tem um valor expressivamente alto, o que caracteriza este grupo de progênies além do alto valor para NS. Com isso, as progênies deste grupo apresentam o maior valor encontrado para PG entre os grupos formados. Isto mostra que a combinação de PCS e NS determina diretamente o potencial produtivo de uma progênie, linhagem ou cultivar.

Para o grupo três, pode-se observar que embora algumas variáveis oscilem em relação média, de modo geral, é caracterizado por progênies que mantêm uma média geral regular. A variação mais expressiva ocorre na variável AIV, o que de certo modo, pode ser uma característica indesejável, ainda mais quando apresenta APM baixo, pois este fator pode diminuir a capacidade produtiva do material influenciando principalmente nas variáveis NR e NV.

No grupo quatro, todas as variáveis possuem valores abaixo da média. Para NDF e NDM isto é uma característica positiva, sendo este o segundo grupo com progênies mais precoces. Possui os menores valores tanto de AIV quanto APM sendo o grupo de progênies com o menor porte. Quando se trata de variáveis componentes de produção como NR, NV, PCS, NS e PG, estes valores abaixo da média, encontrado nesse grupo, são um ponto muito desfavorável, visto que caracterizam em progênies menos produtivas, mesmo que esses valores sejam baixos.

O grupo cinco conta com progênies com valores de NDF, NDM, AIV e APM acima da média, principalmente para APM, porém para que este valor resultasse em melhores resultados, as variáveis NR, NV, PCS e NS também deveriam acompanhar esta média, o que não ocorreu, resultando em um baixo valor de PG.

Por fim, o grupo seis caracteriza-se por apresentar progênies com valores expressivamente abaixo da média para NDF e NDM, sendo este o grupo de progênies caracterizado como as mais precoces de todo grupo estudado, tanto em relação à média quanto se comparado com os demais grupos. Isto significa que este grupo apresenta um ciclo mais curto, podendo ser usada essa informação em estudos posteriores quanto à possibilidade de se obter uma cultivar precoce ou utilizá-la em novas hibridações com o objetivo de se introgridir tal característica.

