

ALISSON SANTOS LOPES DA SILVA

**DIVERSIDADE GENÉTICA ENTRE GENÓTIPOS DE SOJA E
ESTABELECIMENTO DE COLEÇÕES NUCLEARES E DE
MELHORAMENTO PARA TEORES DE ÓLEO E DE PROTEÍNA**

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

**VIÇOSA
MINAS GERAIS – BRASIL
2016**

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

T

S586d
2016
Silva, Alisson Santos Lopes da, 1986-
Diversidade genética entre genótipos de soja e
estabelecimento de coleções nucleares e de melhoramento para
óleo e de proteína / Alisson Santos Lopes da Silva. – Viçosa,
MG, 2016.

ix, 50f. : il. (algumas color.) ; 29 cm.

Orientador: Felipe Lopes da Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 43-50.

1. Soja - Melhoramento genético. 2. Diversidade genética.
I. Universidade Federal de Viçosa. Departamento de Fitotecnia.
Programa de Pós-graduação em Fitotecnia. II. Título.

CDD 22. ed. 633.34

ALISSON SANTOS LOPES DA SILVA

**DIVERSIDADE GENÉTICA ENTRE GENÓTIPOS DE SOJA E
ESTABELECIMENTO DE COLEÇÕES NUCLEARES E DE
MELHORAMENTO PARA TEORES DE ÓLEO E DE PROTEÍNA**

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

APROVADA: 29 de Fevereiro de 2016

Pedro Crescêncio Sousa Carneiro
(Coorientador)

Carlos Nick Gomes

Leiri Daiane Barili

Felipe Lopes da Silva
(Orientador)

*Aos meus pais, irmãs e amigos,
Dedico.*

AGRADECIMENTOS

A Deus, pelo dom da vida, serenidade e coragem nos momentos de insegurança e incertezas.

À Universidade Federal de Viçosa, através do Departamento de Fitotecnia, pela oportunidade de realizar o curso de mestrado.

À coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão da bolsa de estudos.

Ao professor Felipe Lopes da Silva, pelas orientações, ensinamentos e amizade.

Aos Professores Pedro Crescêncio Sousa Carneiro, Cosme Damião Cruz e Tuneo Sedyama, pelas orientações, sugestões e ensinamentos.

Aos funcionários do Laboratório de oleaginosas do Departamento de Fitotecnia pelo apoio na execução dos trabalhos.

Aos estagiários do programa soja pela amizade e ajuda na condução dos experimentos.

Aos amigos Amilton Ferreira, André Ricardo, Daniele Piano, Danúbia Nobre, Francisco Charles, Guilherme Sousa e Stênio Dantas pela amizade, apoio, esclarecimentos e sugestões sempre pertinentes.

Aos demais estudantes de Pós-graduação do Laboratório de Oleaginosas pelo auxílio, oportunidade de conviver e interagir.

Aos amigos Gutierrez Nelson e Marcela Carvalho pela amizade, companheirismo e incentivo incessante em realizar o curso de mestrado.

Aos meus pais, Carlos Lopes e Nilza Maria, pelo amor, carinho, apoio e exemplo de vida.

Às minhas irmãs, Aline Aparecida e Ariane Dianas, pelo amor, carinho, apoio, paciência e estímulo.

Enfim, o meu reconhecimento à todos que contribuíram de forma direta ou indireta, para a realização do presente trabalho.

BIOGRAFIA

Alisson santos Lopes da Silva, filho de Carlos Lopes da Silva e Nilza Maria do Santos Silva, nasceu em vinte cinco de fevereiro de 1986, na cidade de Viçosa, Minas Gerais.

Em janeiro de 2009, graduou-se em Tecnologia de Laticínios pelo Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, Campus Rio Pomba.

Em Agosto de 2009, ingressou na Universidade Federal de Viçosa (UFV) - Minas Gerais, graduando-se em Agronomia em Janeiro de 2014.

Em Março de 2014, iniciou o curso de Mestrado em Fitotecnia, na Universidade Federal de Viçosa, Minas Gerais.

SUMÁRIO

RESUMO	vi
ABSTRACT	viii
INTRODUÇÃO	1
MATERIAL E MÉTODOS	4
RESULTADOS E DISCUSSÃO	13
CONCLUSÕES	42
REFERÊNCIAS BIBLIOGRÁFICA	43

RESUMO

SILVA, Alisson Santos Lopes da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2016. **Diversidade genética entre genótipos de soja e estabelecimento de coleções nucleares e de melhoramento para teores de óleo e de proteína.** Orientador: Felipe Lopes da Silva. Coorientadores: Pedro Crescêncio Sousa Carneiro, Cosme Damião Cruz e Tuneo Sedyama.

A presença de variabilidade genética entre as cultivares de soja é fator determinante para o sucesso de programas de melhoramento da cultura. O objetivo deste trabalho foi avaliar a diversidade genética entre genótipos de soja, estabelecer coleções nucleares e definir estratégias para extração de coleções de melhoramento para teores de óleo e de proteína. A caracterização dos acessos ocorreu por meio de 21 caracteres morfoagronômicos avaliados em experimentos conduzidos em dois locais, utilizando o delineamento de blocos aumentados (DBA). A estimação da diversidade foi realizada por meio do método de agrupamento UPGMA utilizando como medida de dissimilaridade a distância Euclidiana média. No estabelecimento das coleções nucleares utilizou-se inicialmente a estratégia que definiu um número de indivíduos que representasse a média e a variância de todas as características quantitativas, calculada na população original de cada local. A formação dessas coleções aconteceu mediante alocação de acessos tomados ao acaso do conjunto de genótipos analisados. Foram avaliados os impactos das coleções nucleares na manutenção da média e variância original. A estruturação de coleções de melhoramento foi realizada seguindo a estratégia de estratificação dos acessos em relação ao ciclo total e seleção destes, dentro de cada estrato de ciclo, com alta média para cada caractere em questão e divergentes. As análises estatísticas foram realizadas por meio do Programa Computacional GENES. Foi possível verificar variância genética significativa, ao nível de 5% de probabilidade pelo teste F para a maioria dos caracteres quantitativos avaliados no local 1. No local 2 sete dos caracteres apresentaram diferença significativa. Na matriz de dissimilaridade para os dados quantitativos, a distância Euclidiana entre os acessos variou de 0,3753 a 3,2529 no local 1, e de 0,3679 a 3,4008 no local 2. No local 1 os pares de acessos mais divergentes e mais similares foram o BAGSUFV106/BAGSUFV36 ($d_{ii}' = 3,2529$) e BAGSUFV157/BAGSUFV159 ($d_{ii}' = 0,3753$) respectivamente. No local 2 o par BAGSUFV48/BAGSUFV16 ($d_{ii}' = 3,4008$) foram os mais dissimilares e BAGSUFV59/BAGSUFV16 ($d_{ii}' = 0,4211$) os mais similares. Para o conjunto de caracteres qualitativos, os pares de acessos mais divergentes e similares foram os

BAGSUFV38/BAGSUFV75 ($d_{ii}' = 298,0604$) e BAGSUFV2/BAGSUFV22 ($d_{ii}' = 7,9876$), respectivamente. O método de agrupamento UPGMA proporcionou a formação de 57 grupos, sendo trinta destes, composto por apenas um acesso. O número de acessos determinado para compor a coleção nuclear de cada local foi de 80 genótipos. Destes 80, a porcentagem de acessos comuns aos dois ambientes foi de apenas 48,75%. A coleção nuclear do local 1 não foi capaz de representar a variância original somente para o caráter número de hastes laterais. No local 2 a manutenção das variâncias originais não verificada para os caracteres diâmetro do hipocótilo, número hastes laterais e angulo de acamamento. Para as características qualitativas as classes formato do folíolo lanceolado estreito e hábito de crescimento prostrado foram as únicas não representadas pela coleção nuclear do ensaio 1 e 2, respectivamente. Na estruturação da coleção de melhoramento houve uma tendência, no local 1, da alocação de genótipos com maiores teores de óleo e proteína no grupo de ciclo tardio. Já no local 2 esta tendência ocorreu para os genótipos pertencentes ao grupo de ciclo médio. Analisando cada ciclo de maturação verificou-se que o genótipo BAGSUFV23, para o caractere teor de óleo, foi coincidente em ambos os locais. O mesmo ocorreu para os genótipos BAGSUFV17 e BAGSUFV15, em relação ao teor de proteína. Para o ciclo médio o genótipo comum para teor de óleo foi o BAGSUFV50 e para proteína o BAGSUFV82. No grupo do ciclo tardio foi verificado o maior número de genótipos coincidentes para o teor de óleo, sendo eles: BAGSUFV131, BAGSUFV115, BAGSUFV33 e BAGSUFV53. Neste grupo não houve acessos comuns para o caractere teor de proteína. Conclui-se que existe diversidade genética entre os acessos de soja avaliados, foi possível obter coleções nucleares para cada local e, a estratégia adotada se mostrou eficientes na estruturação das coleções de melhoramento.

ABSTRACT

SILVA, Alisson Santos Lopes da, M.Sc., Universidade Federal de Viçosa, February, 2016. **Genetic diversity among soybean and establishing of nuclear and breeding collections for oil and protein contents.** Advisor: Felipe Lopes da Silva. Co-advisors: Pedro Crescêncio Sousa Carneiro, Cosme Damião Cruz and Tuneo Sedyama.

The presence of genetic variability among soybean cultivars is a decisive factor for the success of crop improvement programs. The objective of this study was to evaluate the genetic diversity among soybean genotypes, establish nuclear collections and define strategies to extract improvement collections for oil and protein content. The characterization of the accessions occurred through 21 morphoagronomic characters evaluated in experiments conducted in two locations, using the augmented block design (DBA). The estimation of diversity was performed using the UPGMA grouping method, using as measure of the dissimilarity the average Euclidean distance. In the establishment of core collections was used initially the strategy that has defined a number of individuals who represent the mean and the variance of all quantitative characteristics, calculated on the original population of each location. The formation of these collections happened through the allocation of acesses taken at random from the set of analyzed genotypes. It was evaluated the impacts of nuclear collections in maintaining of the mean and original variance. The structure of the improvement collections was carried out following the strategy of the stratification of the acesses in relation to the total cycle and selection of these, within each stratum of the cycle, with a high average for each trait in question and divergent. The statistical analyzes were performed using the Genes Computational Program. It was possible to verify significant genetic variance at 5% probability by the F test for most quantitative traits evaluated at the location 1. On site 2, seven of the characters showed significant difference. In dissimilarity matrix for quantitative data, the Euclidean distance between accessions ranged from 0,3753 to 3,2529 on site 1, and from 0,3679 to 3,4008 on site 2. In the site 1 the pairs of the acesses more divergent and more similar were BAGSUFV106 / BAGSUFV36 ($d_{ii}' = 3,2529$) and BAGSUFV157 / BAGSUFV159 ($d_{ii}' = 0,3753$), respectively. In the site 2, the pair BAGSUFV48 / BAGSUFV16 ($d_{ii}' = 3,4008$) were the most dissimilar and BAGSUFV59 / BAGSUFV16 ($d_{ii}' = 0,4211$) the most similar. For the set of the qualitative characters, the pairs of the acesses more divergent and similars were BAGSUFV38 / BAGSUFV75 ($d_{ii}' = 298,0604$) and BAGSUFV2/BAGSUFV22 ($d_{ii}' = 7,9876$), respectively. The UPGMA grouping method provided the formation of

57 groups, where thirty of these, being composed of just one access. The number of accesses determined to compose the core collection of each site was 80 genotypes. From these 80, the percentage of common accesses to both environments was only 48,75%. The nuclear collection of the site 1 was not able to represent the original variance only to the character number of lateral rods. In the site 2 the maintenance of the original variances was not checked for the characters of hypocotyl diameter, number of side rods and angle of lodging. For the qualitative characteristics, the classes leaflet format narrow lanceolate and prostrate growth habit were the only ones that was not represented by the core collection of test 1 and 2, respectively. In structuring the collection of the improvement there was a trend, in place 1, of the allocation of genotypes with higher levels of oil and protein in the group of late cycle. Already in site 2, this trend occurred for genotypes belonging to the average cycle group. Analyzing each maturation cycle it was found that the BAGSUFV23 genotype, to the oil content character, was similar in the both locations. The same was true for BAGSUFV17 and BAGSUFV15 genotypes in relation to the protein content. For the mean cycle, the common genotype for oil content was BAGSUFV50 and to the protein the BAGSUFV82 genotype. In the group of the late cycle it was found the largest number of matching genotypes to the oil content, as follows: BAGSUFV131, BAGSUFV115, BAGSUFV33 and BAGSUFV53. In this group there was not common access to the protein content character. It is concluded that there is genetic diversity among the accessions evaluated of soybean, it was possible to obtain nuclear collections for each site and, the strategy adopted proved effective in the structuring of improving collections.

INTRODUÇÃO

A soja [*Glycine max* (L.) Merrill] ocupa uma posição de destaque no cenário do agronegócio mundial. A cultura é responsável por mais de treze por cento de toda produção de grãos no mundo, sendo o quarto cereal mais produzido globalmente (USDA, 2016). No Brasil, a sojicultura faz parte da principal atividade agrícola do país, movimentando mais de 27 bilhões de dólares por ano (ABIOVE, 2016). Segundo dados da Companhia Nacional de Abastecimento (CONAB, 2016), a expectativa é que na safra de 2015/2016 a produção nacional da leguminosa, alcance 100,93 milhões de toneladas, representando 47,99% das 210,3 milhões de toneladas de grãos esperadas para este ano no país. Tais previsões asseguram o Brasil como segundo maior produtor, ficando atrás apenas dos Estados Unidos que esperam uma produção de 106,95 milhões de toneladas na próxima safra (USDA, 2016).

Amplamente difundida, a soja e seus derivados são matérias primas para diferentes segmentos industriais. A busca por uma alimentação humana nutricionalmente equilibrada, visando à manutenção da saúde e à prevenção de doenças, tem elevado a procura por produtos a base de soja. Além do ótimo balanço de aminoácidos essenciais que contribuem ricamente na dieta, estudos tem revelado que a soja apresenta efeitos terapêuticos no tratamento de doenças cardiovasculares, osteoporose e câncer (VIEIRA *et al.*, 1999; PENHA *et al.*, 2007).

A alta fração proteica dos grãos, em torno de 40%, confere a soja considerável propriedades tecnológicas, por exemplo, emulsificantes e estabilizantes na indústria de alimentos (ROSA *et al.*, 2009). A boa digestibilidade das proteínas presentes no grão pode ser comparada, neste quesito, à proteínas de origem animal (CLAPPER *et al.*, 2001; DERSJANT-LI & HILL, 2005; CARCIOFI *et al.*, 2006). Assim, o farelo da soja, uma das formas de comercialização da leguminosa que apresenta alto teor de proteína, tem sido muito empregado na formulação de rações para a nutrição animal, principalmente nos setores da avicultura e suinocultura. A grande oferta do grão e o excelente equilíbrio nutricional deste alimento também contribuem para sua utilização (WILSON, 2004).

Outro produto obtido a partir do processamento da soja é o óleo. Os grãos de variedades cultivadas apresentam em média 20% de óleo em sua composição com base na matéria seca (MORAES *et al.*, 2006). Além da reconhecida aplicabilidade na indústria alimentícia, o óleo de soja faz parte da formulação de produtos cosméticos,

farmacêuticos, vernizes, tintas, lubrificantes, entre outros. Destacam-se ainda, as vantagens da adição deste produto em rações, como por exemplo, a redução na produção de calor pelas aves (BARRETO *et al.*, 2010; CHURCH & POND, 1977). Outro destino do óleo é a produção de biodiesel, que apresenta inúmeros pontos positivos, entre eles a redução na emissão de compostos poluentes e o direcionamento da matriz energética a favor de fontes renováveis (FERRARI *et al.*, 2005).

Procurando atender este diversificado mercado da soja, os programas de melhoramento têm concentrado suas pesquisas na obtenção de cultivares com características específicas para o uso, como aumento dos teores de óleo ou proteína, e que sejam produtivas. Contudo, a eficiência de qualquer programa de melhoramento está condicionada, primeiramente, à existência de variabilidade genética na população base (FERRÃO, 2009). Na soja, verifica-se elevada variabilidade para caracteres fisiológicos, morfológicos e agrônômicos, entretanto, pouco explorada pelos programas de melhoramento. O que tem sido evidenciado é uso extensivo de cultivares muito próximas geneticamente no desenvolvimento das novas cultivares, resultando no estreitamento da base genética da cultura (ALCÂNTARA NETO, 2001; BERTINI *et al.*, 2006; MIRANDA *et al.*, 2007).

Nos métodos utilizados para mensurar a diversidade genética em uma coleção germoplasmas, a caracterização morfoagronômica dos acessos por meio de variáveis quantitativas e qualitativas é a mais frequente (BUENO *et al.*, 2001; FRANCO *et al.*, 2001; GONÇALVES *et al.*, 2009; BARBÉ *et al.*, 2010).

Segundo Cruz *et al.* (2011), a existência de características de diferentes naturezas no estudo da diversidade, permite a utilização de inúmeras estratégias de análise. Uma alternativa é adotar para cada grupo de características uma medida de dissimilaridade apropriada, obtendo assim, várias matrizes de dissimilaridade de diferentes naturezas, sejam elas obtidas por dados quantitativos, qualitativos, multicategóricos ou moleculares. Outra opção é contemplar simultaneamente o conjunto de características em uma medida de dissimilaridade comum, utilizando para isso a estratégia de soma das matrizes, conforme relatado em diversos trabalhos e para diferentes culturas (RAMOS *et al.*, 2000; PEIXOTO *et al.*, 2002; MOHAMMADI & PRASANNA, 2003; ZEWDIE *et al.*, 2004; KARASAWA *et al.*, 2005; SUDRÉ *et al.*, 2006; BENTO *et al.*, 2007; BUZAR *et al.*, 2007; OLIVEIRA *et al.*, 2007; MARTINS *et al.*, 2011; SILVA *et al.*, 2013; GLASENAPP *et al.*, 2015).

Além das etapas de escolha dos acessos a serem avaliados, obtenção e sistematização dos dados e a definição da medida de similaridade ou dissimilaridade, a análise de diversidade genética é constituída da escolha do método de agrupamento (BERTAN *et al.*, 2006; CRUZ *et al.* 2012). Os métodos de agrupamento tem a finalidade de dividir uma coleção inicial de acessos em subgrupos, de modo que exista homogeneidade dentro e heterogeneidade entre os mesmos (BERTAN *et al.*, 2006; KLOSTER *et al.*, 2011). Entre os mais empregados destacam-se os métodos hierárquicos e os de otimização. Nos métodos hierárquicos os acessos são agrupados por um processo que se repete em vários níveis até a formação do dendrograma ou diagrama de árvore (CRUZ *et al.*, 2012). Entre os métodos hierárquicos cita-se o de ligação média entre grupos (Unweighteed Pair Group Method using Arithmetic Averages – UPGMA).

O processo de determinação da diversidade em uma coleção é trabalhoso, visto que, o destaque em torno da conservação dos recursos fitogenéticos tem orientado a formação e manutenção de coleções com um grande número de indivíduos. Ainda que, o elevado número de acessos seja uma forma de representar a diversidade da espécie, a acessibilidade, custos de manutenção e a própria utilidade é inversamente proporcional ao seu tamanho (FRANKEL & BENNET, 1970; FRANKEL & SOULÉ, 1981). Consequentemente, tem sido proposto o estabelecimento de coleções nucleares.

Por definição uma coleção nuclear trata-se de um conjunto de genótipos capazes de representar a máxima variabilidade genética contida em uma coleção inicial, com o mínimo de repetitividade alélica (FRANKEL, 1984; BROWN & SPILLANE, 1999). O estabelecimento dessas coleções implica em uso mais eficiente do potencial genético presente nas coleções base.