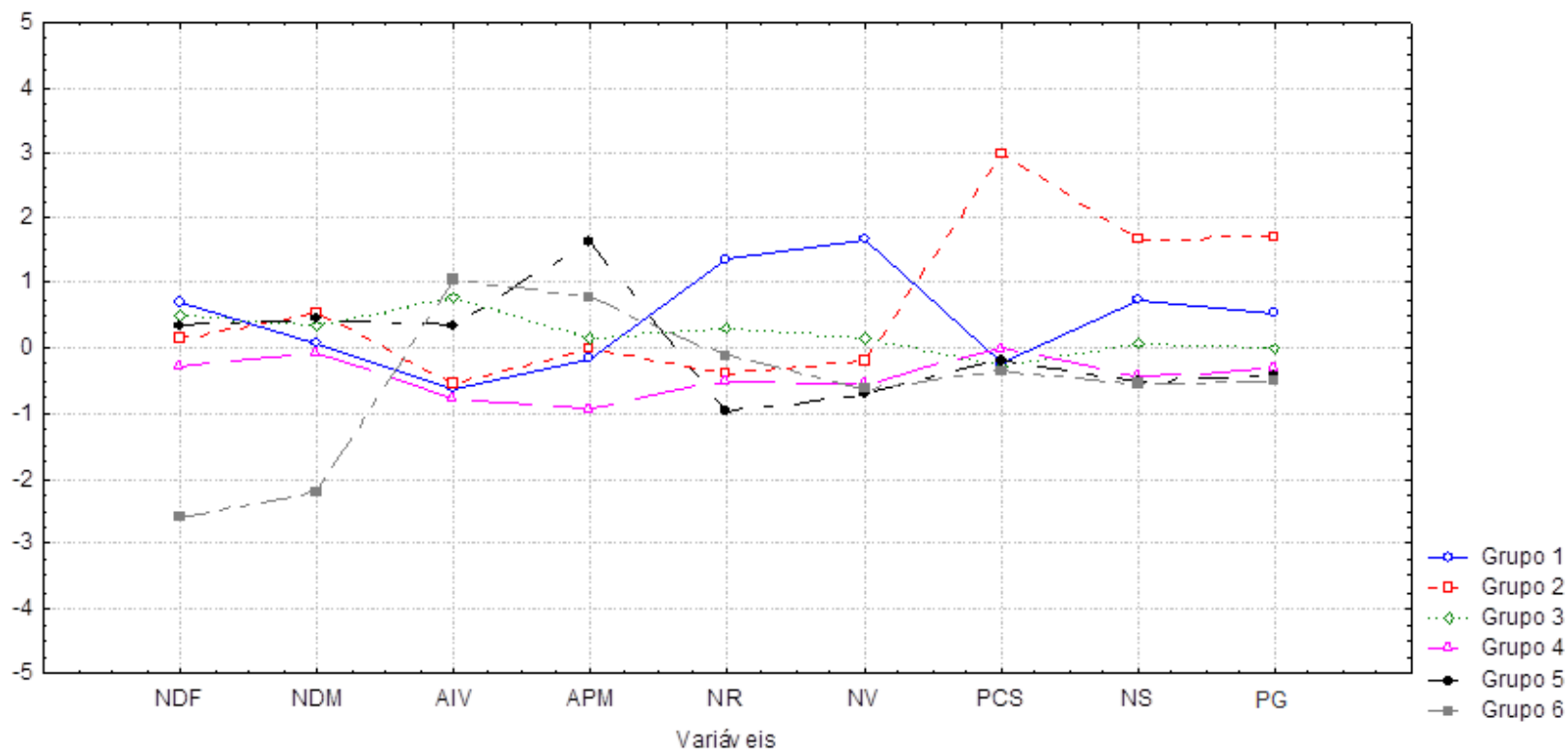


Figura 5. Gráfico biplot do perfil de centroides de cada grupo formado analisadas pelo método k-means e distância Euclidiana a partir das variáveis: número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da planta na maturidade (APM); altura da inserção da primeira vagem (AIV); número de ramos (NR); número de vagens (NV); número de sementes (NS); peso de cem sementes (PCS) e produção de grãos (PG).

Um fato muito comum é encontrar materiais que caracterizam-se por sua precocidade também apresentar baixos valores de AIV e APM justamente pelo fato de possuírem os estádios de desenvolvimento vegetativo e reprodutivo mais curtos. Neste caso, este grupo não apresentou esta característica, pois esses valores são acima da média. Este fator é muito influenciado pelo tipo de crescimento da soja, sendo classificado como determinado, semideterminado ou indeterminado, como descrito primeiramente por Takahashi e Fukuyama (1919) e Etheridge et al., (1929). Ao se tratar das variáveis componentes de produção, apresenta valores abaixo da média nas variáveis NR, NV, PCS, NS e PG resultando no grupo de progênies menos produtivas de todos os grupos analisados.

Tal resultado auxilia o pesquisador quanto à utilização de um grupo específico de progênies, de acordo com uma característica buscada, uma possível cultivar com característica (variável) diferencial ou para cruzamentos direcionados.

A seguir, procedeu-se a Análise de Agrupamento pelo método de Ward para formação do dendrograma objetivando comparar o resultado da formação de grupos entre as diferentes análises, possibilitando inferir, com mais consistência, se as seleções e análises das progênies realizadas pelos métodos já descritos são fundamentados para aplicação em populações segregantes de soja.

Para melhor compreensão e comparação entre o resultado da divisão de grupos feita pelos dois métodos de AA, cada grupo gerado pelo método K-means recebeu uma cor distinta em sua nomenclatura. Assim, o grupo um possui cor marrom, grupo dois cor azul, grupo três cor preta, grupo quatro cor roxa, grupo cinco cor verde e por fim, grupo seis cor vermelha