Uma etapa importante na estruturação das coleções nucleares é a avaliação do quanto da população original, em relação à média e à variância, foi retido no conjunto de genótipos amostrado para cada caráter estudado (CRUZ *et al.*, 2012). Para tanto, comparação entre as variâncias de cada coleção são realizadas empregando o teste F. Já a semelhança ou diferença entre as médias são examinadas pelo teste t de Student ou ainda pelo Índice de coincidência da amplitude (CA), também conhecido como índice de similaridade ou índice de retenção da variabilidade (IRV) (SNEDECOR & COCHRAN, 1980; HOLBROOK *et al.*, 1993; HU *et al.*, 2000; CRUZ, 2006; WANG *et al.*, 2007; VASCONCELOS *et al.*, 2007; DIWAN *et al.*, 1995).

Outra coleção de grande importância em um programa de melhoramento é a coleção de trabalho ou de melhoramento. Estas coleções de tamanho limitado geralmente são compostas por germoplasmas elites que apresentam altas médias, no entanto, frequentemente semelhantes geneticamente. Em virtude da exploração apenas de sua função primária de disponibilizar acessos para os cruzamentos, quase não há relatos de estratégias que definam a estruturação de uma coleção de trabalho com variadas funções e com a presença de diferentes genes para caracteres de interesse, como por exemplo, teor de óleo e de proteína.

Assim, este trabalho teve como objetivos avaliar a diversidade genética entre acessos de soja por meio de caracteres morfoagronômicos, estruturar coleções nucleares e estabelecer estratégia para extração de coleções de melhoramento para teores de óleo e de proteína.

MATERIAL E MÉTODOS

Para realização do trabalho foram avaliados 161 acessos de soja provenientes do Banco Ativo de Germoplasma de Soja (BAG) do Departamento de Fitotecnia (DFT).

A condução dos ensaios ocorreu no ano agrícola de 2014/2015, nas unidades experimentais pertencentes à Universidade Federal de Viçosa: Diogo Alves de Mello (latitude 20°45'14''S, longitude 42°52'55''W e altitude de 648 metros) em Viçosa, Minas Gerais, denominada de local 1; e, Fazenda Experimental de Coimbra (latitude 20° 51'24''S, longitude 42°48'10''W e altitude de 720 metros) em Coimbra, Minas Gerais, denominada de local 2.

O delineamento experimental utilizado foi o de blocos aumentados (DBA) com sete repetições, sete acessos comuns e 154 não-comuns. A escolha do delineamento se deu em função da baixa disponibilidade de sementes e do grande número de acessos a serem avaliados.

O plantio foi realizado conforme as práticas culturais para soja adotada em cada localidade. No local 1, adotou-se o sistema convencional de plantio e as parcelas experimentais foram compostas por uma fileira de um metro de comprimento, espaçadas em 0,70 metros. Já no experimento conduzido no local 2, foi implantado o sistema de plantio direto e as parcelas apresentaram dimensões de um metro linear, espaçadas em 0,45 metros. A densidade de plantio foi de 222 mil plantas/ha para os dois locais.

Foram realizadas todas as operações de manejo conforme as exigências da cultura. O controle de pragas, doenças e plantas daninhas foi feito de forma preventiva e à medida que se fizeram necessário.

Para a caracterização e avaliação dos acessos foram utilizados alguns descritores morfoagronômicos, proposto pela Instrução para Execução dos Ensaio de Distingüibilidade, Homogeneidade e Estabilidade de Cultivares de Soja do Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2002) e caracteres relacionados com a produção e qualidade nutricional, considerados de relevância para a determinação da diversidade genética entre os genótipos.

A caracterização e avaliação dos acessos se resumem na determinação de sete variáveis qualitativas e quatorze quantitativas, avaliadas em campo e laboratório através de dados médios de três plantas representativas dentro da parcela, selecionadas quatorze dias após a emergência das mesmas.

Caracteres qualitativos avaliados em campo:

- a) Cor da flor (CF) - coloração branca ou roxa das pétalas, por ocasião do florescimento;
- b) Tipo de inflorescência (TI) - racimo pedunculado ou séssil nas axilas das folhas e na extremidade do caule;
- c) Formato do folíolo (FF) - formato do folíolo lateral (lanceolado, lanceolado estreito, oval-pontiagudo e oval-arredondado).

Caracteres qualitativos avaliados em laboratório:

- a) Cor da pubescência (CP) - coloração marrom ou cinza dos pelos na haste principal e vagens;
- b) Densidade da pubescência (DP) - concentração alta, média ou baixa de pelos na haste principal e vagens;
- c) Tipo de crescimento (TC) - tipo determinado, semideterminado ou indeterminado de crescimento dos genótipos.
- d) Habito de crescimento (HC) - porte ereto, semiereto ou prostrado, com base na categorização do ângulo de abertura das hastes laterais em relação à haste principal, sendo: ângulo menor que 30° a planta foi caracterizada como ereta; ângulo entre 30° e 60° a planta foi caracterizada como semiereta; e, ângulo maior que 60° a planta caracterizada como prostrada.

Caráter quantitativo avaliado em campo:

- a) Ângulo de acamamento (AC) - ângulo médio formado entre a haste principal das plantas e a superfície do solo, avaliado no momento da colheita (estádio R8), utilizando equipamento com ângulo de 90° dividido em cinco partes iguais (Figura 1).



Figura 1. Equipamento para avaliação do ângulo de acamamento.

Caracteres quantitativos avaliados em laboratório:

- a) Número de dias para o florescimento (NDF) - número de dias contados desde a emergência até o florescimento de mais de 50 % das plantas na parcela;
- b) Número de dias para a maturação (NDM) - número de dias contados desde a emergência até quando mais de 50% das plantas na parcela atingiram o estágio R8;
- c) Diâmetro do hipocótilo (DH) - medida, em centímetro, na região compreendida entre o ponto de inserção dos cotilédones e o início da raiz;

- d) Altura da planta na maturação (APM) - Distância, em centímetro, da marca da superfície do solo até a extremidade da haste principal das plantas no estádio R8;
- e) Altura de inserção da primeira vagem (AIV) - Distância, em centímetro, da marca da superfície do solo à primeira vagem das plantas no estádio R8;
- f) Número de nós na haste principal (NNHP) - número de nós contados a partir da marca da folha unifoliolada até o último nó da haste principal das plantas no estádio R8;
- g) Número de hastes laterais (NHL) - número de hastes laterais contidas em toda extensão da haste principal das plantas no estádio R8;
- h) Número de vagens por planta (NVP) - número médio de vagens por planta, obtido após a colheita;
- i) Número de sementes por vagem (NSV) - número médio de sementes por vagem, obtido após a colheita;
- j) Massa de cem sementes (MCS) - massa, em gramas, de uma amostra de 100 sementes da parcela, obtido após a colheita;
- k) Produção (PROD) - produção, em gramas por planta, baseada na massa total de sementes de cada planta da parcela;
- l) Teor de óleo (TO) - porcentagem de óleo presente no grão, obtido via Ressonância Magnética Nuclear (NMR - Nuclear Magnetic Resonance MQC 2327) utilizando o aparelho da marca Oxford Instruments;
- m) Teor de proteína (TP) - determinação da porcentagem de proteína presente nos grãos utilizando o método de Kjeldahl, conforme recomendação da Association of Official Analytical Chemist (1975) e Vitt *et al.* (2001). Neste método, a porcentagem de proteína é determinada quantificando a concentração de nitrogênio total e multiplicando o valor pelo fator de correção 6,25.

Os dados quantitativos foram submetidos à análise de variância separadamente para cada local. A finalidade deste procedimento foi verificar a existência de variabilidade genética entre os acessos avaliados. A opção pela análise de variância individual ocorreu mediante a hipótese da redução dos efeitos da interação genótipo x ambiente na detecção de variabilidade genética para as características avaliadas.

O modelo estatístico adotado foi:

$$Y_{ij} = \mu + T_i + T_g + B_i + \varepsilon_{ij}$$

em que:

Y_{ij} : valor de cada caráter para i-ésimo genótipo no j-ésimo bloco

μ : média geral

T_i : efeito do i-ésimo genótipo comuns ou testemunha

T_g : efeito do i-ésimo genótipo não-comuns

B_i : efeito do j-ésimo bloco

ε_{ij} : erro experimental

O agrupamento pelo método da ligação média entre grupos (UPGMA) foi utilizado para expressar a diversidade genética existente entre os acessos avaliados. Para isso, foi utilizada como medida de dissimilaridade a distância Euclidiana média.

Em função das diferentes escalas de medidas dos caracteres quantitativos, as estimativas da dissimilaridade foram realizadas a partir de dados padronizados através do estimador:

$$y_{ij} = \frac{Y_{ij}}{S(Y_j)}$$

em que:

Y_{ij} : observação no i-ésimo genótipo ($i = 1, 2, \dots, p$), em relação ao j-ésimo ($j=1, 2, \dots, n$) caracter;

$S(Y_j)$: desvio padrão do j-ésimo caracter.

Já para os caracteres qualitativos, o procedimento utilizado para a obtenção da distância entre as cultivares foi à conversão dos resultados em sistema binário, de forma que a presença da categoria para cada característica recebeu o valor 1, e a ausência, o valor 0.

Para ambos os grupos de caracteres, o cálculo da distância Euclidiana se deu por meio do estimador:

$$d_{ii'} = \sqrt{\frac{1}{n} \sum_j (y_{ij} - y_{i'j})^2}$$

em que:

$d_{ii'}$: distância Euclidiana média entre o i -ésimo genótipo e o i' -ésimo genótipo, sendo $i \neq i'$;

n : número de caracteres estudados;

y_{ij} ou $y_{i'j}$: observação no i -ésimo genótipo ($i = 1, 2, \dots, p$) ou i' -ésimo genótipo ($i' = 1, 2, \dots, p$), em relação ao j -ésimo ($j=1, 2, \dots, n$) caracter, sendo $i \neq i'$.

Estimado as matrizes de dissimilaridade individualmente para cada conjunto de caracteres (quantitativo e qualitativo) e locais, foram identificados os cinco pares de acessos mais dissimilares e os cinco mais similares em cada associação, sendo estabelecida apenas uma associação entre caracteres qualitativos e locais, por serem consideradas como características não influenciadas pelo ambiente. Também foram identificados os acessos que se mostraram mais divergentes e mais similares em relação a todos por meio do somatório das distâncias Euclidianas de cada possível combinação referente a cada acesso.

A fim de verificar a existência de relação da diversidade obtida para cada conjunto de caracteres, tendo como finalidade a soma algébrica das distâncias Euclidianas, o teste Z de Mantel foi utilizado na comparação das matrizes de dissimilaridade. Este teste permite comparar duas matrizes entre si, elemento por elemento, fornecendo um valor de correlação que quantifica o grau de relacionamento entre as matrizes, sendo testada a significância de r por meio de permutações (MANLY, 1997; MARTINS *et al.* 2011).

A matriz de dissimilaridade originada da soma algébrica das matrizes individuais (SOMA) foi empregada no procedimento de agrupamento dos acessos pelo método UPGMA. Neste procedimento a construção do dendrograma foi estabelecida pelos indivíduos com maior semelhança.

Após a análise de diversidade foram estabelecidas coleções nucleares para cada local, utilizando para isso apenas os caracteres quantitativos devido à dificuldade de integração dos dados qualitativos no estabelecimento das coleções.

Neste sentido, inicialmente, a estratégia foi determinar um número de indivíduos que representasse a média e a variância de todas as características quantitativas avaliadas, calculada na população original de cada local. Este procedimento foi realizado através da análise gráfica de dispersão da média e da variância, encontradas em amostras com diferentes números de acessos. Os resultados gráficos foram

produzidos através de 50 simulações de cada amostra, geradas inicialmente, a partir do conjunto de cinco acessos até o conjunto de 160 acessos, sendo o incremento entre os conjuntos de cinco acessos.

O número de acessos que seriam amostrados para formação da coleção nuclear foi definido quando os valores estimados para média e variância nas amostras tomadas ao acaso para cada conjunto de acessos, encontraram-se dentro do intervalo de confiança ($\alpha = 0,05$), para a maioria dos caracteres. Uma vez definido o número de acessos para a formação da coleção nuclear para cada local, os acessos foram tomados ao acaso.

A fim de avaliar a adequação de cada coleção nuclear, ou seja, sua representatividade em relação ao restante das variáveis quantitativas avaliadas, as amostras foram comparadas com a coleção original da qual foram extraídas. Foi estudado o impacto das coleções nucleares na recuperação da variância e da média da coleção inicial.

Na comparação entre as variâncias para cada caractere foi empregado o teste F (SNEDECOR & COCHRAN, 1980) dado pelo estimador:

$$F_{calculado} = \frac{s_x^2}{s_y^2}$$

Sendo s_x^2 a maior variância, para o caractere em questão, entre a estimada para a população original e a estimada para a coleção nuclear. E s_y^2 a de menor valor estimado. Deste modo obtêm-se sempre o F calculado maior que 1. Os valores de F calculado foram comparados com o valor de F tabelado ao nível 5% e 1% de probabilidade, associado aos graus de liberdade de cada variância.

Na comparação das médias foi utilizado o teste t de Student e o índice de coincidência da amplitude (CA).

Pelo teste t de Student em caso de não significância na comparação das variâncias pelo teste f, descrito anteriormente, adotou-se a variância amostral comum obtida pelo estimador:

$$s_c^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2},$$

em que:

s_c^2 = variância amostral comum;

s_x^2 = variância da população original;

s_y^2 = variância da coleção nuclear;

n_x = número de genótipos da população original;

n_y = número de genótipos da coleção nuclear.

Utilizando o valor da variância amostral, a estimativa do t calculado foi determinada pelo estimador:

$$t_{calculado} = \frac{\bar{X} - \bar{Y}}{\sqrt{s_c^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$$

em que:

\bar{X} = média da população original;

\bar{Y} = média da coleção nuclear.

Para as variâncias que se mostraram heterogêneas, o t calculado foi obtido pelo estimador:

$$t_{calculado} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

O índice de coincidência da amplitude, citado por Hu *et al.* (2000) e Wang *et al.* (2007), para cada coleção nuclear foi calculado por meio do estimador:

$$CA = \frac{1}{n} \sum_{i=1}^n \frac{A_i CN}{A_i CI}$$

em que:

CA = índice de coincidência da amplitude;

$A_i CN$ = amplitude do i-ésimo caractere na coleção nuclear;

A_iCI = amplitude do i-ésimo caractere na população original;

n = número de caracteres avaliados.

A conservação nas coleções nucleares da variabilidade para as características qualitativas foi avaliada por meio da estimativa, em porcentagem, da manutenção de cada classe, dentro das mesmas.

Definida as coleções nucleares, a próxima etapa no desenvolvimento do trabalho foi estabelecer coleções de melhoramento.

A estruturação de coleções de melhoramento, com ênfase em teor de óleo e teor de proteína, foi realizada seguindo a estratégia: 1- estratificação dos acessos em relação ao ciclo total, em precoce (até 120 dias de ciclo), médio (de 121 a 140 dias de ciclo) e tardio (com mais de 140 dias de ciclo); 2- seleção de acessos, dentro de cada estrato de ciclo, com alta média para cada caractere em questão e divergentes.

Na estratificação dos acessos, para cada grupo formado em relação ao ciclo, foram selecionados 15 acessos em cada ensaio, com as maiores médias para o teor de óleo e proteína. Dessa forma, foram obtidos doze conjuntos de acessos.

Os conjuntos pertencentes ao mesmo caractere, dentro do mesmo grupo de ciclo e de locais diferentes, foram contrastados a fim de identificar acessos coincidentes. O objetivo foi selecionar 20 acessos distintos, sendo 10 de cada ensaio, para compor uma coleção de melhoramento única. Diante das possíveis semelhanças de acessos selecionados em ambos locais adotou-se a estratégia de inclusão da quantidade de acessos quanto forem o número de semelhanças. Os acessos selecionados foram aqueles de maiores médias para a característica em questão, dentro do mesmo ciclo de maturação e independentemente do local, e que não pertenciam ao grupo dos dez melhores acessos de cada local.

Os mesmos procedimentos foram realizados para os demais pares de conjuntos. Assim, foram estabelecidas três coleções de melhoramento para cada uma das variáveis.

Todas as análises estatísticas foram realizadas com o auxílio do Programa Computacional Genes (CRUZ, 2013).

RESULTADOS E DISCUSSÃO

O resumo das análises de variância individuais e as estimativas dos parâmetros genéticos e ambientais para os 161 acessos avaliados por meio dos caracteres quantitativos encontram-se nas tabelas 1 e 2.

Na tabela 1, foi possível verificar variância genética significativa, ao nível de 5% de probabilidade pelo teste F, para os caracteres número de dias para floração, número de dias para maturidade, altura da planta na maturidade, altura de inserção da primeira vagem, número de nós da haste principal, número de hastes laterais, número de sementes por vagem e teor de óleo.

Analisando os resultados para o local 2 em relação ao local 1, a variância genética significativa também foi verificada para os caracteres diâmetro do hipocótilo e teor de proteína, não sendo diagnosticada para o número de dias para floração e número de dias para maturidade (Tabela 2).

A variância genética foi estatisticamente igual a zero, em ambos os locais, para os caracteres ângulo de acamamento, número de vagens por planta, massa de cem sementes e produção em gramas por planta.

A significância do efeito de genótipos é um indicativo de que as constituições genéticas são divergentes para os caracteres morfoagronômicos avaliados. Segundo Bertan et al. (2006), a significância para esta fonte de variação é fator essencial para a realização de estudos de diversidade genética.

Os valores das variâncias fenotípicas (σ^2_f) observadas entre os acessos para os caracteres NDF, NDM, APM, AIV, NNHP, NHL e TO, no local 1, possuem predominantemente influência das variâncias genéticas (σ^2_g) em detrimento às variâncias ambientais (σ^2_e). Isso também foi verificado para os caracteres DH, APM, AIV, NHL, TO e TP, no local 2. Assim, caracteres que apresentaram mais influência das variâncias ambientais (σ^2_e) no cômputo da variância fenotípica (σ^2_f), refletiram baixo valor de herdabilidade.

Os valores de herdabilidade estimados para cada caractere no local 1 variaram de 1,99 a 90,21, com destaque para o NDF, NDM, APM e NNHP que registraram valores superiores a 80. No local 2 os caracteres NNHP, AC e MCS apresentaram variância genotípica nula, conseqüentemente, os valores de herdabilidade, bem como os de coeficiente de variação genotípica (CVg), foram iguais a 0. Neste mesmo local a maior herdabilidade foi observada para a APM (77,49).

Tabela 1. Resumo da análise de variância e estimativa de parâmetros genéticos e ambientais para os caracteres número de dias para floração (NDF); número de dias para maturidade (NDM); diâmetro do hipocótilo (DH); altura da planta na maturidade (APM); altura de inserção da primeira vagem (AIV); número de nós da haste principal (NNHP); número de hastes laterais (NHL); ângulo de acamamento (AC); número de vagens por planta (NVP); número de sementes por vagem (NSV); massa de cem sementes (MCS); produção em g/planta (PROD); teor de óleo (TO) e teor de proteína (TP) avaliados em 161 genótipos de soja na Unidade Experimental Diogo Alves de Mello, em Viçosa, Minas Gerais.