Essa análise resultou na divisão das progênies em dois grandes grupos, sendo denominados grupos A e B, conforme Figura 6. Com mais critério, utilizando-se como base a distância de ligação de valor cinco, pode-se observar ainda a formação de seis subgrupos, sendo, três subgrupos dentro do grupo A denominados de A1, A2 e A3, destacados por retângulos de cor vermelha e três subgrupos dentro do grupo B denominados de B1, B2 e B3, destacados por retângulos de cor azul. Os grupos e subgrupos podem ser

divididos de acordo com a distância de ligação indicada pelo pesquisador, ficando a critério do mesmo, determinar a distância que mais lhe convém de acordo com o genótipo, rigor da seleção e objetivo final ou ainda, obedecer algum critério específico de cada método. O grupo A é representado por 40 progênies, sendo 10 no subgrupo A1, 19 no subgrupo A2 e 11 no subgrupo A3. O grupo B é constituído por 45 progênies, divididas em 17 para o subgrupo B1, 13 para o subgrupo B2 e 15 para o subgrupo B3.

Embora o número de progênies por grupo variasse de 40 e 45 e nos subgrupos variasse de 10 a 19, houve uma boa distribuição entre grupos e subgrupos, indicando a eficiência da análise na determinação e separação dos mesmos.

Em função da diferenciação de grupos por cores, resultante da AA pelo método K-means, é possível verificar no dendrograma gerado pelo método de Ward que houve semelhança entre os métodos quando comparados com a localização das progênies dentro dos grupos A e B. No grupo A apenas quatro progênies (37.2.C4, 174.2.C19, 172.1.C19 e 228.6.C25) ou 10%, deslocaram-se dos demais representantes de seu grupo que restaram no grupo B. No grupo B apenas 2 progênies (119.3.C13 e 317.2.C34) ou 4,4% deslocaram-se dos demais representantes de seu grupo que restaram no grupo A. Quando se compara os subgrupos, verifica-se que houve uma dispersão de progênies em diferentes subgrupos, como por exemplo, progênies do subgrupo A1 alocadas no subgrupo B3, do subgrupo B3 alocadas no A3, do subgrupo B1 alocadas no A3. Esta pequena variação pode ser explicada pelo fato de que quanto menor a distância genética buscada, maior será a variação e formação de subgrupos. Além disso, deve-se considerar que existem diferenças nas metodologias de análise em cada AA. Essa pequena diferença existente entre os métodos não caracterizou a impossibilidade de comparação dos resultados gerados entre ambos.

Resultado semelhante foi encontrado por José et al. (2012) quando compararam a análise de agrupamento hierarquizado, não hierarquizado (K-means) e análise de componentes principais das variáveis físico-hídricas do solo, as quais mostraram-se semelhantes aos agrupamentos e suas interações.

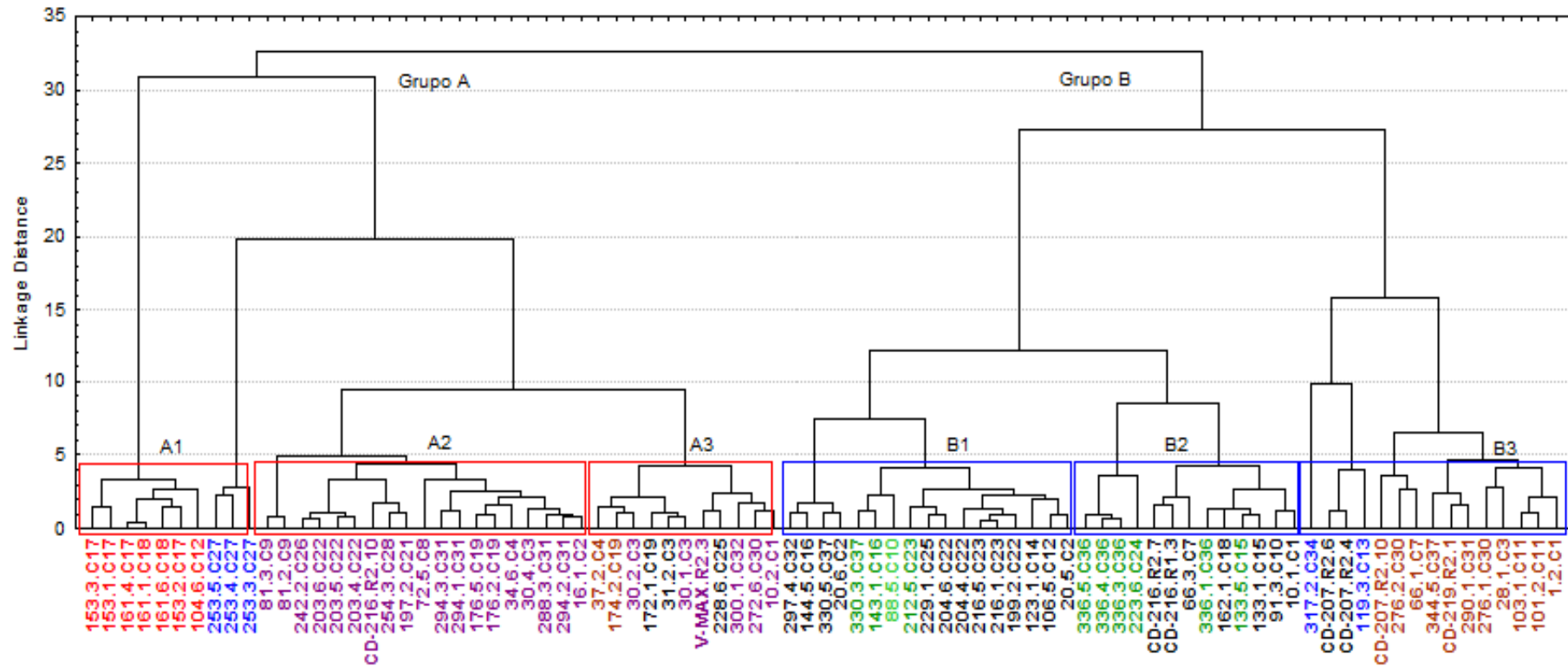


Figura 6. Dendrograma da análise de agrupamento hierárquico utilizando a distância Euclidiana e a ligação entre os grupos pelo método de Ward, com indicação das progênes e do grupo de classificação a qual pertencem. O eixo X representa as progênes e o eixo Y representa as distâncias para os caracteres agrônômicos de acordo com as variáveis: número de dias para o florescimento (NDF); número de dias para a maturidade (NDM); altura da planta na maturidade (APM); altura da inserção da primeira vagem (AIV); número de ramos (NR); número de vagens (NV); número de sementes (NS); peso de cem sementes (PCS) e produção de grãos (PG).

No estudo da compactação de um latossolo vermelho cultivado com milho Freddi, Ferraudo e Centurion (2008) aplicaram AA pelo método K-means para a confirmação da ordenação obtida na AA hierárquica e os resultados confirmaram a ordenação, reforçando o resultado obtido neste trabalho.

Também Seidel et al. (2008) no estudo com agrupamento por similaridade de produtores de leite pelos métodos Ward e K-médias, relataram que os resultados encontrados tiveram alta concordância evidenciando a eficiência da robustez dos agrupamentos formados pelos dois métodos.

Costa (2008) destaca que os resultados do agrupamento pelo método K-means foram similares com o método de Ward no estudo quali-quantitativo da resistência de genótipos de soja à ferrugem asiática.

Após as progênes atingirem um nível satisfatório de homozigose será realizada a seleção especificamente para o gene RR por meio da aplicação de glifosato e verificação à resistência e suscetibilidade destas.

5. CONCLUSÕES

A Análise de Componentes Principais auxiliou na seleção de 77 progênies com caracteres superiores e de importância agrônômica em soja, principalmente para os componentes relacionados à produção de grãos.

Possibilitou a identificação dos genitores JAB.04-1/5A4D, JAB.05-1/5C3B e M 8360 RR como mais eficientes na obtenção de progênies superiores. E concluiu-se que o cruzamento C22 (JAB.05-1/5C3B x M 7639 RR) foi o que obteve maior número de progênies selecionadas neste estudo.

A Análise de Agrupamentos pelo método de K-means permitiu identificar e caracterizar os diferentes grupos de progênies de acordo com os caracteres agrônômicos que mais influenciaram o processo de seleção. A análise de Agrupamento pelo método de Ward possibilitou a formação do dendrograma e caracterização da similaridade e divergência genética existente entre os grupos de progênies selecionadas. Os resultados encontrados tiveram boa concordância, evidenciando a eficiência dos agrupamentos formados pelos dois métodos.