FV	GI	Quadrado médio													
		NDF	NDM	DH	APM	AIV	NNHP	NHL	AC	NVP	NSV	MCS	PROD	TO	TP
Bloco	6	166,83	416,95	2,47	1263,88	89,11	10,91	9,22	1210,63	5000,5	0,15	32,37	522,94	2,64	120,42
Genótipo	160	108,41**	231,46**	7,57 ^{ns}	670,71**	23,34**	15,18**	9,46**	433,49 ^{ns}	4833,48 ^{ns}	0,07*	17,13 ^{ns}	314,92 ^{ns}	2,82**	13,10 ^{ns}
Resíduo	36	12,89	21,34	4,96	114,27	9,08	2,95	4,04	440,08	5699,19	0,04	14,5	312,24	1,26	9,04
Parâmetros ¹															
Média comuns		67,71	152,22	14,21	85,30	14,94	18,57	8,45	59,82	198,42	1,99	12,70	49,61	19,35	44,00
Média não comuns		61,58	142,87	12,57	80,01	14,32	17,68	8,07	61,91	149,63	2,11	13,67	41,98	19,66	45,19
Média ajustada		61,90	143,35	12,55	79,68	14,31	17,58	8,00	62,69	148,74	2,10	13,74	41,91	19,57	45,53
CV% geral		5,69	3,18	17,18	13,15	20,82	9,59	24,64	34,16	46,77	9,72	28,34	40,32	5,73	6,69
σ^2_f		103,30	217,87	6,83	735,44	25,18	17,24	10,88	502,50	5815,04	0,08	16,42	382,04	3,15	15,00
σ^2_e		12,89	21,34	4,96	114,27	9,08	2,95	4,04	440,08	5699,19	0,04	14,50	312,24	1,26	9,04
σ^2_g		90,40	196,53	1,87	621,17	16,11	14,30	6,84	62,42	115,85	0,04	1,92	69,80	1,89	5,96
h^2		87,52	90,21	27,33	84,46	63,95	82,91	62,88	12,42	1,99	46,71	11,71	18,27	60,08	39,74
CVg		15,44	9,81	10,87	31,15	28,02	21,39	32,43	12,76	7,19	8,98	10,14	19,90	7,00	5,40
CVg/CV geral		2,71	3,08	0,63	2,37	1,35	2,23	1,32	0,37	0,15	0,92	0,36	0,49	1,22	0,81

*,** Significativo a 5% e 1% de probabilidade, respectivamente, pelo teste F.

¹ média dos genótipos comuns (Média comuns), média dos genótipos não comuns (Média não comuns), média ponderada de Federer (Média ponderada), coeficiente de variação ambiental geral (CV%), variância fenotípica (σ^2_f), variância ambiental (σ^2_e), variância genotípica (σ^2_g), herdabilidade (h^2), coeficiente de variação genotípica (CVg) e razão coeficiente de variação genotípica/coeficiente de variação geral (CVg/CV geral).

Tabela 2. Resumo da análise de variância e estimativa de parâmetros genéticos e ambientais para os caracteres número de dias para floração (NDF); número de dias para maturidade (NDM); diâmetro do hipocótilo (DH); altura da planta na maturidade (APM); altura de inserção da primeira vagem (AIV); número de nós da haste principal (NNHP); número de hastes laterais (NHL); ângulo de acamamento (AC); número de vagens por planta (NVP); número de sementes por vagem (NSV); massa de cem sementes (MCS); produção em g/planta (PROD); teor de óleo (TO) e teor de proteína (TP) avaliados em 161 genótipos de soja na Unidade Experimental Fazenda Experimental de Coimbra, em Coimbra, Minas Gerais.

FV	GI	Quadrado médio													
		NDF	NDM	DH	APM	AIV	NNHP	NHL	AC	NVP	NSV	MCS	PROD	TO	TP
Bloco	6	370,5	179,61	5,84	3079,92	112,13	31,24	8,57	46,98	2664,34	0,06	38,12	173,6	1,49	69,64
Genótipo	160	112,02 ^{ns}	137,15 ^{ns}	7,06 ^{**}	555,43 ^{**}	49,05 ^{**}	26,41 ^{ns}	6,24 ^{**}	70,13 ^{ns}	1809,99 ^{ns}	0,06 [*]	17,93 ^{ns}	137,37 ^{ns}	2,49 ^{**}	13,91 ^{**}
Resíduo	36	99,36	148,87	2	129,73	15,59	93,28	2,56	129,31	1928,06	0,03	21,57	124,38	1,07	5,99
Parâmetros ¹															
Média comuns		54,04	137,33	11,68	100,02	18,19	21,99	6,32	77,69	141,63	2,11	11,77	34,13	21,41	40,39
Média não comuns		53,63	134,80	11,26	89,12	18,69	18,77	6,21	78,77	114,07	2,06	14,30	32,14	21,08	42,04
Média ajustada		53,93	134,85	11,22	89,65	18,61	18,70	6,20	78,48	113,87	2,06	14,08	32,04	21,09	41,98
CV% geral		18,55	9,01	12,46	12,41	21,26	49,42	25,63	14,48	36,37	8,32	33,94	34,19	4,88	5,88
σ^2_f		141,36	167,29	7,37	576,35	51,61	33,92	6,52	80,82	2074,57	0,06	18,96	150,02	2,55	13,44
σ^2_e		99,36	148,87	2,00	129,73	15,59	93,28	2,56	129,31	1928,06	0,03	21,57	124,38	1,07	5,99
σ^2_g		42,00	18,43	5,37	446,62	36,03	0,00	3,96	0,00	146,50	0,03	0,00	25,64	1,48	7,46
h^2		29,71	11,01	72,82	77,49	69,80	0,00	60,77	0,00	7,06	48,58	0,00	17,09	58,16	55,46
CVg		12,09	3,18	20,58	23,71	32,11	0,00	32,04	0,00	10,61	8,14	0,00	15,76	5,77	6,49
CVg/CV geral		0,65	0,35	1,65	1,91	1,51	0,00	1,25	0,00	0,29	0,98	0,00	0,46	1,18	1,11

*,** Significativo a 5% e 1% de probabilidade, respectivamente, pelo teste F.

¹ média dos genótipos comuns (Média comuns), média dos genótipos não comuns (Média não comuns), média ponderada de Federer (Média ponderada), coeficiente de variação ambiental geral (CV%), variância fenotípica (σ^2_f), variância ambiental (σ^2_e), variância genotípica (σ^2_g), herdabilidade (h^2), coeficiente de variação genotípica (CVg) e razão coeficiente de variação genotípica/coeficiente de variação geral (CVg/CV geral).

Quanto aos coeficientes de variação genotípica (CVg), com exceção dos caracteres NDM, NVP, NSV, TO e TP, no local 1 e, NDM, NNHP, AC, NSV, MCS, TO e TP, no local 2, o valores encontrados foram superior a 10%. Segundo Rezende (2002), coeficiente de variação genotípica acima de 10% é suficiente para proporcionar ganhos genéticos com a seleção.

As relações CVg/CVgeral foram maiores que 1 para os caracteres NDF, NDM, APM, AIV, NNHP,NHL e TO, no local 1 e, DH, AIV, NHL, TO e TP, no local 2. De acordo com Vencovsky (1987), a relação CVg/CVgeral, juntamente com a herdabilidade, caracterizam importantes informações a respeito do entendimento da variação existente entre os acessos, sendo indicativos de ganhos genéticos. Assim, quando a razão CVg/CVe é maior que 1,0 pode presumir que as condições à seleção são favoráveis, já que a variação genética supera a ambiental.

Em relação à precisão experimental, no local 1, os coeficientes de variação (CV% geral) oscilaram de 3,18% (NDM) à 46,77% (NVP), tendo os caracteres AC (34,16%), PROD (40,32%) e NVP (46,77%) apresentado os maiores valores. No local 2 a variação do CVgeral foi de 4,88% (TO) à 49,42% (NNHP), sendo observado os valores mais elevados para os caracteres MCS (33,94%), PROD (34,19%), NVP (36,37%) e NNHP (49,42%). Os altos coeficientes de variação encontrados em cada local, para os caracteres descritos, são comumente relatados para a cultura da soja (LAÍNEZ-MEJIA, 1996; ROSSMANN, 2001; FILHO *et al.*, 2009; ALMEIDA *et al.* 2011; NOGUEIRA *et al.*, 2012; MARCONATO, 2014; TEODORO *et al.*, 2015). Esta medida de precisão apresenta flexibilidade e segundo Steel *et al.* (1997), a definição do valor de CV como alto ou baixo para um determinado caractere, depende do conhecimento e experiência do pesquisador com a espécie em estudo e com resultados de experimentos semelhantes.

A distância Euclidiana média (dii') entre os genótipos foram calculadas por conjuntos de variáveis, quantitativos e qualitativos, e por local, gerando matrizes de dissimilaridade quantitativa para cada local e uma qualitativa comum aos dois locais.

Na matriz de dissimilaridade para os dados quantitativos, a distância Euclidiana entre os acessos variou de 0,3753 a 3,2529 no local 1, e de 0,3679 a 3,4008 no local 2. Para este mesmo grupo de variáveis foram identificados os cinco pares de acessos mais divergentes e os cinco mais similares em cada local (Tabela 3).

Tabela 3. Cinco pares de acessos mais divergentes e os cinco mais similares com base na distância Euclidiana média entre 161 acessos avaliados em dois locais em Minas Gerais para o grupo de caracteres quantitativos.

Local	Pares de acessos mais dissimilares	Distância Euclidiana	Pares de acessos mais similares	Distância Euclidiana
Local 1	BAGSUFV106 - BAGSUFV36	3,2529	BAGSUFV157 - BAGSUFV159	0,3753
	BAGSUFV124 - BAGSUFV63	3,0657	BAGSUFV9 - BAGSUFV12	0,3860
	BAGSUFV106 - BAGSUFV45	3,0021	BAGSUFV43 - BAGSUFV44	0,3886
	BAGSUFV36 - BAGSUFV18	2,9761	BAGSUFV56 - BAGSUFV157	0,3922
	BAGSUFV36 - BAGSUFV23	2,9512	BAGSUFV37 -BAGSUFV116	0,3930
Local 2	BAGSUFV48 - BAGSUFV16	3,4008	BAGSUFV105 - BAGSUFV112	0,3679
	BAGSUFV117 - BAGSUFV16	3,3613	BAGSUFV1 - BAGSUFV157	0,3914
	BAGSUFV48 - BAGSUFV19	3,2178	BAGSUFV129 - BAGSUFV159	0,3976
	BAGSUFV81 - BAGSUFV16	3,1675	BAGSUFV126 - BAGSUFV132	0,3999
	BAGSUFV59 - BAGSUFV16	3,1536	BAGSUFV82 - BAGSUFV126	0,4211

No local 1 a maior distância ($d_{ii}' = 3,2529$) foi verificada entre os acessos BAGSUFV106 e BAGSUFV36, conseqüentemente, os mais divergentes. Já os mais similares foram os acessos BAGSUFV157 e BAGSUFV159 ($d_{ii}' = 0,3753$). Para este local os acessos que se mostram mais divergentes e mais similares em relação a todos foi o BAGSUFV36 e BAGSUFV157, respectivamente.

No local 2, o par de acessos BAGSUFV48 e BAGSUFV16 ($d_{ii}' = 3,4008$) foi o mais divergente. A menor distância foi identificada entre os acessos BAGSUFV59 e BAGSUFV16 ($d_{ii}' = 0,4211$). Os acessos BAGSUFV16 e BAGSUFV155 foram os mais divergentes e os mais similares, respectivamente, em relação a todos.

As medidas de dissimilaridade determinadas através dos dados qualitativos transformados oscilaram entre 7,9876 e 298,0604 (Tabela 4). Para este conjunto de caracteres, os pares de acessos mais divergentes e similares foram os BAGSUFV38/BAGSUFV75 ($d_{ii}' = 298,0604$) e BAGSUFV2/BAGSUFV22 ($d_{ii}' = 7,9876$), respectivamente. Para este conjunto de caracteres os acessos BAGSUFV75 e BAGSUFV9 foram os mais divergentes e os mais similares em relação a todos, respectivamente.

Tabela 4. Cinco pares de acessos mais divergentes e os cinco mais similares obtidos após a determinação da distância Euclidiana média entre 161 acessos avaliados em Minas Gerais para o grupo de caracteres qualitativos.

Local	Pares de acessos mais dissimilares	Distância Euclidiana	Pares de acessos mais similares	Distância Euclidiana
Local 1 e 2	BAGSUFV38 - BAGSUFV75	298,0604	BAGSUFV2 - BAGSUFV22	7,9876
	BAGSUFV1 - BAGSUFV75	282,9365	BAGSUFV2 - BAGSUFV65	8,0879
	BAGSUFV125 - BAGSUFV75	276,7355	BAGSUFV2 - BAGSUFV71	10,0078
	BAGSUFV128 - BAGSUFV75	268,2762	BAGSUFV2 - BAGSUFV3	10,9357
	BAGSUFV53 - BAGSUFV75	244,2744	BAGSUFV6 - BAGSUFV127	11,7199

Os valores de dissimilaridade, verificados tanto para o conjunto de caracteres quantitativos, quanto os qualitativos, podem ser considerados um indicativo de ampla variabilidade genética entre os acessos. Vários autores relatam a presença diversidade genética entre acessos de soja associados à ocorrência de elevada magnitude nas medidas de dissimilaridade (BROWN-GUERIDA *et al.*, 2000; SIHAG *et al.*, 2004; CHETTRI *et al.*, 2005; MALIK *et al.*, 2007; PELUZIO *et al.*, 2009; VIEIRA *et al.*, 2009; MULATO *et al.*, 2010; ALMEIDA *et al.*, 2011; RIGON *et al.*, 2012; FERREIRA JÚNIOR *et al.*, 2015 e TORRES *et al.*, 2015).

Os valores dos coeficientes de correlação (r) estimados para cada par de matrizes de dissimilaridades e suas respectivas significâncias pelo teste Z de Mantel são apresentadas na Tabela 5.

Tabela 5. Correlação entre as matrizes de dissimilaridade de variáveis quantitativas do local 1 (MD Quant. local 1), quantitativas do local 2 (MD Quant. local 2), qualitativas (MD Quali. Geral) e a matriz SOMA, pelo teste Z de Mantel.

	MD Quant. Local 1	MD Quant. Local 2	MD Quali. Geral	MD SOMA
MD Quant. Local 1	1	0,3025**	- 0,08857 ^{ns}	0,72**
MD Quant. Local 2		1	- 0,05938 ^{ns}	0,71**
MD Quali. Geral			1	0,47**
MD SOMA				1

** significativo a 1% de probabilidade, pelo teste Z de Mantel baseado em 1000 simulações.

Os valores de correlação estimados entre a matriz SOMA e as matrizes de dissimilaridade referentes a cada conjunto de características foram significativas a 1% de probabilidade pelo teste Z de Mantel. No entanto, foi verificada baixa correlação (0,47) entre a matriz de dissimilaridade obtida por meio dos dados qualitativos e a matriz SOMA. Já os coeficientes de correlação entre os pares de matrizes SOMA e as matrizes referentes aos dados quantitativos de cada ensaio, apresentaram valores superiores a 0,7.

De acordo com Tatineni *et al.* (1996), valores de correlação são estatisticamente significativos, se r é superior a 0,5 quando o número de unidades comparadas for maior que 15. Segundo Rohlf (2000), valores de correlação inferiores a 0,7 têm sido considerados inconsistentes, o que justificaria a inadequação da matriz SOMA na representatividade da matriz de dissimilaridade determinada pelas variáveis qualitativas. Entretanto, os autores Tsivelikas *et al.* (2009) afirmam que a análise de diversidade, considerando simultaneamente variáveis quantitativas e qualitativas, pode fornecer uma melhor compreensão da diversidade genética em um conjunto de acessos. Nesse sentido, foi utilizado neste trabalho a matriz SOMA para a realização do agrupamento via método UPGMA.

O dendrograma resultante da análise de 161 acessos de soja, obtido pelo método de agrupamento UPGMA está representado na Figura 2.

Para a definição dos grupos formados por meio do dendrograma, foi utilizado um exame visual e definiu-se a região de corte onde havia pontos de alta mudança de nível.

O processo de agrupamento resultou na formação de 57 grupos distintos, sendo trinta destes, composto por apenas um acesso (BAGSUFV75, BAGSUFV45, BAGSUFV36, BAGSUFV46, BAGSUFV161, BAGSUFV38, BAGSUFV62, BAGSUFV117, BAGSUFV106, BAGSUFV63, BAGSUFV122, BAGSUFV128, BAGSUFV14, BAGSUFV19, BAGSUFV40, BAGSUFV55, BAGSUFV137, BAGSUFV39, BAGSUFV13, BAGSUFV110, BAGSUFV48, BAGSUFV32, BAGSUFV109, BAGSUFV82, BAGSUFV152, BAGSUFV77, BAGSUFV73, BAGSUFV98, BAGSUFV51 e BAGSUFV25).

Os acessos BAGSUFV12, BAGSUFV31, BAGSUFV33, BAGSUFV34, BAGSUFV37, BAGSUFV47, BAGSUFV58, BAGSUFV60, BAGSUFV61, BAGSUFV64, BAGSUFV69, BAGSUFV83, BAGSUFV92, BAGSUFV94, BAGSUFV95, BAGSUFV96, BAGSUFV102, BAGSUFV111, BAGSUFV113, BAGSUFV115, BAGSUFV119, BAGSUFV127, BAGSUFV130, BAGSUFV132,

BAGSUFV134, BAGSUFV155, BAGSUFV157 e BAGSUFV158, foram alocados no grupo 44, estabelecendo o maior conjunto. O par de acessos mais similares BAGSUFV9/BAGSUFV12 foi reunido no grupo 28. Já os acessos mais dissimilares BAGSUFV75 e BAGSUFV48 integraram os grupos 1 e 27, respectivamente.

Val Pedroso *et al.* (2014), avaliando a diversidade genética de trinta genótipos de soja por meio de caracteres agromorfológicos distinguiu seis grupos pelo método UPGMA. Vieira *et al.* (2009), utilizando o mesmo método de agrupamento, obteve a formação de sete grupos na análise da variabilidade genética em cinquenta e três cultivares de soja determinada por meio marcadores microsatélites. Brown-Guerida *et al.* (2000), obtiveram resultados semelhante e com a alocação de cento e quinze acessos em onze grupos.

A diferença elevada no número de grupos formados está associada a vários fatores como: número inicial de acessos, características consideradas na avaliação, acessos que compunham coleção inicial caracterizando a própria diversidade e a escolha do ponto de corte, sendo este dependente da experiência do pesquisador.

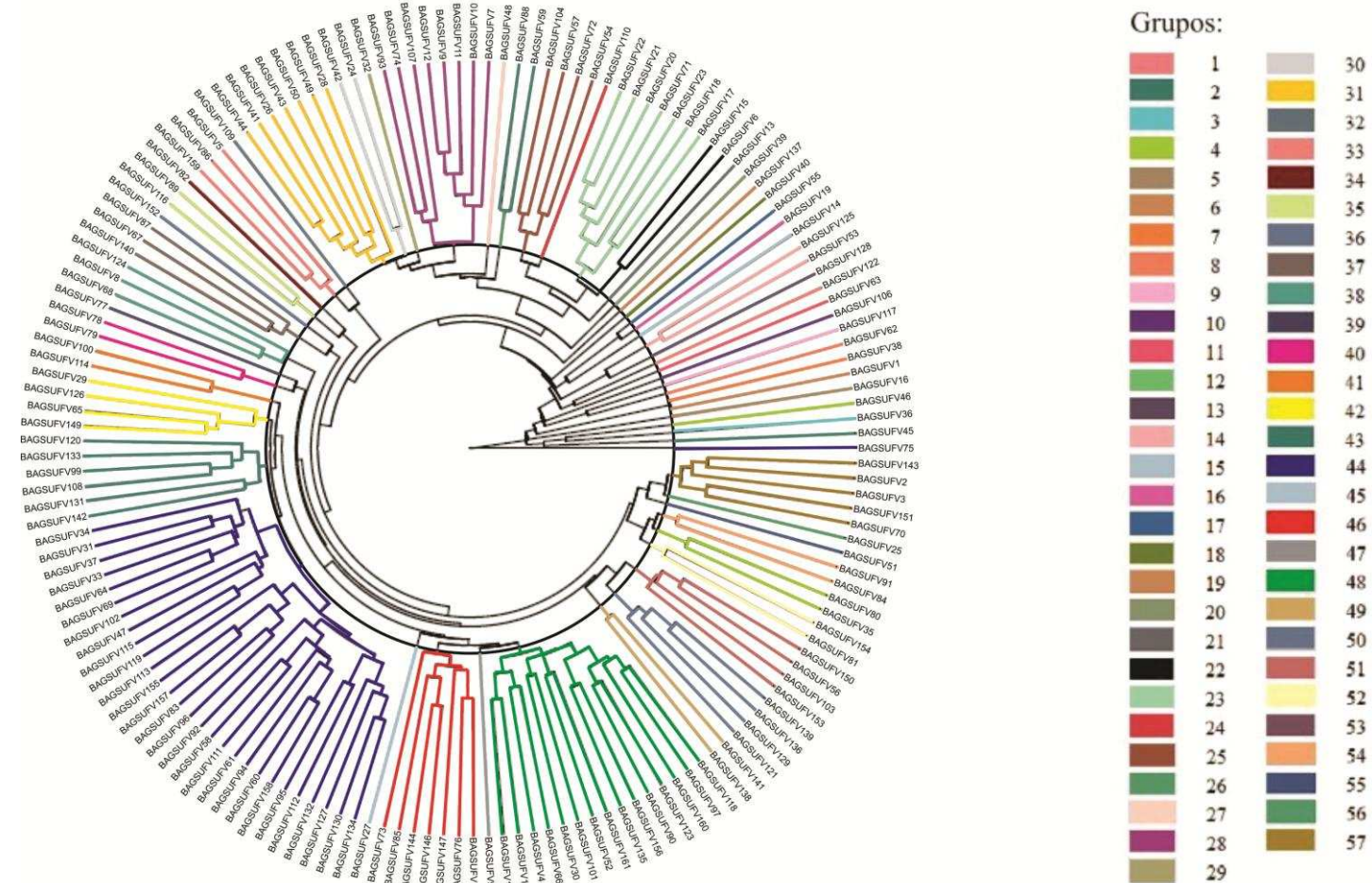


Figura 2. Dendrograma obtido pelo método UPGMA, representativo da dissimilaridade genética entre 161 genótipos de soja avaliados em dois locais no estado de Minas Gerais, em função de 14 caracteres quantitativos e sete qualitativos.