Os métodos multivariados foram eficientes, auxiliando a seleção de progênies com múltiplos caracteres de interesse agrônômico, confirmando a potencialidade do uso dos mesmos em programas de melhoramento da soja.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALCANTARA NETO, F.; GRAVINA, G. A.; MONTEIRO, M. M. S.; MORAIS, F. B.; PETTEN, F. A.; ALBUQUERQUE, J. A. A. Análise de trilha do rendimento de grãos de soja na microrregião do Alto Médio Gurguéia. **Comunicata Scientiae**, Piauí, v.2, p.107-112, 2011.

ALLARD, R. W. **Princípios de melhoramento genético das plantas**. Rio de Janeiro: Edgar Blücher, p. 381. 1971.

ALMEIDA, L.A.; KIIHL, R.A.S de. Melhoramento da soja no Brasil – Desafios e perspectivas. In: CÂMARA, G.M.S. **Soja: tecnologia da produção**. Piracicaba: USP, 1998. p. 1-25.

ANDRADE, L.P. **Procedimento Interativo de Agrupamento de dados**. [Rio de Janeiro] 2004 VIII, 193 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Civil, 2004) Tese - Universidade Federal do Rio de Janeiro, COPPE

BACKES, R. L.; REIS, M. S.; SEDIYAMA, T.; CRUZ, C. D.; TEIXEIRA, R.C. Estimativas de parâmetros genéticos em populações F5 e F6 de soja. **Revista Ceres**, v. 49, n. 2, p. 201-216, 2002.

BARROS, H. B.; SILVA, A. A.; SEDIYAMA, T. Manejo de plantas daninhas. In: SEDIYAMA, T. (ed.). **Tecnologias de produção e usos da soja**. Londrina: Mecenias, 2009. 314p.

BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury: Stemma Press, 2002, p. 360.

BERTINI, C. H. C. M.; ALMEIDA, W. S.; SILVA, A. P. M.; LIMA, J. W.; TEÓFILO, E. M. Análise multivariada e índice de seleção na identificação de genótipos superiores de feijão-caupi. **Acta Scientiarum. Agronomy**. Maringá, v. 32, n. 4, p. 613-619, 2010.

BORÉM, A.; ALMEIDA, L.A de.; KIIHL, R.A.S. Hibridação em soja. In: BORÉM, A. (Ed). **Hibridação artificial de plantas**. 2. Ed. atual. ampl. – Viçosa, MG : Ed. UFV, 2009. 625p. : il. 22 cm.

BORÉM, A.; MIRANDA, G. V. **Melhoramento de plantas**. 4.ed. – Viçosa : UFV, 2005. 525 p.. : il. 22 cm.

CÂMARA, G.M.S. Origem, difusão geográfica e importância da soja. In: CÂMARA, G.M.S. **Soja: tecnologia da produção**. Piracicaba: USP, 1998. p. 1-25.

CELERES – CELERES Consultoria. **Relatório de Biotecnologia, segundo levantamento de área cultivada com transgênicas**. 2012. Disponível em: < http://celeres.com.br/wordpress/wp-content/uploads/2012/12/RelBiotecBrasil_1202_por.pdf > Acesso em: 30 maio 2013.

CHUANG, G.; SINGH, R.J. Broadening the genetic base of soybean: A multidisciplinary approach. **Critical Reviews in Plant Science**, v.2, n.27, p.295-341, 2008.

COLE, D. J. Mode of action of glyphosate – a literature analysis. In: GROSSBARD, E.; ATKINSON, D. (Ed.). **The herbicide glyphosate**, pp. 49-54. Londres: Butterworths, 1985

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Séries históricas de produtividade de grãos**. Disponível em: <http://www.conab.gov.br> >. Acesso em: 15 set 2011.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileira: grãos, sexto levantamento, março 2013** / Companhia Nacional de Abastecimento. – Brasília: Conab, 2013. Disponível em: < http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_03_07_10_39_19_levantamento_safras_graos_6.pdf >. Acesso em: 08 jun 2013.

COSTA, J.A. **Cultura da soja**. Porto Alegre: I. Manica, J.A. COSTA, 1996. 233 p. : il.

COSTA, M. M. **Herança quali-quantitativa e marcadores moleculares para seleção assistida de genótipos de soja resistentes à ferrugem asiática**. Jaboticabal, 2008, 77 p. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2008.

COSTA, M. M.; MAURO, A. O.; UNÊDA-TREVISOLI, S. H.; ARRIEL, N. H. C.; BÁRBARO, I. M.; MUNIZ, F. R. S. Ganho genético por diferentes critérios de seleção em populações segregantes de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 11, p. 1095-1105, 2004.

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

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

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

CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa: UFV, 2006. v. 2

CUNHA, E. P. C. **Avaliação de ensaio final de linhagens UFV de soja de ciclo tardio nas condições de Uberlândia, Minas Gerais**. 1999. 4af. Monografia (Conclusão do curso de graduação em Agronomia) – Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, 1999.

DALL'AGNOL, A. **A soja no brasil: evolução, causas, impactos e perspectivas**. In: CONGRESO DE LA SOJA DEL MERCOSUR, 5.; FORO DE LA SOJA ASIA, 1., 2011, Rosário. Un grano: un universo. [Rosário: Asociación de la Cadena de la Soja Argentina], 2011. 4 p. 1 CD-ROM. MERCOSOJA 2011.

DAVIS, J. C. **Statistics and data analysis in geology**. New York: John Wiley & Sons, 1986. 646 p.

DEUBER, R. **Ciência das plantas daninhas: fundamentos**. Funep, Jaboticabal. 1992. 431 p.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Tecnologias de produção de soja – Região Central do Brasil - 2005**. Londrina: Embrapa Soja: Embrapa Cerrados: Embrapa Agropecuário Oeste: Fundação Meridional, 2004. 239p. (Sistemas de Produção / Embrapa Soja, ISSN 1677-8499; n.6)

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Tecnologias de produção de soja – Região Central do Brasil, 2012 e 2013**. - Londrina: Embrapa Soja, 2011. 261p. (Sistemas de Produção / Embrapa Soja, ISSN 2176-2902; n.15)

EMBRAPA. Centro Nacional de Pesquisa de Soja. **Tecnologia de Produção de Soja – Região Central do Brasil 2007**. Londrina: Embrapa Soja: Embrapa Cerrados: Embrapa Agropecuária Oeste. 2006. 225p.

ETHERIDGE, W.C.; HELM, C.A & KING B.M. **A classification of soybeans.** Mo. Agric. Exp. Sta., 1929. 154p (Research Bulletin, 131).

FEHR, W. R. **Principles of Cultivar Development.** 3ª ed. v.1 Ames: Macmillan. Publishing Company, 1993. 527p.

FEHR, W.R. **Principles of cultivar development.** Theory and technique. New York: Macmillan Publishing Company, 1987. 536p.

FEHR, W.R.; CAVINESS, C.E. **Stages of soybean development.** Ames: State University of Science and Technology, 1977. 11 p. (Special report, 80).

FERRAUDO, A.S. **Técnicas de Análise Multivariada – uma introdução.** Treinamento. StatSoft South América, São Caetano, SP, 2010.

FREDDI, O. S.; FERRAUDO, A. S.; CENTURION, J. F. Análise multivariada na compactação de um latossolo vermelho cultivado com milho. **Rev. Bras. Ciênc. Solo** [online]. 2008, vol.32, n.3, pp. 953-961. ISSN 0100-0683.

GUIDINI, M.B.; NASCIMENTO, A.M.; BONE, R.B.; ALVES, T.W. **Aplicação do K-Means Cluster para classificação de estilos gerenciais.** XXVIII Encontro Nacional de Engenharia de Produção. Rio de Janeiro, RJ, Brasil, 2008. Disponível em: http://www.abepro.org.br/biblioteca/enegep2008_TN_STO_076_539_11138.pdf <http://
acesso em julho de 2012.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. BABIM, B. J. **Análise multivariada de dados.** SANT'ANNA, A. S. (trad.). 5. ed. – Porto Alegre : BOOKMAN, 2005. 688p. : il. ; 28 cm.