Devido a possibilidade da verificação de comportamento distinto de um mesmo genótipo em locais diferentes (análises não realizadas), para caracteres quantitativos, adotou-se a estratégia de definição de coleção nuclear para cada local de avaliação

A determinação do número mínimo de acessos, em cada coleção nuclear, capazes de representar a média e a variância da população original para cada caractere quantitativo, ocorreu mediante a análise dos gráficos de dispersão destes parâmetros (Figuras 3, 4, 5 e 6).

Analisando os resultados para o local 1, o número de acessos estimado para representar a média da população original, para cada caractere, variou de 55 a 85 acessos. Os caracteres altura de inserção da primeira vagem (Figura 3e) e número de vagens por planta (Figura 3i) foram os que exigiram o menor número de acessos. Já os caracteres altura da planta na maturidade (Figura 3d) e massa de cem sementes (Figura 3l) foram os que demandaram os maiores números de acessos.

Neste mesmo local o número mínimo de acessos necessário para a representação da variância da população original foi encontrado para a característica número de dias para floração (Figura 4a), sendo este valor de 60 acessos. Para os caracteres massa de cem sementes (Figura 4l) e teor de proteína (Figura 4o) foram necessários um número elevado de acessos. Na característica número de hastes laterais (Figura 4g) nenhuma amostra com número de acessos inferior a 160 foi suficiente para a representação da variância original.

Considerando os resultados obtidos para o local 2, de maneira geral, a quantidade de acessos amostrados para a representação da média original foi superior àquela encontrada para os caracteres correspondentes ao local 1. No entanto, a maioria dos valores estiveram abaixo de 85 indivíduos, com exceção àqueles definidos para os caracteres número de sementes por vagem (Figura 5j) e ângulo de acamamento (Figura 5h), sendo necessário 110 e 115 acessos respectivamente.

Na representação da variância da população original avaliada no local 2, para nove dos quatorze caracteres avaliados, um número de 85 indivíduos foi capaz de representar a variância original. Para os caracteres restantes, em três não houve possibilidade de representação da variância por meio de um conjunto com menos de 160 indivíduos, sendo eles: diâmetro do hipocótilo (Figura 6c), número de hastes laterais (Figura 6g) e ângulo de acamamento (Figura 6h).

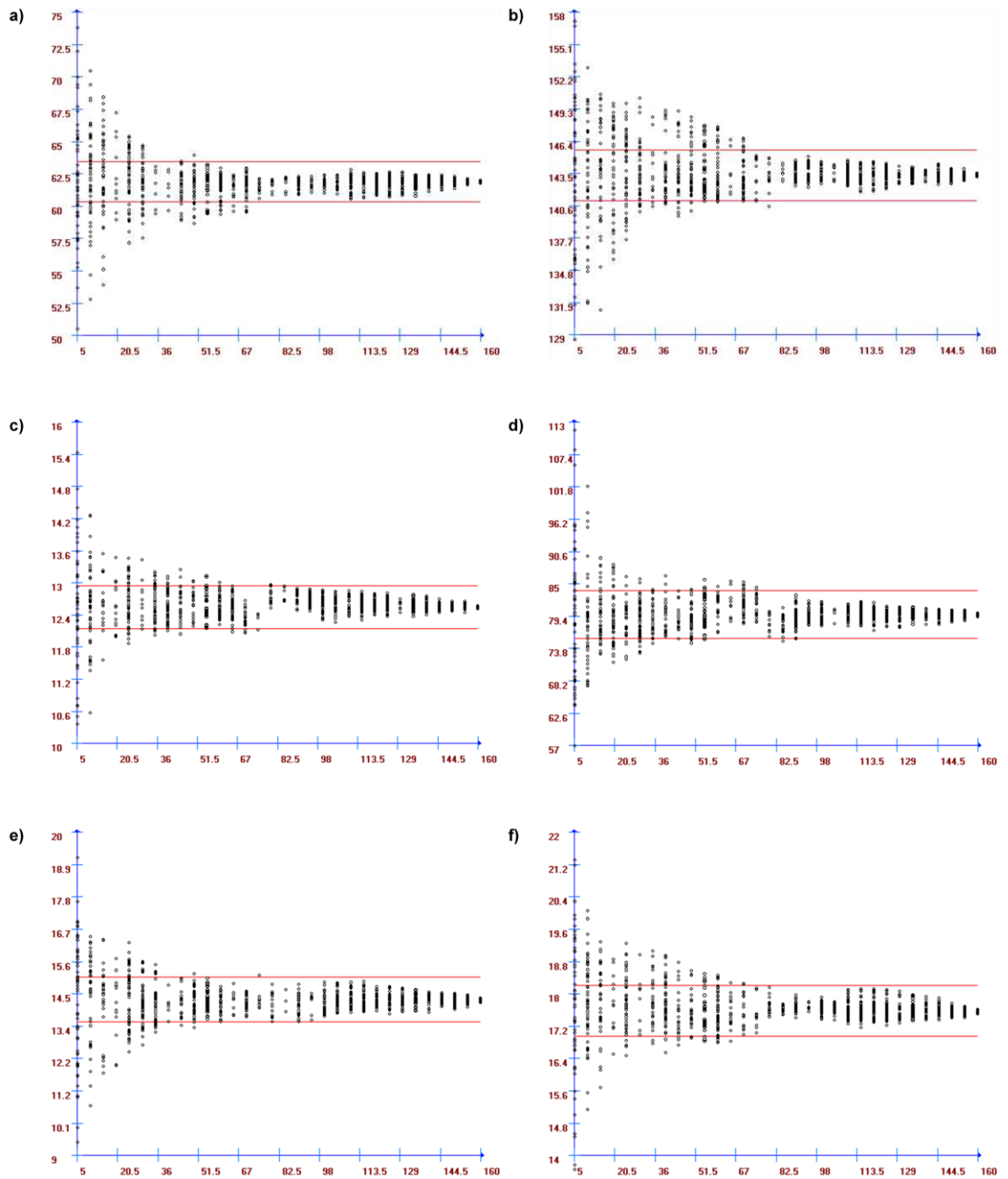
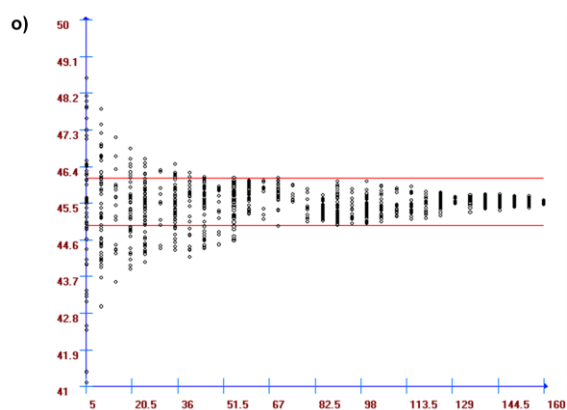
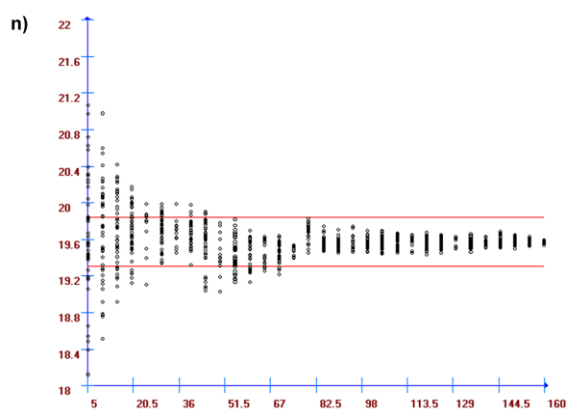
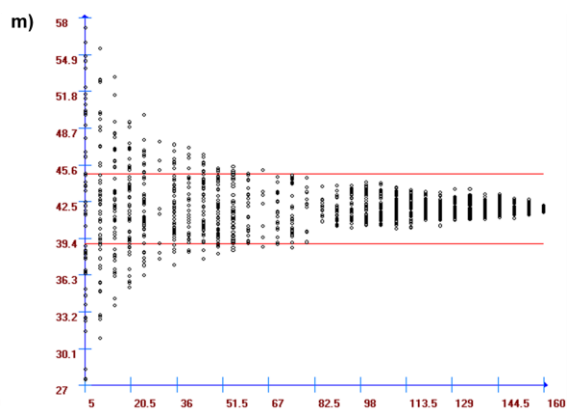
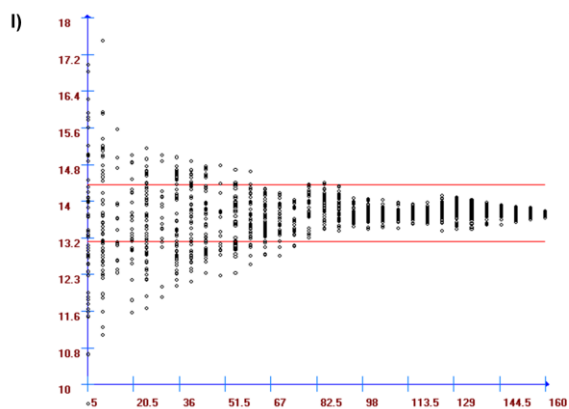
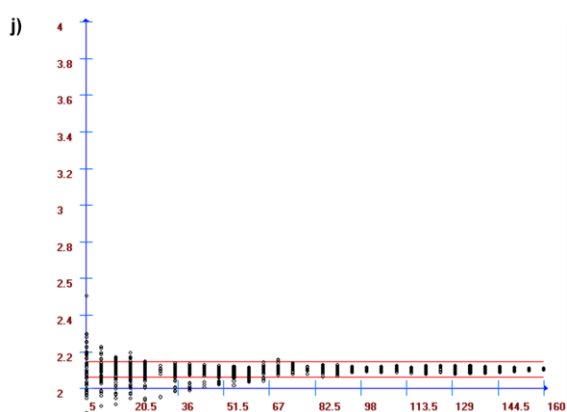
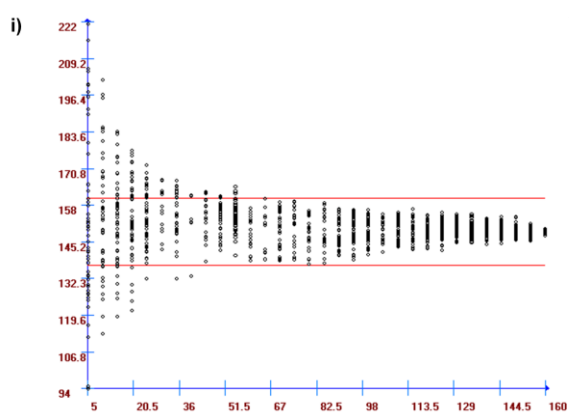
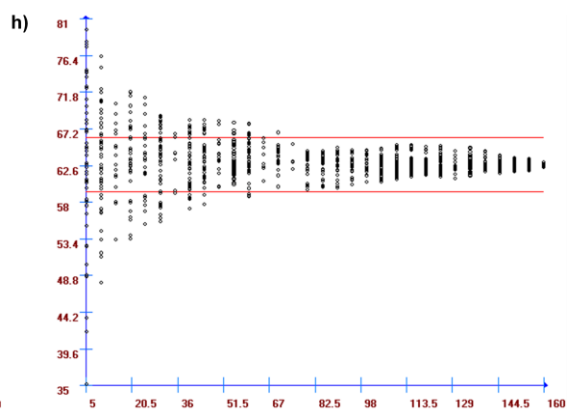
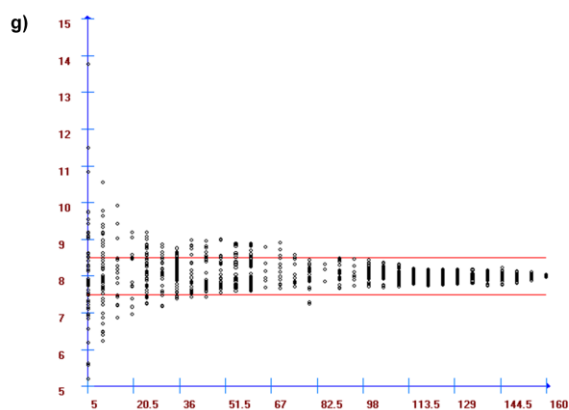


Figura 3. Dispersão gráfica dos valores da média (eixo y) obtida em amostras constituídas por diferentes números de indivíduos (eixo x) para as variáveis: número de dias para floração (a); número de dias para maturidade (b); diâmetro do hipocótilo (c); altura da planta na maturidade (d); altura de inserção da primeira vagem (e); número de nós da haste principal (f); número de hastes laterais (g); ângulo de acamamento (h); número de vagens por planta (i); número de sementes por vagem (j); massa de cem sementes (l); produção (m); teor de óleo (n) e teor de proteína (o) avaliadas no local 1.

Continuação da figura 3.



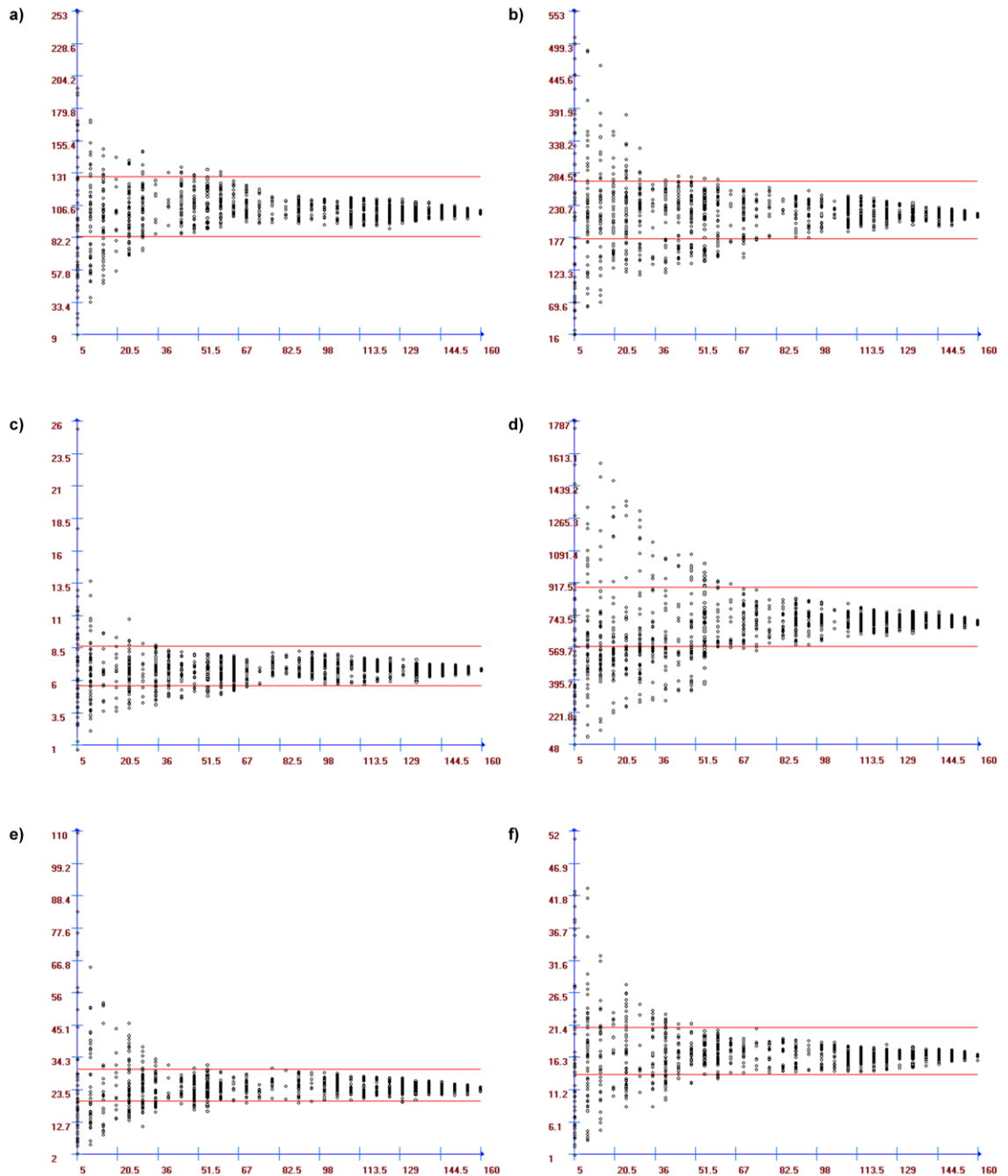
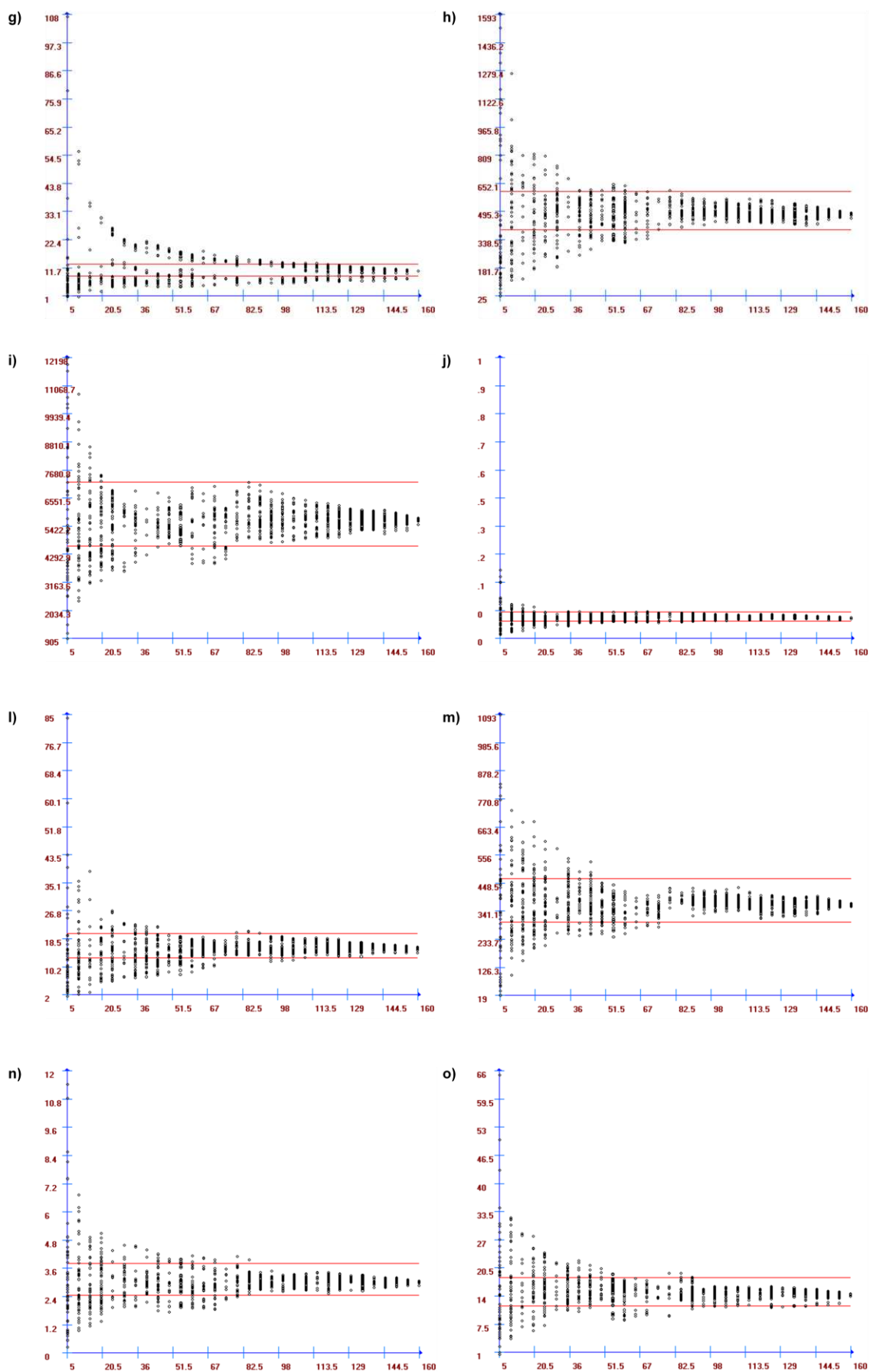


Figura 4. Dispersão gráfica dos valores da variância (eixo y) obtida em amostras constituídas por diferentes números de indivíduos (eixo x) para as variáveis: número de dias para floração (a); número de dias para maturidade (b); diâmetro do hipocótilo (c); altura da planta na maturidade (d); altura de inserção da primeira vagem (e); número de nós da haste principal (f); número de hastes laterais (g); ângulo de acamamento (h); número de vagens por planta (i); número de sementes por vagem (j); massa de cem sementes (l); produção (m); teor de óleo (n) e teor de proteína (o) avaliados no local 1.

Continuação da figura 4.



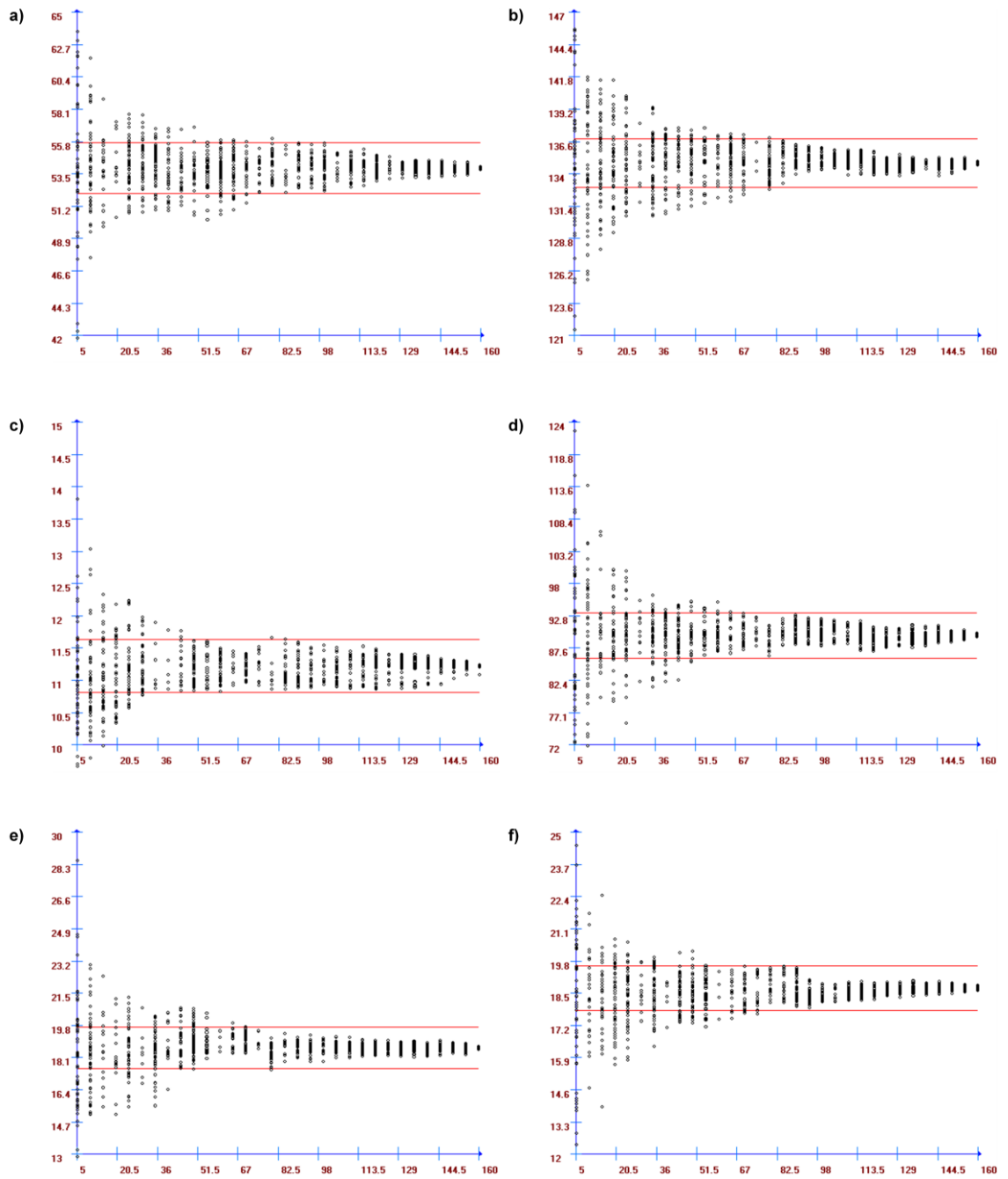
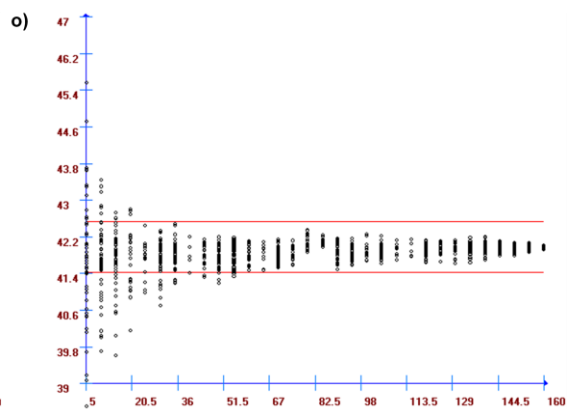
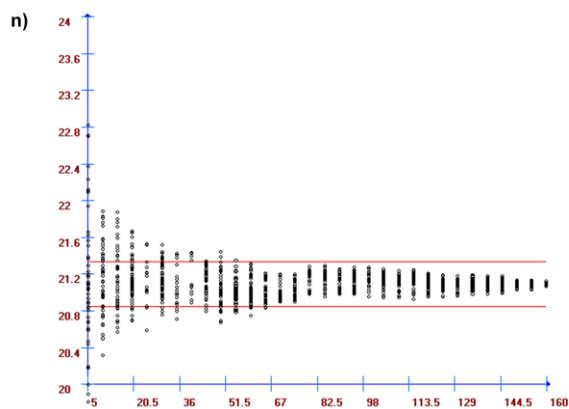
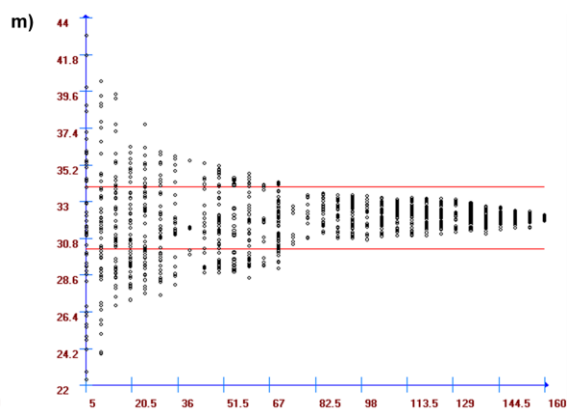
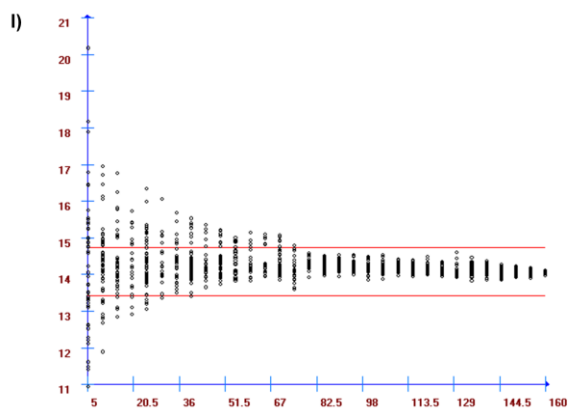
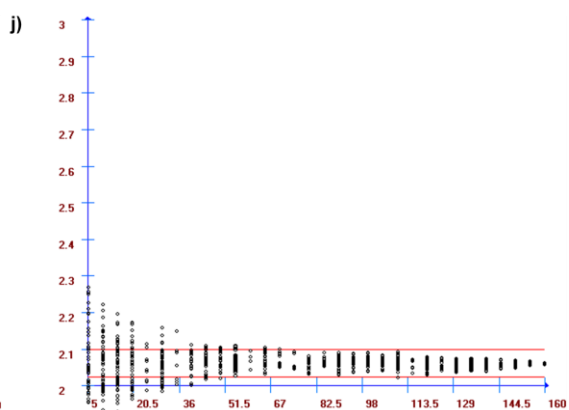
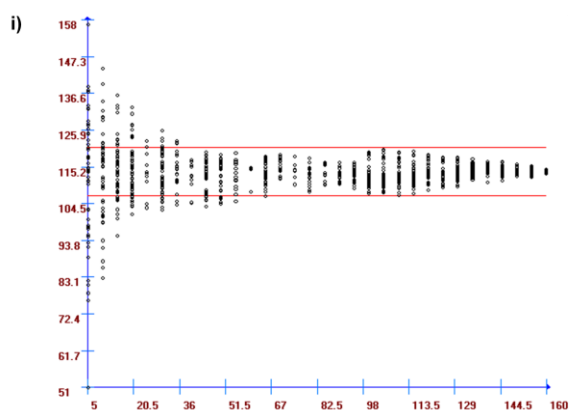
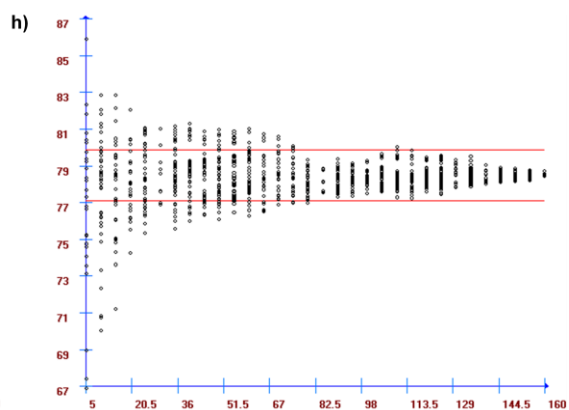
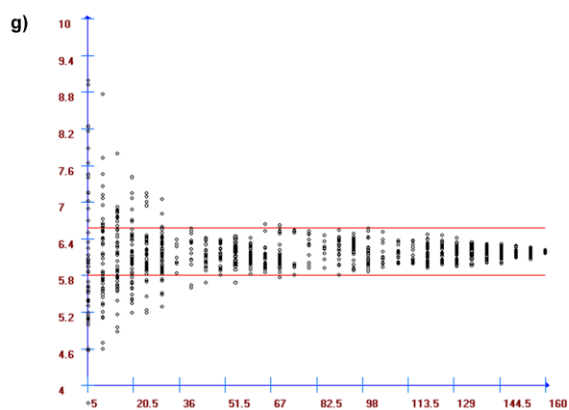


Figura 5. Dispersão gráfica dos valores da média (eixo y) obtida em amostras constituídas por diferentes números de indivíduos (eixo x) para as variáveis: número de dias para floração (a); número de dias para maturidade (b); diâmetro do hipocótilo (c); altura da planta na maturidade (d); altura de inserção da primeira vagem (e); número de nós da haste principal (f); número de hastes laterais (g); ângulo de acamamento (h); número de vagens por planta (i); número de sementes por vagem (j); massa de cem sementes (l); produção (m); teor de óleo (n) e teor de proteína (o) avaliadas no local 2.

Continuação da figura 5.



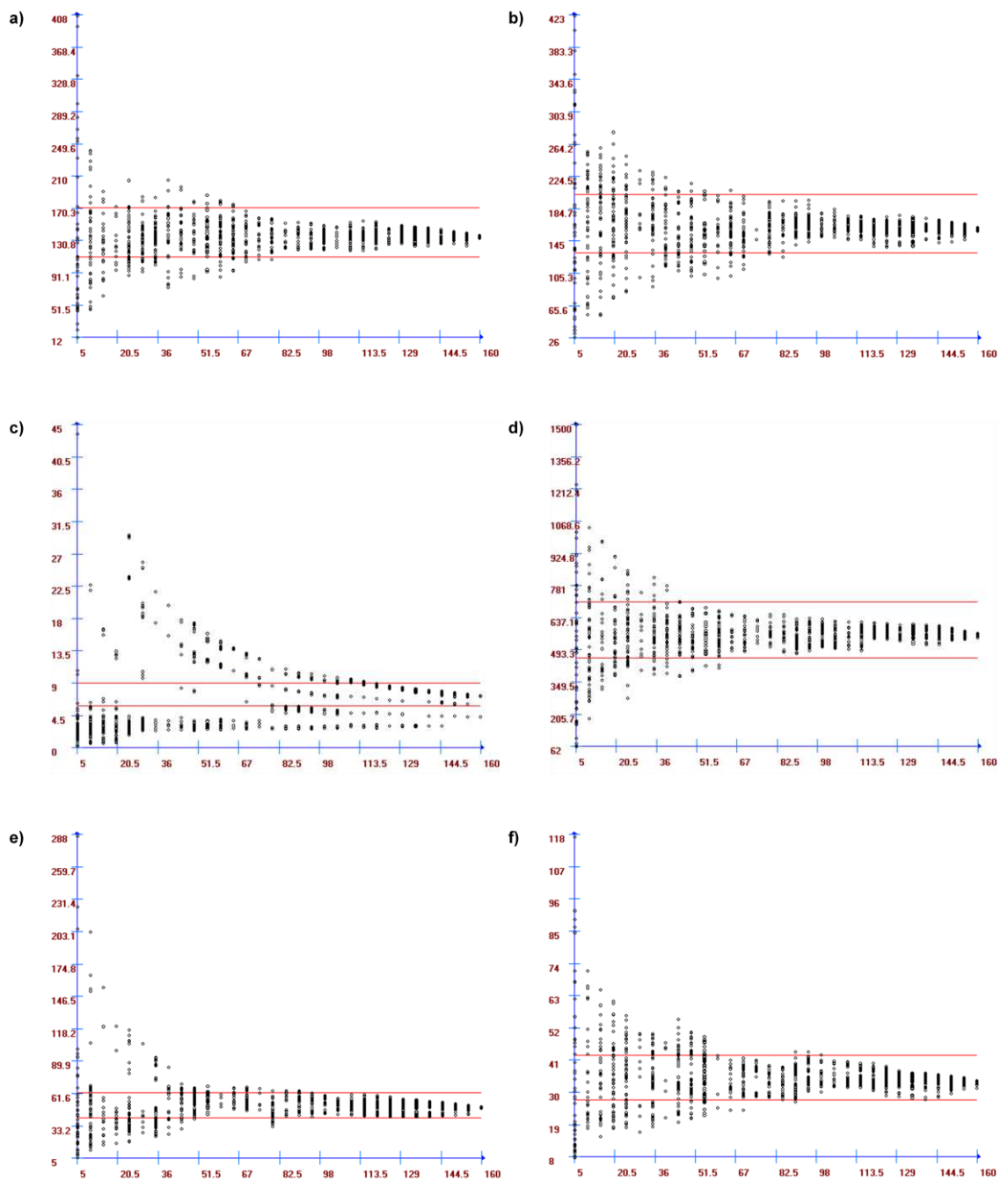
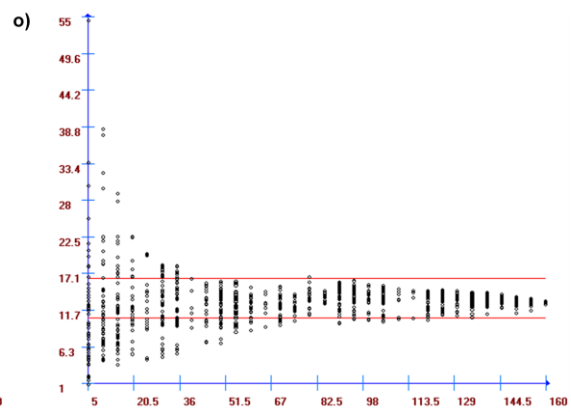
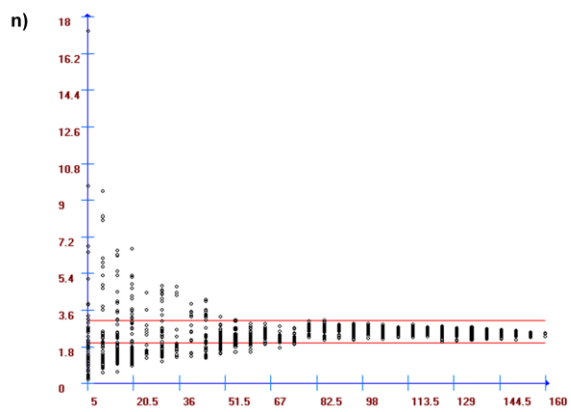
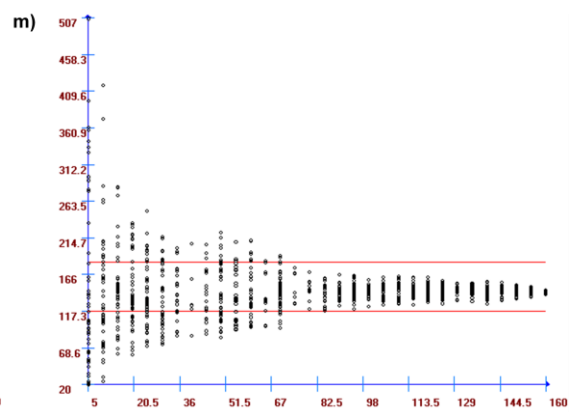
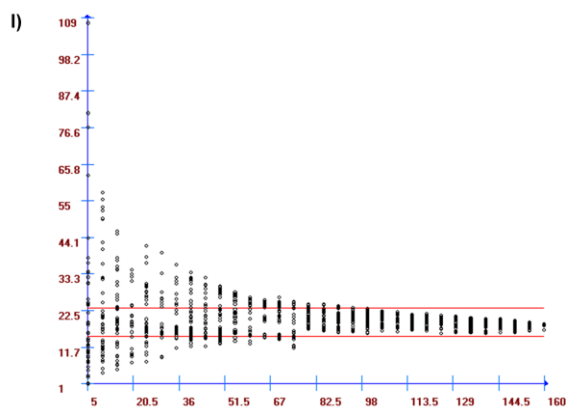
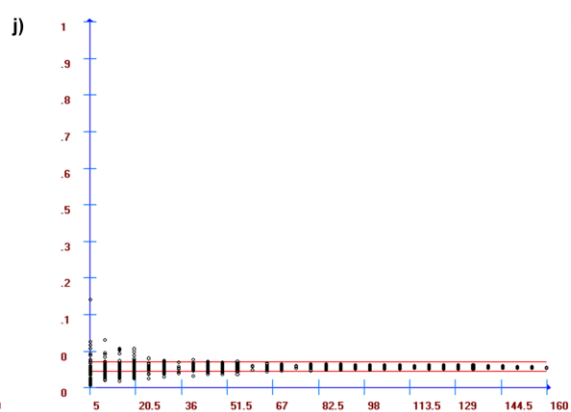
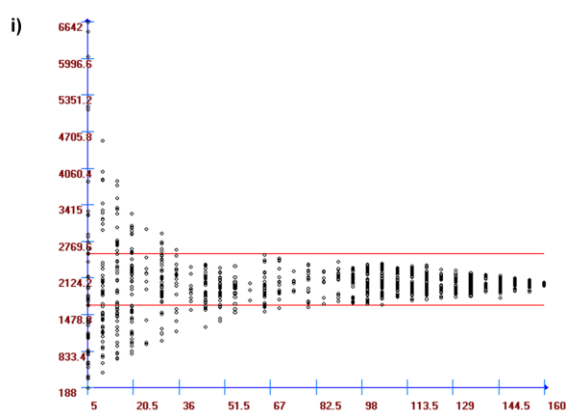
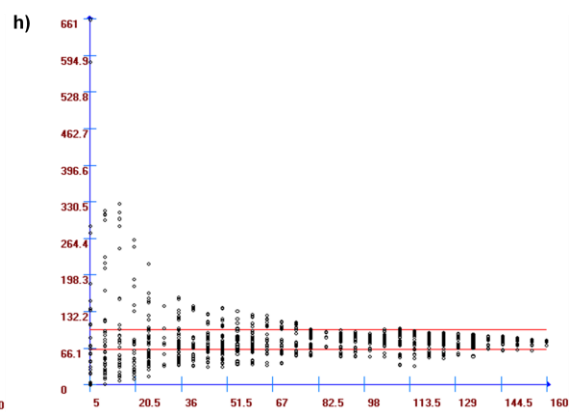
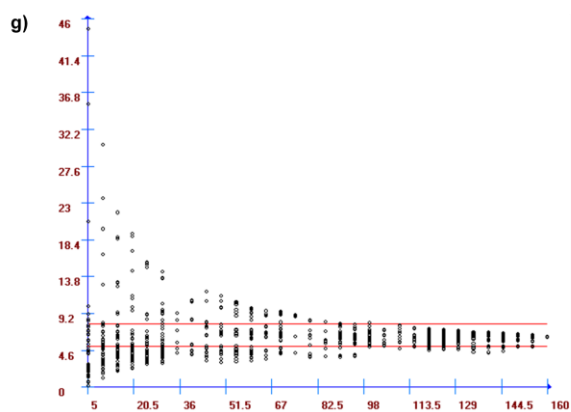


Figura 6. Dispersão gráfica dos valores da variância (eixo y) obtida em amostras constituídas por diferentes números de indivíduos (eixo x) para as variáveis: número de dias para floração (a); número de dias para maturidade (b); diâmetro do hipocótilo (c); altura da planta na maturidade (d); altura de inserção da primeira vagem (e); número de nós da haste principal (f); número de hastes laterais (g); ângulo de acamamento (h); número de vagens por planta (i); número de sementes por vagem (j); massa de cem sementes (l); produção (m); teor de óleo (n) e teor de proteína (o) avaliados no local 2.

Continuação da figura 6.



Após análise gráfica, verificou-se que, para a maioria das características, quando o número de acessos foi superior a 80, as quantidades de amostras simuladas que se encontravam fora do intervalo de confiança são pequenas.

Cruz *et al.* (2012) relata que o tamanho mínimo de uma coleção nuclear é aquele capaz de representar a potencialidade e a variabilidade da população original, na coleção nuclear, para todas ou para um determinado conjunto de características. Nesse sentido o número de genótipos estabelecidos para compor a coleção nuclear de cada local foi de 80 acessos conforme apresentado na Tabela 6.

De acordo com Van Hintum *et al.* (2000) as coleções nucleares apresentam tamanhos entre 5% e 20% daquelas populações as quais foram retiradas. Na tentativa da manutenção da média e da variância de uma população inicial em uma coleção nuclear, tais parâmetros são fixados, dessa forma, a heterogeneidade entre os acessos tende a ser maior quanto menor for seu número de genótipos (Vasconcelos *et al.*, 2007).

No presente trabalho, a determinação do número de acessos para compor as coleções nucleares, seguindo o limite máximo de 20% proposto por Van Hintum *et al.* (2000), resultaria na seleção de 33 acessos, o que provavelmente significaria maior variabilidade entre os genótipos, entretanto com baixa representatividade da média e da variância da população inicial, conforme representado nas Figuras 3, 4, 5 e 6.

Dos oitenta genótipos selecionados para cada local, os acessos BAGSUFV16, BAGSUFV18, BAGSUFV21, BAGSUFV22, BAGSUFV25, BAGSUFV28, BAGSUFV31, BAGSUFV32, BAGSUFV33, BAGSUFV37, BAGSUFV42, BAGSUFV48, BAGSUFV51, BAGSUFV53, BAGSUFV83, BAGSUFV87, BAGSUFV94, BAGSUFV97, BAGSUFV100, BAGSUFV101, BAGSUFV104, BAGSUFV105, BAGSUFV109, BAGSUFV110, BAGSUFV116, BAGSUFV122, BAGSUFV126, BAGSUFV128, BAGSUFV129, BAGSUFV141 e BAGSUFV143 são comuns aos dois ambientes. Esta coincidência relativamente baixa, 48,75%, demonstra a especificidade de cada coleção nuclear para determinado ambiente. Assim, além dos vários aspectos que podem estar envolvidos na definição de uma coleção nuclear como, origem geográfica dos acessos, caracteres agrônômicos, grupos taxonômicos, regiões ecogeográficas e marcadores genéticos, todos citados por Brown & Spillane (1999), a definição da coleção nuclear deve ser ambiente específico.

Tabela 6. Acessos selecionados para compor a coleção nuclear de cada local.

Acessos							
Local 1			Local 2				
BAGSUFV	1	BAGSUFV	75	BAGSUFV	7	BAGSUFV	76
BAGSUFV	4	BAGSUFV	79	BAGSUFV	10	BAGSUFV	80
BAGSUFV	5	BAGSUFV	83	BAGSUFV	12	BAGSUFV	81
BAGSUFV	6	BAGSUFV	87	BAGSUFV	13	BAGSUFV	83
BAGSUFV	8	BAGSUFV	92	BAGSUFV	16	BAGSUFV	84
BAGSUFV	11	BAGSUFV	94	BAGSUFV	18	BAGSUFV	85
BAGSUFV	15	BAGSUFV	95	BAGSUFV	19	BAGSUFV	86
BAGSUFV	16	BAGSUFV	96	BAGSUFV	20	BAGSUFV	87
BAGSUFV	18	BAGSUFV	97	BAGSUFV	21	BAGSUFV	90
BAGSUFV	21	BAGSUFV	100	BAGSUFV	22	BAGSUFV	93
BAGSUFV	22	BAGSUFV	101	BAGSUFV	25	BAGSUFV	94
BAGSUFV	23	BAGSUFV	104	BAGSUFV	28	BAGSUFV	97
BAGSUFV	24	BAGSUFV	105	BAGSUFV	31	BAGSUFV	98
BAGSUFV	25	BAGSUFV	109	BAGSUFV	32	BAGSUFV	100
BAGSUFV	27	BAGSUFV	110	BAGSUFV	33	BAGSUFV	101
BAGSUFV	28	BAGSUFV	111	BAGSUFV	35	BAGSUFV	104
BAGSUFV	29	BAGSUFV	116	BAGSUFV	36	BAGSUFV	105
BAGSUFV	30	BAGSUFV	118	BAGSUFV	37	BAGSUFV	107
BAGSUFV	31	BAGSUFV	120	BAGSUFV	38	BAGSUFV	109
BAGSUFV	32	BAGSUFV	121	BAGSUFV	40	BAGSUFV	110
BAGSUFV	33	BAGSUFV	122	BAGSUFV	42	BAGSUFV	115
BAGSUFV	34	BAGSUFV	126	BAGSUFV	47	BAGSUFV	116
BAGSUFV	37	BAGSUFV	128	BAGSUFV	48	BAGSUFV	117
BAGSUFV	39	BAGSUFV	129	BAGSUFV	50	BAGSUFV	122
BAGSUFV	42	BAGSUFV	131	BAGSUFV	51	BAGSUFV	124
BAGSUFV	43	BAGSUFV	135	BAGSUFV	52	BAGSUFV	126
BAGSUFV	44	BAGSUFV	139	BAGSUFV	53	BAGSUFV	127
BAGSUFV	46	BAGSUFV	140	BAGSUFV	56	BAGSUFV	128
BAGSUFV	48	BAGSUFV	141	BAGSUFV	57	BAGSUFV	129
BAGSUFV	51	BAGSUFV	142	BAGSUFV	59	BAGSUFV	132
BAGSUFV	53	BAGSUFV	143	BAGSUFV	61	BAGSUFV	133
BAGSUFV	58	BAGSUFV	144	BAGSUFV	62	BAGSUFV	136
BAGSUFV	59	BAGSUFV	147	BAGSUFV	63	BAGSUFV	141
BAGSUFV	60	BAGSUFV	149	BAGSUFV	65	BAGSUFV	143
BAGSUFV	61	BAGSUFV	150	BAGSUFV	66	BAGSUFV	144
BAGSUFV	62	BAGSUFV	151	BAGSUFV	67	BAGSUFV	145
BAGSUFV	63	BAGSUFV	156	BAGSUFV	68	BAGSUFV	152
BAGSUFV	65	BAGSUFV	158	BAGSUFV	70	BAGSUFV	154
BAGSUFV	73	BAGSUFV	159	BAGSUFV	71	BAGSUFV	158
BAGSUFV	74	BAGSUFV	161	BAGSUFV	74	BAGSUFV	161

Uma vez estabelecida as coleções nucleares para cada local, o impacto na recuperação da variância e média da população original para cada caractere foi avaliado (Tabelas 7 e 8).

Analisando o impacto da coleção nuclear obtida para o local 1, foi observada diferença significativa pelo teste F a 1% de probabilidade apenas para a variável número de hastes laterais (Tabela 7), indicando que a variância estimada para os acessos da coleção nuclear foi menor que a da população original. Já para o local 2, a diferença significativa foi verificada para as variáveis diâmetro do hipocótilo, número hastes laterais e angulo de acamamento, sendo, o nível de significância de 1% para as duas primeiras e 5% para a última, pelo teste F (Tabela 8). Esses resultados foram reflexos do estabelecimento do número de acessos que compuseram a coleção nuclear, uma vez que, para esses caracteres o número de acessos inferior a 160 não representaram a variância da população original.

Em relação às médias, as coleções nucleares de cada local favoreceram a manutenção das mesmas em relação às populações originais (Tabelas 7 e 8), uma vez que as diferenças entre as médias das populações originais e as das coleções nucleares não foram significativas pelo teste t.

Na comparação das médias levando em consideração o índice de coincidência da amplitude (CA), os valores encontrados para o local 1 e 2 foram 0,90 e 0,93, respectivamente. De acordo com Hu *et al.* (2000) e Wang *et al.* (2007) índices de coincidência da amplitude de no mínimo 80% caracterizam uma coleção nuclear como representativa.

Tabela 7. Impacto da coleção nuclear, referente ao ensaio 1, na recuperação da variância e média da coleção original.

Variável	Coleção original					Coleção nuclear						
	N	Média	Máximo	Mínimo	DP	N	Média	Máximo	Mínimo	DP	F	t
NDF	161	61,9023	83,1394	41,568	10,0953	80	61,6442	83,1394	42,1394	9,8155	1,0578 ^{ns}	0,1886 ^{ns}
NDM	161	143,347	173,504	112,361	14,6791	80	144,4134	173,5035	112,3606	14,3995	1,0392 ^{ns}	0,5344 ^{ns}
DH	161	12,546	21,1549	6,1392	2,6225	80	12,2924	19,41	6,1392	2,6538	1,024 ^{ns}	0,7041 ^{ns}
APM	161	79,6751	152,588	19,9224	26,6658	80	79,7352	143,5924	19,9224	28,8233	1,1684 ^{ns}	0,016 ^{ns}
AIV	161	14,3091	30,588	4,328	4,9258	80	14,4935	30,588	4,588	5,0544	1,0529 ^{ns}	0,2714 ^{ns}
NNHP	161	17,5824	29,9065	7,5694	4,0875	80	17,4257	28,5694	7,5694	4,234	1,073 ^{ns}	0,2769 ^{ns}
NHL	161	7,9975	29,5412	1,4427	3,238	80	7,714	13,7384	1,4427	2,5095	1,6649 ⁺⁺	0,7476 ^{ns}
AC	161	62,6894	106,102	-1,898	21,9817	80	63,7228	106,102	3,2449	23,7454	1,1669 ^{ns}	0,3346 ^{ns}
NVP	161	148,741	396,474	-12,5337	75,5364	80	146,6247	396,4735	-4,8637	76,9459	1,0377 ^{ns}	0,2036 ^{ns}
NSV	161	2,1049	2,85	1,17	0,2735	80	2,099	2,85	1,3314	0,2818	1,0614 ^{ns}	0,1573 ^{ns}
MCS	161	13,7395	29,9235	3,8535	3,9979	80	14,2551	29,9235	6,0835	4,1555	1,0804 ^{ns}	0,9307 ^{ns}
PROD(g/plt)	161	41,9078	84,8461	-2,1167	19,2195	80	41,7351	84,7561	5,2533	18,609	1,0667 ^{ns}	0,0664 ^{ns}
TO	161	19,5738	24,9378	13,3006	1,7423	80	19,5913	24,9378	15,4649	1,8112	1,0806 ^{ns}	0,0724 ^{ns}
TP	161	45,5342	55,7324	31,321	3,8091	80	45,2595	55,7324	32,571	3,8357	1,014 ^{ns}	0,5258 ^{ns}

^{+,++} Variância heterogênea, pelo teste F a 5% e 1% de probabilidade, respectivamente

*,** Médias diferentes estatisticamente pelo teste t a 5% e 1% de probabilidade, respectivamente.

DP: Desvio padrão

Tabela 8. Impacto da coleção nuclear, referente ao ensaio 2, na recuperação da variância e média da coleção original.

Variável	Coleção original					Coleção nuclear						
	N	Média	Máximo	Mínimo	DP	N	Média	Máximo	Mínimo	DP	F	t
NDF	161	53,9261	87,3265	20,7551	11,6597	80	52,8987	77,1837	20,7551	12,2448	1,1029 ^{ns}	0,6335 ^{ns}
NDM	161	134,854	160,612	106,327	12,6831	80	134,8719	158,4694	106,3265	12,7144	1,005 ^{ns}	0,0102 ^{ns}
DH	161	11,2249	32,5971	4,9629	2,6796	80	11,1998	32,5971	4,9629	3,3989	1,609 ⁺⁺	0,0577 ^{ns}
APM	161	89,6498	137,533	25,7845	23,7201	80	88,3788	135,9459	25,7845	25,2622	1,1343 ^{ns}	0,3833 ^{ns}
AIV	161	18,6133	51,0978	7,4378	7,0444	80	19,1063	51,0978	8,0278	7,9775	1,2825 ^{ns}	0,4893 ^{ns}
NNHP	161	18,7043	32,2871	-1,1971	5,815	80	18,5964	32,2871	-1,1971	6,0638	1,0874 ^{ns}	0,1336 ^{ns}
NHL	161	6,1961	20,8	1,6571	2,5107	80	6,2325	11,3271	2,3671	1,8378	1,8664 ⁺⁺	0,1275 ^{ns}
AC	161	78,4831	87,9796	23,6939	8,8899	80	77,1658	87,9796	23,6939	10,4108	1,3714 ⁺	0,9696 ^{ns}
NVP	161	113,874	242,677	-10,7988	45,0704	80	111,4879	240,3469	-10,7988	41,9357	1,1551 ^{ns}	0,396 ^{ns}
NSV	161	2,0613	2,648	1,4608	0,2363	80	2,05	2,648	1,4608	0,2497	1,1167 ^{ns}	0,3426 ^{ns}
MCS	161	14,0769	32,7702	3,3659	4,3137	80	14,0018	32,7702	3,3659	4,6476	1,1608 ^{ns}	0,1241 ^{ns}
PROD(g/plt)	161	32,0426	68,8571	2,2714	12,0301	80	30,4873	66,9257	2,2714	11,5255	1,0895 ^{ns}	0,9582 ^{ns}
TO	161	21,0897	27,049	15,779	1,57	80	20,9873	24,649	15,779	1,7748	1,2779 ^{ns}	0,4564 ^{ns}
TP	161	41,9771	56,4596	34,8467	3,6168	80	42,2481	56,4596	34,8467	3,8804	1,1511 ^{ns}	0,5346 ^{ns}

^{+,++} Variância heterogênea , pelo teste F a 5% e 1% de probabilidade, respectivamente

*,** Médias diferentes estatisticamente pelo teste t a 5% e 1% de probabilidade, respectivamente.

DP: Desvio padrão

Analisando as coleções nucleares obtidas por meio dos caracteres qualitativos, observou-se que de forma geral houve uma tendência de manutenção das mesmas porcentagens observadas para cada classe na população original (Tabela 9). As classes formato do folíolo lanceolado estreito e hábito de crescimento prostrado foram as únicas não representadas pela coleção nuclear do local 1 e 2, respectivamente.

Tabela 9. Análise da manutenção de variabilidade fenotípica para as características qualitativas, nas coleções nucleares de cada ensaio, por meio do percentual de representatividade de cada classe das características.

Características	Classes	Percentual		
		PO	CNL 1	CNL 2
Cor da flor	Branca	56,52%	53,75%	57,5%
	Roxa	43,48%	46,25%	42,5%
Tipo de inflorescência	Pedunculada	72,67%	75%	67,5%
	Séssil	27,33%	25%	32,5%
Formato do folíolo	Oval pontiagudo	91,30%	90%	91,25%
	Oval arredondado	4,34%	6,25%	5%
	Triangular	1,24%	1,25%	1,25%
	Lanceolado	2,50%	2,5%	1,25%
	Lanceolado estreito	0,62%	-	1,25%
Cor da pubescência	Marrom	53,42%	51,25%	57,5%
	Cinza	46,58%	48,75%	42,5%
Dens. da pubescência	Alta	4,34%	5%	3,75%
	Média	79,50%	85%	83,75%
	Baixa	16,16%	10%	12,5%
Tipo de crescimento	Determinado	65,22%	61,25%	61,25%
	Semideterminado	18,63%	22,5%	20%
	Indeterminado	16,15%	16,25%	18,75%
Hábito de crescimento	Ereto	86,33%	85%	91,25%
	Semiereto	11,80%	11,25%	8,75%
	Prostrado	1,87%	3,75%	-

PO - População original;
 CNL 1- Coleção nuclear do local 1;
 CNL 2 - Coleção nuclear do local 2.

Em relação ao formato do folíolo, a predominância da classe oval pontiaguda na coleção inicial foi refletida na representatividade desta característica pelas coleções nucleares, fazendo com que genótipos que possuem folíolos lanceolados estreitos não fossem selecionados para compor ou estavam em baixa proporção, nas coleções nucleares dos locais 1 e 2, respectivamente. Segundo Souza *et al.* (2014), a realidade das novas cultivares de soja são plantas com folhas com formato dos folíolos lanceolado. Nesse sentido, futuras introduções de genótipos com esse fenótipo devem ser realizadas às coleções para condução de programas de melhoramento.

A porcentagem de retenção de variabilidade para o caractere hábito de crescimento não se mostrou semelhante entre as coleções nucleares. A ausência de genótipos com hábito de crescimento da classe prostrada na coleção nuclear do local 2, pode ser explicada pela baixa porcentagem de cultivares com este tipo de porte na coleção inicial. O porte ereto, observado em maior proporção nas coleções, é uma característica desejável em programas de melhoramento. A arquitetura das plantas com este hábito de crescimento reduz a propensão ao acamamento, favorece a circulação de máquinas e facilita a aplicação de defensivos, aumentando a eficiência no controle de pragas, doenças e plantas daninhas.

Na análise de manutenção da variabilidade para a característica tipo de crescimento, a classe do tipo de crescimento determinado apresentou as mesmas porcentagens de representatividade nas duas coleções nucleares (61,25%), sendo o tipo mais comum observado em todas as coleções. Este tipo de crescimento também é observado entre a maioria das cultivares de soja lançadas no Brasil. No entanto, nos últimos anos o cultivo de genótipos com crescimento indeterminado tem se expandido devido ao prolongamento da fase de floração, fazendo com que estes genótipos sejam recomendados para áreas de plantio que sejam acometidas por adversidades edafoclimáticas.

Diante do exposto, verificou-se que as coleções nucleares foram bem definidas e representaram as médias e as variâncias para a maioria dos caracteres em relação às populações originais.

A classificação dos quinze melhores genótipos para teores de óleo e de proteína em cada local dentro de cada ciclo de maturação foi utilizada para iniciar a estruturação das coleções de melhoramento (Tabelas 10, 11 e 12).

Tabela 10. Acessos de ciclo precoce com as maiores médias para teor de óleo e proteína, selecionados no local 1 e no local 2.

Precoce							
Óleo				Proteína			
Local 1		Local 2		Local 1		Local 2	
Acesso	Média	Acesso	Média	Acesso	Média	Acesso	Média
BAGSUFV40 ¹	23,13	BAGSUFV20	23,37	BAGSUFV12	52,84	BAGSUFV152	51,45
BAGSUFV39 ¹	21,95	BAGSUFV110 ¹	23,35	BAGSUFV7 ²	52,22	BAGSUFV15*	46,1
BAGSUFV93	21,15	BAGSUFV128	23,11	BAGSUFV9	52,22	BAGSUFV46 ²	44,76
BAGSUFV23*	20,78	BAGSUFV23*	23,05	BAGSUFV17*	52,22	BAGSUFV17*	44,23
BAGSUFV95	20,61	BAGSUFV21	22,79	BAGSUFV67	50,73	BAGSUFV6 ²	42,98
BAGSUFV107	20,51	BAGSUFV133	22,75	BAGSUFV11 ²	50,34	BAGSUFV49	42,88
BAGSUFV114	20,22	BAGSUFV99 ¹	22,47	BAGSUFV107	49,98	BAGSUFV73	42,8
BAGSUFV117	19,8	BAGSUFV73	22,31	BAGSUFV114	49,36	BAGSUFV111 ²	41,72
BAGSUFV11	19,72	BAGSUFV111	22,05	BAGSUFV23	48,47	BAGSUFV110	41,1
BAGSUFV7	18,9	BAGSUFV22	22,03	BAGSUFV32 ²	47,39	BAGSUFV99	41,1
BAGSUFV26	18,64	BAGSUFV113 ¹	21,98	BAGSUFV15*	47,22	BAGSUFV113	41,1
BAGSUFV15	18,52	BAGSUFV16	21,97	BAGSUFV26	46,14	BAGSUFV16 ²	41,1
BAGSUFV10	18,5	BAGSUFV152	21,93	BAGSUFV10	45,97	BAGSUFV25 ²	41,1
BAGSUFV88	18,24	BAGSUFV14	21,79	BAGSUFV95	45,88	BAGSUFV21	40,48
BAGSUFV12	18,11	BAGSUFV151	21,55	BAGSUFV88	45,88	BAGSUFV14	40,48

¹Acessos coincidentes para o teor de óleo e ciclos distintos;

²Acessos coincidentes para o teor de proteína e ciclos distintos;

*Acessos coincidentes dentro do mesmo ciclo, para uma mesma variável e de ensaios distintos.

Dos acessos relacionados como precoce (Tabela 10), os acessos BAGSUFV40 no local 1 e o BAGSUFV20 no local 2, apresentaram as maiores médias para teor de óleo, 23,13 e 23,37%, respectivamente. A variação da média neste grupo foi de 18,11 a 23,13% e 21,5 a 23,37%, nos locais 1 e 2, respectivamente. Em relação ao teor de proteína, os acessos que se destacaram foram o BAGSUFV12, com média de 52,84% no local 1, e o BAGSUFV152, com média de 51,45% no local 2.

No ciclo médio (Tabela 11) os teores de óleo variaram de 19,67 a 22,51% entre os acessos selecionados no local 1 e de, 22,07 a 27,05% no local 2. Os acessos BAGSUFV104 e BAGSUFV103 registraram as maiores médias em cada local, sendo 22,51 e 27,07%, respectivamente. Já os acessos BAGSUFV62 e BAGSUFV117 alcançaram as maiores médias para o teor de proteína (55,73 e 54,22%, respectivamente).

Tabela 11. Acessos de ciclo médio com as maiores médias para teor de óleo e proteína, selecionados no local 1 e no local 2.

Médio							
Óleo				Proteína			
Local 1		Local 2		Local 1		Local 2	
Acesso	Média	Acesso	Média	Acesso	Média	Acesso	Média
BAGSUFV104 ¹	22,51	BAGSUFV103	27,05	BAGSUFV62 ²	55,73	BAGSUFV117	54,22
BAGSUFV129	22,24	BAGSUFV109 ¹	24,65	BAGSUFV24	53,47	BAGSUFV93	49,05
BAGSUFV110 ¹	22,12	BAGSUFV108	24,48	BAGSUFV111 ²	51,23	BAGSUFV107	47,97
BAGSUFV113 ¹	21,79	BAGSUFV39 ²	24,36	BAGSUFV6 ²	50,97	BAGSUFV87	47,8
BAGSUFV106	21,78	BAGSUFV56	24,03	BAGSUFV51	49,48	BAGSUFV45	47,26
BAGSUFV59	21,27	BAGSUFV52	23,42	BAGSUFV16 ²	48,47	BAGSUFV78	47,17
BAGSUFV44	20,54	BAGSUFV13	22,9	BAGSUFV18	48,47	BAGSUFV43	46,63
BAGSUFV57 ¹	20,47	BAGSUFV37	22,56	BAGSUFV82*	48,38	BAGSUFV92	46,55
BAGSUFV130	20,17	BAG5SUFV0*	22,48	BAGSUFV20	47,22	BAGSUFV120	46,1
BAGSUFV20	20,16	BAGSUFV120 ¹	22,44	BAGSUFV76	47,13	BAGSUFV36 ²	46,01
BAGSUFV58	20,06	BAGSUFV47	22,39	BAGSUFV55	46,98	BAGSUFV82*	45,92
BAGSUFV50*	19,93	BAGSUFV157	22,37	BAGSUFV50	46,77	BAGSUFV80	45,92
BAGSUFV124	19,91	BAGSUFV42	22,26	BAGSUFV132	46,32	BAGSUFV149	45,83
BAGSUFV102	19,85	BAGSUFV160	22,25	BAGSUFV106 ²	46,23	BAGSUFV59	45,83
BAGSUFV132	19,67	BAGSUFV105	22,07	BAGSUFV58	45,73	BAGSUFV74	45,3

¹Acessos coincidentes para o teor de óleo e ciclos distintos;

²Acessos coincidentes para o teor de proteína e ciclos distintos;

*Acessos coincidentes dentro do mesmo ciclo, para uma mesma variável e de ensaios distintos

Para o número de dias para a maturação acima de 140 (Tabela 12), os acessos que apresentaram os maiores teores de óleo nos locais 1 e 2 foram o BAGSUFV141 e BAGSUFV40, respectivamente. A variação dos teores para esta variável foi de 21,31 a 24,94% no local 1, e de 21,72 a 23,74% no local 2. Para o teor de proteína os acessos BAGSUFV36 e BAGSUFV62 foram os que apresentaram os maiores teores nos locais 1 e 2, respectivamente.

Segundo Moraes *et al.* (2006), sementes de cultivares de soja contém cerca de 40% de proteína e 20% de óleo com base na matéria seca. No entanto, alguns autores relataram a presença de acessos na coleção norte americana com frações de proteínas variando de 31 a 57,9% e a de óleo de 14,7 a 28,4% (CASTRO *et al.*, 1973; COSTA *et al.*, 1974; WILSON, 2004 e OLIVEIRA *et al.* 2007)

Tabela 12. Acessos de ciclo tardio com as maiores médias para teor de óleo e proteína, selecionados no local 1 e no local 2.

Tardio							
Óleo				Proteína			
Local 1		Local 2		Local 1		Local 2	
Acesso	Média	Acesso	Média	Acesso	Média	Acesso	Média
BAGSUFV141	24,94	BAGSUFV40 ¹	23,74	BAGSUFV36 ²	54,89	BAGSUFV62 ²	56,46
BAGSUFV138	23,63	BAGSUFV33*	23,62	BAGSUFV3	53,84	BAGSUFV147	51,45
BAGSUFV108	22,95	BAGSUFV57 ¹	23,36	BAGSUFV5	53,47	BAGSUFV137	51,1
BAGSUFV131*	22,83	BAGSUFV153	23,14	BAGSUFV146	52,49	BAGSUFV106 ²	47,97
BAGSUFV120 ¹	22,69	BAGSUFV154	23,1	BAGSUFV143	51,24	BAGSUFV122	47,97
BAGSUFV115*	22,25	BAGSUFV145	23,09	BAGSUFV119	51,23	BAGSUFV7 ²	46,73
BAGSUFV60	22,06	BAGSUFV100	23	BAGSUFV46 ²	51,14	BAGSUFV28	46,63
BAGSUFV48	22,01	BAGSUFV131*	22,98	BAGSUFV63	50,73	BAGSUFV126	46,1
BAGSUFV142	21,99	BAGSUFV104 ¹	22,98	BAGSUFV115	50,61	BAGSUFV140	45,2
BAGSUFV33*	21,89	BAGSUFV121	22,79	BAGSUFV109	50,61	BAGSUFV33	44,76
BAGSUFV89	21,88	BAGSUFV101	22,54	BAGSUFV4	50,34	BAGSUFV11 ²	44,23
BAGSUFV109 ¹	21,82	BAGSUFV53*	22,48	BAGSUFV13	49,72	BAGSUFV124	44,23
BAGSUFV53*	21,53	BAGSUFV115*	22,25	BAGSUFV25 ²	49,72	BAGSUFV48	44,13
BAGSUFV99 ¹	21,5	BAGSUFV148	21,94	BAGSUFV69	48,85	BAGSUFV32 ²	44,13
BAGSUFV72	21,31	BAGSUFV64	21,72	BAGSUFV66	48,85	BAGSUFV118	43,6

¹Acessos coincidentes para o teor de óleo e ciclos distintos;

²Acessos coincidentes para o teor de proteína e ciclos distintos;

*Acessos coincidentes dentro do mesmo ciclo, para uma mesma variável e de ensaios distintos

Em relação à seleção realizada no local 1 houve uma tendência da alocação de acessos com maiores teores de óleo e proteína no grupo de ciclo tardio. Já no local 2 esta tendência ocorreu para os acessos pertencentes ao grupo de ciclo médio. Cavalcante *et al.* (2009), em avaliação de linhagens de soja de ciclo precoce e tardio, quanto ao teor óleo para produção de biodiesel, encontrou as maiores médias para as de ciclo tardio.

A classificação para teor de óleo revelou a existência de genótipos com diferentes ciclos nos diferentes locais, sendo eles: BAGSUFV40, BAGSUFV39, BAGSUFV110, BAGSUFV99, BAGSUFV113, BAGSUFV104, BAGSUFV57, BAGSUFV109 e BAGSUFV120. O mesmo ocorreu em relação ao teor de proteína para os acessos BAGSUFV7, BAGSUFV11, BAGSUFV32, BAGSUFV46, BAGSUFV6, BAGSUFV111, BAGSUFV16, BAGSUFV25, BAGSUFV62, BAGSUFV106 e BAGSUFV36. O comportamento destes acessos indica a existência de interação genótipo ambiente para o caractere ciclo.

Analisando cada ciclo de maturação verificou-se que a seleção do acesso BAGSUFV23, para o caractere teor de óleo, foi coincidente em ambos os locais. O mesmo ocorreu para os acessos BAGSUFV17 e BAGSUFV15, em relação ao teor de proteína. Para o ciclo médio o acesso comum para teor de óleo foi o BAGSUFV50 e para proteína o BAGSUFV82. No grupo do ciclo tardio foi verificado o maior número de acessos coincidentes para o teor de óleo, sendo eles: BAGSUFV131, BAGSUFV115, BAGSUFV33 e BAGSUFV53. Neste grupo não houve acessos comuns para o caractere teor de proteína.

Além disso, não foi verificada a classificação de nenhum acesso que possuía tanto alto teor de óleo quanto de proteína, evidenciando a relação inversamente proporcional entre esses caracteres.

Os acessos selecionados para comporem as coleções de melhoramento para os caracteres teores de óleo e de proteína dentro de cada ciclo de maturação se encontram na Tabela 13.

Tabela 13. Coleções de melhoramento para teor de óleo e proteína, estruturadas por ciclo da soja.

Precoce		Médio		Tardio	
Óleo	Proteína	Óleo	Proteína	Óleo	Proteína
BAGSUFV40	BAGSUFV12	BAGSUFV104	BAGSUFV62	BAGSUFV141	BAGSUFV36
BAGSUFV39	BAGSUFV7	BAGSUFV129	BAGSUFV24	BAGSUFV138	BAGSUFV3
BAGSUFV93	BAGSUFV9	BAGSUFV110	BAGSUFV111	BAGSUFV108	BAGSUFV5
BAGSUFV95	BAGSUFV17	BAGSUFV113	BAGSUFV6	BAGSUFV131	BAGSUFV146
BAGSUFV114	BAGSUFV67	BAGSUFV106	BAGSUFV51	BAGSUFV120	BAGSUFV143
BAGSUFV117	BAGSUFV11	BAGSUFV59	BAGSUFV16	BAGSUFV115	BAGSUFV119
BAGSUFV11	BAGSUFV114	BAGSUFV44	BAGSUFV18	BAGSUFV60	BAGSUFV46
BAGSUFV7	BAGSUFV23	BAGSUFV57	BAGSUFV82	BAGSUFV48	BAGSUFV63
BAGSUFV26	BAGSUFV32	BAGSUFV130	BAGSUFV20	BAGSUFV142	BAGSUFV115
BAGSUFV15	BAGSUFV15	BAGSUFV20	BAGSUFV76	BAGSUFV89	BAGSUFV109
BAGSUFV20	BAGSUFV152	BAGSUFV103	BAGSUFV117	BAGSUFV40	BAGSUFV62
BAGSUFV110	BAGSUFV15	BAGSUFV109	BAGSUFV93	BAGSUFV33	BAGSUFV147
BAGSUFV128	BAGSUFV46	BAGSUFV108	BAGSUFV107	BAGSUFV57	BAGSUFV137
BAGSUFV23	BAGSUFV6	BAGSUFV39	BAGSUFV87	BAGSUFV153	BAGSUFV106
BAGSUFV21	BAGSUFV49	BAGSUFV56	BAGSUFV45	BAGSUFV154	BAGSUFV122
BAGSUFV133	BAGSUFV73	BAGSUFV52	BAGSUFV78	BAGSUFV145	BAGSUFV7
BAGSUFV99	BAGSUFV111	BAGSUFV13	BAGSUFV43	BAGSUFV100	BAGSUFV28
BAGSUFV73	BAGSUFV110	BAGSUFV37	BAGSUFV92	BAGSUFV104	BAGSUFV126
BAGSUFV111	BAGSUFV99	BAGSUFV50	BAGSUFV120	BAGSUFV121	BAGSUFV140
BAGSUFV22	BAGSUFV113	BAGSUFV120	BAGSUFV36	BAGSUFV101	BAGSUFV33

Diante da estratégia adotada para obtenção das coleções de melhoramento foi possível ressaltar que, ao selecionar acessos diferentes em cada local e com altos teores de óleo ou de proteína, genes associados a esses caracteres estão sendo diferencialmente expressos por local, dando um indicativo de diversidade genética. De acordo com Kang e Gauch Júnior (1996) esta expressão diferenciada de um caractere ocorre quando a contribuição de alelos dos diferentes genes que controlam esse caractere ou o nível e expressão dos mesmos não são coincidentes entre os ambientes. Segundo Squilassi (2003), se o ranking de genótipos com base no fenótipo de desempenho individual muda entre os ambientes, significativa variância aditiva poderá ser mantida a partir da seleção de acessos distintos.

A estruturação das coleções de melhoramento forneceu subsídio para a escolha de genitores a serem cruzados, em esquema de dialelo parcial, com cultivares elites altamente produtivas e que possuem ciclo de maturidade precoce, médio ou tardio. Diante disso, o estabelecimento dessas coleções, com acessos que apresentam alta média para teores de óleo ou proteína, divergentes e adaptados a diferentes regiões produtoras de soja, se mostrou boa estratégia para o melhoramento genético.

CONCLUSÕES

Existe variabilidade genética entre os acessos para a maioria dos caracteres avaliados, sendo esta variabilidade condicionada à expressão diferencial de genes nos diferentes ambientes.

A estruturação das coleções nucleares favoreceu a representatividade das médias e das variâncias, para a maioria dos caracteres, em relação às populações originais.

Os genótipos que compuseram as coleções de melhoramento apresentaram alta média para teores de óleo ou proteína, diversidade genética e estabilidade de produção à diferentes regiões produtoras de soja.

REFERÊNCIAS BIBLIOGRÁFICA

ABIOVE. Associação Brasileira das Indústrias de Óleos Vegetais. **Exportações do complexo soja.** 2015. Disponível em: <<http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=>>> acesso em: 12/02/2016

ALCÂNTARA NETO, F. Caracterização genético-molecular de um banco ativo de germoplasma de soja. 2005. 75p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa.

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

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

ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS (Arlington, Estados Unidos). **Official methods of analysis.** Washington, DC, 1975. 1094p

BARBÉ, T.C.; AMARAL JÚNIOR, A.T.; GONÇALVES, L.S.A.; RODRIGUES, R.; SCAPIM, C.A. Association between advanced generations and genealogy in inbred lines of snap bean by the Ward- Modified Location Model. **Euphytica**, v.173, n.3, p.337-343, 2010.

BARRETO, L. S. T.; MOURA, W. C. O.; REIS, R. S.; HOSADA, L. R.; MAIA, G. V. C.; PENA, G. M. Soja integral processada em dietas para codornas japonesas em postura. **Revista brasileira de Zootecnia**, Viçosa, v.39, n.9, p. 1978-1983, set. 2010.

BENTO, C.S.; SUDRÉ, C.P.; RODRIGUES, R.; RIVA, E.M.; PEREIRA, M.G. Descritores qualitativos e multicatégoricos na estimativa da variabilidade fenotípica entre acessos de pimentas, **Scientia Agraria**, v.8, n.2, p.149-156, 2007.

BERTAN, I.; CARVALHO, F.I.F.; OLIVEIRA, A.C.; VIEIRA, E.A.; HARTWIG, I.; SILVA, J.A.G.; SHIMIDT, D.A.M.; VALERIO, I.P.; BUSATO, C.C.; RIBEIRO, G. Comparação de métodos de agrupamento na representação da distância morfológica entre genótipos de trigo. Pelotas: **Revista Brasileira de Agrociência**, v.12, n.3, p.279-286, jul./set. 2006.

BERTINI, C.H.C.M.; SCHUSTER, I.; SEDIYAMA, T.; BARROS, E.G. de; MOREIRA, M.A. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. **Genetics and Molecular Biology**, v.29, p.321-329, 2006.

BRASIL. Ministério da Agricultura Pecuária e Abastecimento –MAPA . Instrução para Execução dos Ensaios de Distingibilidade, Homogeneidade e Estabilidade de Cultivares de Soja. **Diário Oficial** , 2002. P. 8 e 9.

BROWN, A.H.D.; SPILLANE, C. Implement core collections - principles, procedures, progress, problems and promise. In: JOHNSON, R. C.; HODGKIN, T. (Ed.) **Core collections for today and tomorrow**. Rome, Italy: IPGRI, 1999. P.1-9.

BROWN-GUERIDA, G.L.; THOMPSON, J.A.; NELSON, R.L.; WARBURTON, M.L. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. **Crop Science**, v.40, p.815-823, 2000.

BUENO, L.C.S.; MENDES, A. N. G.; CARVALHO, S. P. **Melhoramento de plantas: princípios e procedimentos**, Lavras: UFLA, 2001, 282 p.

BUZAR, A.G.R.; OLIVEIRA, V.R.; BOITEUX, L.S. Estimativa da diversidade genética de germoplasma de cebola via descritores morfológicos, agronômicos e bioquímicos. **Horticultura Brasileira**, v.25, n.4, p.513-518, 2007.

CARCIOFI, A.C.; PONTIERI, R.; FERREIRA, C.F.; PRADA, F. Avaliação de dietas com diferentes fontes protéicas para cães adultos. **Revista Brasileira de Zootecnia**, v. 35. n. 3, p.754-760, 2006.

CASTRO, A.T.B.; MILLAN,A.; LAGO, R.C.A. **Contribuição ao estudo da soja no Brasil**. Rio de Janeiro : Centro de Tecnologia Agrícola e Alimentar/Ministério da Agricultura, 1973. 28p. (Boletim técnico, 10).

CAVALCANTE, A.K.; ESPINDOLA, S. M.C.G.; HAMAWAKI, O.T.; BISINOTTO, F.F.; COSTA, E.G.; GONÇALVES.F.A. Avaliação e seleção de linhagens de soja quanto ao teor de óleo para a produção de biodiesel. Uberaba: **FAZU em Revista**, n.6, p.11-52, 2009.

CHEATTRI, M.; MONDAL, S.; NATH, R. Studies on genetic variability in Soybean (*Glycine max*(L.) Merrill) in the mid hills of Darjeeling District. **Journal of Interacademia**, v.9, n.1, p.175-178, 2005.

CHURCH, D. C; POND, W. G. Bases científicas para la nutriciôn y alimentaciôn de los animales domèsticos. Zaragoza: **Acribia**, 1977. 462 p.

CLAPPER, G.M.; GRIESHOP, C.M.; MERCHEN, N.R.; RUSSETT, J.C.; BRENT, J.L. Jr.; FAHEY, G.C. Jr. Ileal and total tract nutrient digestibilities and fecal characteristics of dogs as affect by soybean protein inclusion in dry, extruded diets. **J. Anim. Sci.**, v.79, p. 1523-1532, 2001.

CONAB. Companhia Nacional de Abastecimento. **Acompanhamento da Safra Brasileira de soja 2015/2016 quinto levantamento**, fevereiro/2016. Brasília: Conab, 2016. 182p. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_02_04_11_21_34_boletim_graos_fevereiro_2016_ok.pdf> acesso em: 12/02/2016

COSTA, S.L.; MIYA, E.E.; FUJITA. J.T. Composição química e qualidades organolépticas e nutricionais das principais variedades de soja cultivadas no Estado de São Paulo. **Coletânea do Instituto de Tecnologia de Alimentos**, v.5, p.305-319, 1974.

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

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

CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco-MG: Suprema, 2011. 620p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 4.ed. Viçosa: Ed. da UFV, 2012. 514p.

DERSJANT-LI Y.; HILL, D. A. Why soy? The use of soy protein concentrate in pet foods. **Petfood Ind.**, v. 47, n. 1, p.16-18, 2005.

DIWAN, N.; MCINTOSH, M.; BAUCHAN, G. Methods of developing a core collection of annual Medicago species. **Theoretical and Applied Genetics**, v.90, n.6, p.755-761, 1995.

FERRÃO, M. A. G.; FONSECA, A. F. A.; FERRÃO, R. G.; BARBOSA, W. M.; SOUZA, E. M. R. Genetic divergence in Conilon coffee revealed by RAPD markers. **Crop Breeding and Applied Biotechnology**. v.9, p.63-74, 2009.

FERRARI, R.A.; OLIVEIRA, V.S.; SCABIO, A. Biodiesel de soja – taxa de conversão em ésteres etílicos, caracterização físicoquímica e consumo em gerador de energia. **Química Nova**, v.28, n.1, p.19-23, 2005.

FILHO, A. C.; EVANGELISTA, D. H. R.; GONÇALVES, E. C. P.; STORCK, L. Tamanho de amostra de caracteres de genótipos de soja. **Ciência Rural**, v. 39, n. 4, p. 983-991, 2009.

FRANCO, J.; CROSSA, J.; RIBAUT, J.; BERTAN, J.; WARBURTON, M.; KHAIRALLAH, M. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. **Theoretical and Applied Genetics**, v.103, n.6, p.944-952, 2001.

FRANKEL, O.H.; BENNETT, E. **Genect resources in plants: their explorations and conservation**. Oxford: Blackwell, 1970. 554p. (International Biological Programme, Handbock, 11).

FRANKEL, O.H.; Genect perspectives of germplasm conservation. In: ARBER, W.K.; LLIMENSAA, K.; PEACOCK, W.J.; STAELINGER, P(Eds) **Genetic manipulation: impact on man and society**. Cambridge: Cambridge University, 1984. P. 161-170

FRANKEL, O.H.; SOULE, M.E. **Conservation and evolution**. Londres: Cambridge University Press, 1981. 327p.

GLASENAPP, J.S.; SEDIYAMA, T.; CRUZ, C.D.; MATSUO, E.; BROMMONSCHENKEL, S.H.; OLIVEIRA, R.C.T.; HAMAWAKI, O.T. Diversidade de características agronômicas e moleculares em cultivares de soja com diferentes graus de resistência à phakopsora pachyrhizi. Uberlândia: **Bioscience Journal**, v.31, n.1, p.25-36, Jan./Feb. 2015.

GONÇALVES, L.; RODRIGUES, R.; AMARAL, A.T.; KARASAWA, M.; SUDRE. Heirloom tomato gene bank : assessing genetic divergence based on morphological, agronomic and molecular data using a Ward- Modified Location Mode. **Genetics and Molecular Research**, v.8, n.1, p.364-374, 2009.

HOLBROOK, C.C.; ANDERSON, W.F.; PITTMAN, R.N. Selection of a core collection from the US germplasm collection of peanut. **Crop Science**, v.33, n.4, p.859-861, 1993.

HU, J.; ZHU, J.; XU, H.M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Berlin: **Theoretical and Applied Genetics**, v.101, p.264-268, 2000.

JÚNIOR FERREIRA, J. A.; UNÊDA-TREVISOLI, S. H.; ESPINDOLA, S.M.C.G.; VIANA, V.F.; MAURO, A.O. Di. Diversidade genética em linhagens avançadas de soja oriundas de cruzamentos biparentais, quádruplos e óctuplos. **Revista Ciência Agronômica**, v.46, n.2, 2015.

KANG, M.S.; GAUCH JÚNIOR, H.M. Methods of testing inbred lines of maize in crossbred combinations. Maidison: **Journal of the American Society of Agronomy**, v.24, n.7, p.523-530, july 1932.

KARASAWA, M.; RODRIGUES, R.; SUDRÉ, C.P.; SILVA, M.P.; RIVA, E.M.; AMARAL JUNIOR, A.T. Aplicação de métodos de agrupamento na quantificação da divergência genética entre acessos de tomateiro. **Horticultura Brasileira**, v.23, n.4, p.1000-1005, 2005.

KLOSTER, G.S.; BARELLI, M.A.A.; SILVA, C.R.; NEVES, L.G.; PAIVA, SOBRINHO. S.; LUZ, P.B. Análise da divergência genética através de caracteres morfológicos em cultivares de feijoeiro. **Revista Brasileira de Ciências Agrárias**, v. 6, p. 452-459, 2011.

LAÍNEZ-MEJIA, J. R. **Implicações da interação genótipos x ambientes na seleção de progênies de soja com ênfase nas produtividades de grãos e óleo**. 1996. 145p. Tese de Doutorado. Universidade de São Paulo.

MALIK, M.F.A.; ASHRAF, M.; QURESHI, A.S.; GHAFOOR, A. Assessment of genetic variability, correlation and path analyses for yield and its components in soybean. **Pakistan Journal of Botany**, v.39, n.1, p. 405-413, 2007.

MANLY, B.F.J. **Randomization, bootstrap and Monte Carlo methods in Biology**. London: Chapman & Hall, 1997.

MARCONATO, M. B. **Diversidade fenotípica por meio de caracteres agronômicos em acessos de soja**. 2014. 50p. Dissertação. Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2014.

MARTINS, F.A.; CARNEIRO, P.C.S.; DA SILVA, D.J.H.; CRUZ, C.D.; CARNEIRO, J.E.S. Integração de dados em estudos de diversidade genética de tomateiro. **Pesquisa Agropecuária Brasileira**, v.46, n.11, p.1496-1502, 2011.

MIRANDA, Z.F.S.; ARIAS, C.A.A.; PRETE, C.E.C.; KIHIL, R.A. de S.; ALMEIDA, L.A. de A.; TOLEDO, J.F.F. de; DESTRO, D. Genetic characterization of ninety elite soybean cultivars using coefficient of parentage. **Pesquisa Agropecuária Brasileira**, v.42, p.363-396, 2007.

MOHAMMADI, S.; PRASANNA, B. Analysis of genetic diversity in crop plants salient statistical tools and considerations. **Crop Science**, v.43, n.4, p.1235-1248, 2003.

MORAES, R.M.A.; JOSE, I.C.; RAMOS, F.G.; BARROS, E.G. ; MOREIRA, M.A. Caracterização bioquímica de linhagens de soja com alto teor de proteína. Brasília: **Pesquisa Agropecuária Brasileira**, v.41, n.5, p.725-729, maio 2006.

MULATO, B.M.; MÖLLER, M.; ZUCCHI, M.I.; QUECINI, V.; PINHEIRO, J.P. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. Brasília: **Pesquisa Agropecuária Brasileira**, v.45, n.3, p.276-283, mar. 2010.

NOGUEIRA, A. P. O.; SEDIYAMA, T.; DE SOUSA, L. B.; HAMAWAKI, O. T.; CRUZ, C. D.; PEREIRA, D. G.; MATSUO, É. Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. **Bioscience Journal**, v. 28, n. 6, 2012.

OLIVEIRA, M.F. **Avaliação de cinco estratégias de amostragem para obtenção da coleção nuclear de soja (*Glycine max*(L.) Merrill)**. Piracicaba, Sp : ESALQ, 2007. 143p. Tese (Doutorado em Agronomia : Genética e Melhoramento de Plantas) – Escola superior de Agricultura ‘Luiz de Queiros’.

OLIVEIRA, M.S.P.; AMORIM, E.P.; SANTOS, J.B.; FERREIRA, D.F. Diversidade genética entre acessos de açaizeiro baseada em marcadores RAPD. **Ciência e Agrotecnologia**, v.31, n.6, p.1645-1653, 2007.

PEDROSO VAL, B.H.; FERREIRA JUNIOR, J.A.; BIZARI, E.H.; MAURO, A.O. DI.; TREVISOLI, S.H.U. Diversidade genética de genótipos de soja por meio de caracteres agromorfológicos. Jaboticabal (SP): **Ciência & Tecnologia**, FATEC-JB, v. 6, n. 1, p. 72-83, 2014.

PEIXOTO, N.; BRAZ, L.T.; BANZATTO, D.A.; MORAES, E.A.; MOREIRA, F.M. Características agronômicas, produtividade, qualidade de vagens e divergência genética em feijão-vagem de crescimento indeterminado. **Horticultura Brasileira**, v.20, n.3, p.447-451, 2002.

PELUZIO, J.M.; VAZ-DE-MELO, A.; AFFÉRI, F.S.; SILVA, R.R.; BARROS, H.B.; NASCIMENTO, I.R.; FIDELIS, R.R. Variabilidade genética entre cultivares de soja, sob diferentes condições edafoclimáticas. **Pesquisa Aplicada & Agrotecnologia**, v.2, p.21-29, 2009.

PENHA, L. A. O.; FONSECA, I. C. B.; MANDARINO, J. M.; BENASSI, V.T. A soja como alimento: valor nutricional, benefícios para a saúde e cultivo orgânico. Curitiba: **Boletim do Centro de Pesquisa e Processamento de Alimentos - B.CEPPA**. v. 25, n. 1, p. 91-102, jan./jun. 2007.

RAMOS, S.R.R.; QUEIROZ, M.A.; CASALI, V.W.D.; CRUZ, C.D. Divergência genética em germoplasma de abóbora procedente de diferentes áreas do Nordeste. **Horticultura Brasileira**, v.18, n.3, 2000.

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

RIGON, J.P.G.; CAPUANI, S.; NETO, J.F.B.; ROSA, G.M.; WASTOWSKI, A.D.; RIGON, C.A.G. Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. Viçosa : **Revista Ceres**, v. 59, n.2, p. 233-240, mar/abr, 2012.

ROHLF, J. F. **Numerical taxonomy and multivariate analysis system**: version 1.70. New York: Exeter software, 2000. 38p. (version 2.1).

ROSA, A.M.; CLAVISO, J.; PASSOS, L.; AGUIAR, C.L. Alimentos fermentados à base de soja (*Glycine max* (Merrill) L.): importância econômica, impacto na saúde e efeitos associados às isoflavonas e seus açúcares. **Revista Brasileira de Biociência**, Porto Alegre, v.7, n.4, out./dez. 2009.

ROSSMANN, H. **Estimativas de parâmetros genéticos e fenotípicos de uma população de soja avaliada em quatro anos**. 2001. 80p. Tese de Doutorado. Universidade de São Paulo.

SIHAG, R.; HOODA, J.S.; VASHISHTHA, R.D.; MALIK, B.P.S. Genetic divergence in soybean [*Glycine max*(L.) Merrill]. **Annual Biological**, v.20, n.1, p.17-21, 2004.
Silva, F. L.; Baffa, D. C. F.; de Oliveira, A. C. B.; Pereira, A. A.; Bonomo, V. S. Integração de dados quantitativos e multicategóricos na determinação da divergência genética entre acessos de cafeeiro. **Bragantia**, v. 72, n. 3, p. 224-229, 2013.

SNEDECOR, G.W.; COCHRAN, W.G. **Statistical methods**. 7.ed. Ames : Iowa State University, 1980. 507p.

SOUZA, V.Q.; NARDINO, M.; FOLLMANN, D.N.; BAHRY, C.A.; CARON, B.O.; ZIMMER, P.D. Caracteres morfofisiológicos e produtividade da soja em razão da desfolha no estágio vegetativo. Jaboticabal: **Científica : Revista de ciência Agrárias**, v.42, n.3, p.216–223, 2014.

SQUILASSI, M.G. **Interação de genótipos com ambientes**. Embrapa Tabuleiros Costeiros. 2003, 47p. Disponível em: <<http://www.cpatc.embrapa.br>> acesso em: 25/01/2016.

STELL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics: with reference to the biological sciences**. New York: McGraw-Hill, 1997. 666 p.

SUDRÉ, C.P.; CRUZ, C.D.; RODRIGUES, R.; RIVA, E.M.; AMARAL JUNIOR, A.T. SILVA, D.J H.; PEREIRA, T.N.S. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. **Horticultura Brasileira**, v.24, n.1, p.88-93, 2006.

TATINENI, V.; CANTRELL, R. G.; DAVIS, D. D. Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. Madison: **Crop Science**, v.36, p. 186-192, 1996.

TEODORO, P. E.; RIBEIRO, L. P.; CORRÊA, C. C. G.; da LUZ JÚNIOR, R. A. A.; dos SANTOS ZANUNCIO, A.; CAPRISTO, D. P.; TORRES, F. E. Path analysis in soybean genotypes as function of growth habit. **Bioscience Journal**, v. 31, n. 3, 2015.

TORRES, F.E.; DAVID, G.V.; TEODORO, P.E.; RIBEIRO, L.P.; CAIO G. CORREA, C.G.; LUZ JÚNIOR, R.A. Desempenho agronômico e dissimilaridade genética entre genótipos de soja. Lisboa: **Revista de Ciências Agrárias**, v.38 n.1, mar. 2015.

TSIVELIKAS, A.L.; KOUTITA, O.; ANASTASIADOU, A.; SKARACIS, G.N.; EKATERINI TRAKA-MAVRONA, E.; KOUTSIKA-SOTIRIOU, M. Description and Analysis of Genetic Diversity among Squash Accessions. **Brazilian Archives Biology Technology**, v.52 n.2: p. 271-283, Mar./Apr. 2009.

USDA. United States Department of Agriculture. **Approved by the World Agricultural Outlook Board**. February 9, 2016 Disponível em: <<http://www.usda.gov/oce/commodity/wasde/latest.pdf>> acesso em: 12/02/2016.

VAN HINTUM, T.J.L., BROWN, A.H.D., SPILLANE, C. & HODGKIN, T. **Core collection of plant genetic resources**. Roma: IPGRI Technical Bulletin n.3. International Plant Genetic Resources Institute, 2000. 48p.

VASCONCELOS, E.S. et al. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa Agropecuária Brasileira**, v. 42, n. 04, p. 507-514, 2007.

VENCOVSKY, R. Herança quantitativa. In: PATERNIANI, E. (Coord.). Melhoramento e a produção de milho no Brasil. Piracicaba: Fundação Cargill, 1987. p.137-214.

VIEIRA, C.R.; CABRAL, L.C.; PAULA, A.C.O. de. Proximate composition and amino acid, and fatty acid and mineral contents of six soybean cultivars for human consumption. **Pesquisa Agropecuária Brasileira**, v.34, p.1277-1283, 1999.

VIEIRA, E.S.N.; SCHUSTER,I.; SILVA, R.B.; OLIVEIRA, M.A.R. Variabilidade genética em cultivares de soja determinada com marcadores microssatélites em gel de agarose. Brasília: **Pesquisa Agropecuária Brasileira**, v.44, n.11, p.1460-1466, nov. 2009.

VITTI, G.C.; CAMARGO, M.A.F.; LARA, C. **Síntese de análise químicas em tecido vegetal**. Piracicaba: Departamento de Solos e Nutrição de Plantas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 2001.

WANG, J.; HU, J.; XU, H.; ZHANG, S. A strategy on constructing core collections by least distance stepwise sampling. **Theoretical and Applied Genetics**, v.115, n.1, p.1-8, 2007.

WILSON, R. F. Seed composition. In: BOERMA, H.R.; SPECHT, J. E. (Ed.). **Soybeans: Improvement, Production, and Uses**. 3. Madison: American Society of agronomy, Crop Science Society of American, and Soil Science Society of American, 2004. Cap. 3, p. 621- 677.

ZEWDIE, Y.; TONG, N.; BOSLAND, P. Establishing a core collection of capsicum using a cluster analysis with enlightened selection of accessions. **Genetic Resources and Crop Evolution**, v.51, n.2, p.147-151, 2004.