HAIR, J. F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. 2009. **Análise multivariada de dados.** 6. ed. Porto Alegre: Bookman, 2009. 688 p.

HAN, J.; KAMBER, M., **Data Mining: Concepts And Techniques,** Second Edition, Morgan Kaufmann, 2006.

JAMES, C. **Global Status of Commercialized Biotech/GM Crops: 2010.** ISAAA Brief Nº. 42. ISAAA: Ithaca, NY. 2010. ISBN: 978-1-892456-49-4

JAMES, C. **Global Status of Commercialized Biotech/GM Crops: 2012**. ISAAA Brief No. 44. ISAAA: Ithaca, NY. 2012

JEZOVSEK, G.K. Soja transgênica no Brasil: situação atual e futura. In: CÂMARA, G.M.S. **Soja: tecnologia da produção**. Piracicaba: USP, 1998. p. 55-64.

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

JOSÉ, J. V.; Wolff, W.; DIOTTO, A. V.; ALVES, D. S.; MARQUES, P. A. A.; GAVA, R. **Análise multivariada de variáveis físico-hídricas de diferentes solos**. IV WINOTEC – WORKSHOP INTERNACIONAL DE INOVAÇÕES TECNOLÓGICAS NA IRRIGAÇÃO, 4, Fortaleza. Ceará, Brasil. 2012. 5p.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**. 1958. 23:187-200

LANDIM, P.M.B. **Análise estatística de dados geológicos multivariados**. Lab. Geomatemática, DGA, IGCE, UNESP/Rio Claro, Texto Didático 03, 128pp. 2000. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>> Acesso em julho de 2012.

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).

LOURENÇO, A.; MATIAS, R.P. **Estatística Multivariada**. Instituto Superior de Engenharia do Porto, 2000.

MAGALHÃES, C.M. Introdução e evolução da soja no Brasil No estado do Rio Grande do Sul. In: MIYASAKA, S.; MEDINA, J.C (Eds.) **A soja no Brasil**. Campinas: Instituto de Tecnologia de Alimentos. 1981. p.18-20

MANLY, B.F.J. Discriminant function analysis. In:____. **Multivariate statistical methods: a primer**. London: Chapman and Hall, 1986, p. 87-89.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academic Press, 1979. 521 p.

MARINIS, G. Ecologia das plantas daninhas. In: **texto básico de controle químico de plantas daninhas**, 3. ed., Piracicaba: ESALQ, 1971. p.1-74.

MENDELSON, J. Round up: o herbicida mais vendido no mundo. **Ecologist**, v. 28, p. 24-27, 1998.

MORCELI, A. A. J. **Análise genéticas de populações de soja com genitores resistentes ao nematóide do cisto raça 3** – Jaboticabal, 2009 iv, 62 f.; 28 cm. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2009

NAVARRO JÚNIOR, H.; COSTA, J. A. Contribuição relativa dos componentes do rendimento para produção de grãos em soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.3, p.269-274, 2002.

PEEL, M.D.; RASMUSSEN, D.C. Improvement strategy for mature plant breeding programs. **Crop Science**, Madison, v.40, n.5, p.1241-1246, 2000.

PROCÓPIO, S.O.; PIRES, F.R.; MENEZES, C.C.E.; BARROSO, A.L.L.; MORAES, R.V.; SILVA, M.V.V.; QUEIROZ, R.G.; CARMO, M.L Efeitos de dessecantes no controle de plantas daninhas na cultura da soja. **Planta Daninha**, v. 24, n. 1, p. 193-197, 2006

RAMOS, S.F.C. **Utilização de Técnicas de *Data Mining* para Apoio aos Agentes dos Mercados Retalhistas de Energia Elétrica**. 2006 (IST - Engenharia Electrotécnica e de Computadores, 2006) 242 p. Dissertação – Universidade Técnica de Lisboa – Instituto Superior Técnico.

REDDY, B.V.S.; COMSTOCK, R.S. Simulation of the backcross method: I. Effects of heritability and gene number on fixation of desired alleles. **Crop Science**, Madison, v.16, n.6, p.825- 830, 1976.

REZENDE, P. M.; CARVALHO, E. A. Avaliação de cultivares de soja [*Glycine max* (L.) merrill] para o sul de minas gerais. **Ciência e Agrotecnologia**, Lavras-MG , v. 31, n. 6, p. 1616-1623, nov./dez., 2007.

RODRIGUES, J. D. **Absorção, translocação e modo de ação de defensivos (glifosato e alachlor)**. Apostila. Botucatu: Unesp, 1994. 10 p.

SAAD, D.S. **Aplicação de técnicas estatísticas multivariadas em dados de cerâmica vermelha produzida na região central do Rio Grande do Sul**. Santa Maria – RS. 2009. 166 p. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Santa Maria, 2009.

SEDIYAMA, T.; PEREIRA, M. G.; SEDIYAMA, C. S.; GOMES, J. L. L. **Cultura da soja**: 1ª parte. Viçosa: UFV, 1989.

SEDIYAMA, T.; TEIXEIRA, R. C.; REIS, M. S. Melhoramento da soja. In: BORÉM, A. (Ed.) **Melhoramento de espécies cultivadas**. Viçosa: Ed. UFV, 2005. 969p. : il. ; 23 cm.

SEIDEL, E.J.; MOREIRA JÚNIOR, F. de J.; ANSUJ, A.P.; NOAL, M.R.C. Comparação Entre o Método Ward e o Método K-Médias no Agrupamento de Produtores de Leite. **Ciência e Natura**, v. 30, n. 1, p. 7-15. 2008.

SILVA, A. A.; FERREIRA, F. A.; FERREIRA, L. R.; SANTOS, J. B. Biologia das plantas daninhas. In: SILVA, A. A.; SILVA, J. F. (eds). **Tópicos em manejo de plantas daninhas**. Viçosa : Ed. UFV, 2007, 367p. : il. ; 22 cm.

STATSOFT, Inc. (2004). STATISTICA (data analysis software system), version 7. Disponível em: www.statsoft.com.

TAKAHASHI, Y. E FUKUYAMA, J. **Morphological and genetic studies on the soybean**. Hokkaido. Agr. Exp. Sta. Rep., 1919. 100p

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. **Production, Supply and Distribution Online. 2011**. Disponível em: <http://www.fas.usda.gov/psdonline/psdQuery.aspx> >. Acesso em: 30 mar 2013.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. **Production, Supply and Distribution Online. 2013**. Disponível em: <http://www.fas.usda.gov/psdonline/psdQuery.aspx> >. Acesso em: 15 Jun. 2013.

VENCATO, A. Z. **Anuário Brasileiro da Soja 2010**. Santa Cruz do Sul: Ed. Gazeta Santa Cruz, p. 144, 2010.

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

VERNETTI, E J.; BONATTO, E.R. GASTAL, M.F.C.; DALLAGNOLL, A; RAUPP, A.A.A.; VELOSO, J.A.R.O. Melhoramento de cultivares no Brasil. No Estado do Rio Grande do Sul, 1.2.1- Pelo Instituto de Pesquisas Agropecuárias do Sul. In: MYIASAKA, S.; MEDINA, J.C. (Eds.). **A soja no Brasil**. Campinas: ITAL, 1981. p. 283-292

VIANNA, V. F. **Estudo genético e influência de caracteres na seleção de genótipos superiores de soja** – Jaboticabal, 2013 xi, 92 p.; 28 cm. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2013

VICINI, L. **Análise multivariada da teoria à prática**. (monografia especialização – Universidade Federal de Santa Maria) UFSM, CCNE, 2005. 215p. : il.

YUYAMA, K. **Avaliação de algumas características agronômicas e morfofisiológicas de cinco cultivares de soja (*Glycine max* (L.) Merrill), cultivados em solo de várzea e de terra firme da Amazônia Central**.1991, 123f, Tese (Doutorado em Produção Vegetal)-Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal.