

RAFAEL PAULO DA SILVA

**RECONHECIMENTO DE PADRÕES E ORGANIZAÇÃO DA DIVERSIDADE
GENÉTICA EM ACESSOS DE SOJA (*Glycine max* (L.) Merrill) DO SUL DO BRASIL**

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

Orientador: Cosme Damião Cruz

Coorientador: Ivan Ricardo de Carvalho

**VIÇOSA – MINAS GERAIS
2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S586r
2023
Silva, Rafael Paulo da, 1997-
Reconhecimento de padrões e organização da diversidade genética em acessos de soja (*Glycine max* (L.) Merrill) do Sul do Brasil / Rafael Paulo da Silva. – Viçosa, MG, 2023.
1 dissertação eletrônica (61 f.): il. (algumas color.).

Orientador: Cosme Damião Cruz.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Biologia Geral, 2023.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2023.286>

Modo de acesso: World Wide Web.

1. Soja - Melhoramento genético - Métodos estatísticos.
2. Soja - Melhoramento genético - Brasil, Sul. I. Cruz, Cosme Damião, 1958-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 631.52

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

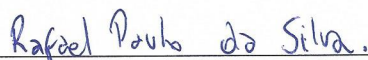
RAFAEL PAULO DA SILVA

**RECONHECIMENTO DE PADRÕES E ORGANIZAÇÃO DA DIVERSIDADE
GENÉTICA EM ACESSOS DE SOJA (*Glycine max* (L.) Merrill) DO SUL DO BRASIL**

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

APROVADA: 25 de fevereiro de 2023.

Assentimento:



Rafael Paulo da Silva
Autor



Cosme Damião Cruz
Orientador

Aos meus pais e irmãs.

AGRADECIMENTOS

Agradeço primeiramente a Deus por todas as graças alcançadas e pela oportunidade de estabelecer relações importantes nessa vida.

Aos meus pais Elzira Maria da Silva e Jandir Lino da Silva, às minhas irmãs Ivanete, Joziane, Raquel e Sueli, ao Samuel, e aos meus sobrinhos Maria Eduarda, João Lucas, João Miguel, Gustavo, Gabrielly, que foram minha base e sempre estiveram comigo, entenderam minhas ausências e me proporcionaram todo amor, sinceridade, carinho e tantas alegrias.

Ao professor Cosme Damião Cruz, orientador, educador e exemplo a ser seguido. Obrigado pela oportunidade de continuar a aprender com o senhor, pela confiança e paciência, e pelos bons momentos de conversa.

Ao professor Ivan Ricardo de Carvalho, coorientador, educador e um grande amigo, sempre disposto e proativo em me ajudar sempre que precisei.

Ao Laboratório de Bioinformática, que virou minha segunda casa durante a graduação e o mestrado.

Aos amigos adquiridos durante a graduação e a pós-graduação, por todas as aventuras, conversas, risos e desafios enfrentados juntos, vocês farão sempre parte de mim também.

À Universidade Federal de Viçosa (UFV) que me permitiu concluir mais uma etapa acadêmica.

À Universidade Regional do Noroeste do Rio Grande do Sul (UNIJUÍ), por disponibilizar o banco de dados e por disponibilizar meu coorientador.

Às agências financiadoras CNPq, CAPES e FAPEMIG, que contribuíram para a realização desse projeto.

À banca por aceitarem meu convite e pela participação.

RESUMO

SILVA, Rafael Paulo da, M.Sc., Universidade Federal De Viçosa, fevereiro de 2023. **Reconhecimento de padrões e organização da diversidade em acessos de soja (*Glycine max* (L.) Merrill) do sul do Brasil.** Orientador: Cosme Damião Cruz. Coorientador: Ivan Ricardo de Carvalho.

A soja (*Glycine max* (L.) Merrill) é uma cultura anual, herbácea com ciclo de 70 a 100 dias e autógama. A partir dos anos 2000 o cultivo da soja aumentou em todo o território brasileiro e, tornando o Brasil o principal produtor desta cultura no mundo. No caso das plantas autógamas perfeitas, como a soja, o programa de melhoramento é conduzido por meio de hibridações artificiais e posteriores avanços das populações segregantes, por autofecundações, visando a obtenção de melhores genótipos. Tendo em vista a extensão territorial do Brasil e a grande diversidade edafoclimática, a busca por genótipos adequados a cada ambiente se faz necessária. O objetivo do trabalho foi caracterizar e estimar o potencial produtivo e a diversidade de genótipos de soja com base em caracteres quantitativos. Foram utilizadas quatro diferentes coleções totalizando 301 genótipos, provindos da Universidade Regional Noroeste do Rio Grande do Sul (UNIJUÍ). As características foram massa de cem sementes (MCS), proteína (PTN), óleo (OLE), Fibra (FIB), Cinza (CIN), Ácido Palmítico (APA), Ácido Esteárico (AES), Ácido Oleico (AOL), Ácido Linoleico (ALI), Ácido Linolênico (ALN). Todas as análises foram realizadas utilizando o software Genes. O potencial foi avaliado através da análise descritiva das características e aplicado o teste de Tukey e a estatística de Lilliefors, depois os dados foram submetidos à uma análise de correlação de Pearson. A rede de correlações foi gerada através da matriz de distância euclidiana, e por fim os melhores genótipos foram analisados de acordo com Diagramas de Venn. A diversidade foi fundamentada através da análise discriminante e agrupamento pelo método de Tocher. A distribuição da variabilidade entre e dentro das coleções foi estimada pelo método da AMOVA e a discriminação dos genótipos foi realizada pelas funções linear e quadrática de Anderson. As análises mostraram que a MCS esteve inferior, média de 15.66, requerendo esforços do melhorista afim de elevar esta característica. Além disso, dentre os constituintes da semente, apenas PTN se apresentou baixo, com média no patamar de 33,40%, PTN e OLE apresentaram curva de distribuição próximo da Normal, indicando possível controle poligênico de natureza aditiva. A correlação mais expressiva foi entre PTN e OLE, valor de -0,7, e já é muito conhecida. MCS se correlacionou positivamente com PTN e negativamente com FIB, valores de 0,18 e 0,31 respectivamente. As redes de correlações individualizadas mostraram que a coleção 4 apresentou maior número de interações

com alta intensidade, sejam elas positivas ou negativas e a coleção 3 foi a que apresentou a menor. A coleção de interesse, coleção 3 apresentou genótipos semelhantes entre si, 3 e 4, e 20 e 38, e ambos se mostraram divergentes do genótipo 38, e a população base apresentou bons genótipos para PTN e OLE, e MCS e OLE. Os diagramas apresentados foram concordantes em apontar a dificuldade de evidenciar genótipos que tenham bom desempenho considerando simultaneamente as características MCS, PTN e OLE. A análise de componentes principais mostrou dois componentes foram suficientes para explicar 88.38% da variância disponível nas características analisadas e a dispersão gráfica evidenciou que as coleções 3 e 4 são dissimilares. Nenhum grupo foi estabelecido com exclusividade ou prevalência de uma determinada coleção pelo agrupamento de Tocher. A AMOVA evidenciou que a coleção 1 é mais heterogênea e a coleção 3 é mais homogênea, além disso, 80.45% da variação encontrada está dentro das coleções. As análises discriminantes de Anderson tiveram taxa de erro de 26.44% e 19.96% para as análises linear e quadrática, respectivamente, evidenciando que determinados fatores podem dificultar a classificação correta das coleções. Sendo assim concluiu-se que os genótipos avaliados apresentaram bom potencial em termos de PTN, atingindo valor médios de 33,40%, porém a coleção 3 apresentou 34.29% de PTN, com baixa variabilidade exigindo esforços adicionais para incluir formas alélicas mais favoráveis recorrendo a genótipos de outras coleções, e foram identificados genótipos de bom desempenho que reúnam características de interesse tais como MCS e PTN ou MCS e OLE. Associado a isto, existem coleções que podem ser diferenciadas e com concentração de alelos favoráveis para características de qualidade industrial diversificada.

Palavras-chave: Produtividade. Caracteres quantitativos. Melhoramento genético. Diagramas de Venn.

ABSTRACT

SILVA, Rafael Paulo da, M.Sc., Universidade Federal De Viçosa, February 2023. **Pattern recognition and organization of diversity in soybean (*Glycine max* (L.) Merrill) accessions from southern Brazil**. Advisor: Cosme Damião Cruz. Co-advisor: Ivan Ricardo de Carvalho.

Soybean (*Glycine max* (L.) Merrill) is an annual, herbaceous crop with a cycle of 70 to 100 days and autogamous. From the 2000s the cultivation of soybeans increased throughout the Brazilian territory and, making Brazil the main producer of this crop in the world. In the case of perfect autogamous plants, such as soybeans, the breeding program is conducted through artificial hybridizations and subsequent advances of the segregating populations, by self-fertilization, aiming at obtaining better genotypes. In view of the territorial extension of Brazil and the great edaphoclimatic diversity, the search for genotypes appropriate to each environment is necessary. The objective of this work was to characterize and estimate the productive potential and diversity of soybean genotypes based on quantitative traits. Four different collections were used, totaling 301 genotypes, from the Northwest Regional University of Rio Grande do Sul (UNIJUÍ). The characteristics were mass of one hundred seeds (MCS), protein (PTN), oil (OLE), Fiber (FIB), Ash (CIN), Palmitic Acid (APA), Stearic Acid (AES), Oleic Acid (AOL), Linoleic Acid (ALI), Linolenic Acid (ALN). All analyses were performed using the Genes software. The potential was evaluated through the descriptive analysis of the characteristics and applied the Tukey test and the Lilliefors statistic, then the data were submitted to a Pearson correlation analysis. The correlation network was generated through the Euclidean distance matrix, and finally the best genotypes were analyzed according to Venn diagrams. Diversity was based on discriminant analysis and grouping by the Tocher method. The distribution of variability between and within the collections was estimated by the AMOVA method and genotype discrimination was performed by Anderson's linear and quadratic functions. The analyses showed that the MCS was lower, with an average of 15.66, requiring efforts from the breeder in order to raise this characteristic. In addition, among the constituents of the seed, only PTN was low, with an average of 33.40%, PTN and OLE presented a distribution curve close to normal, indicating possible polygenic control of an additive nature. The most significant correlation was between PTN and OLE, a value of -0.7, and is already well known. MCS correlated positively with PTN and negatively with FIB, values of 0.18 and 0.31 respectively. The individualized correlation networks showed that collection 4 presented the highest number of interactions with high intensity, whether positive

or negative, and collection 3 presented the lowest. The collection of interest, collection 3 presented genotypes similar to each other, 3 and 4, and 20 and 38, and both were divergent from genotype 38, and the base population presented good genotypes for PTN and OLE, and MCS and OLE. The diagrams presented were in agreement in pointing out the difficulty of evidencing genotypes that have good performance considering simultaneously the characteristics MCS, PTN and OLE. The principal component analysis showed two components were sufficient to explain 88.38% of the variance available in the analyzed characteristics and the graphic dispersion showed that collections 3 and 4 are dissimilar. No group was established with exclusivity or prevalence of a given collection by the Tocher grouping. AMOVA showed that collection 1 is more heterogeneous and collection 3 is more homogeneous, in addition, 80.45% of the variation found is within the collections. Anderson's discriminant analyses had an error rate of 26.44% and 19.96% for the linear and quadratic analyses, respectively, evidencing that certain factors may hinder the correct classification of the collections. Thus, it was concluded that the genotypes evaluated presented good potential in terms of PTN, reaching a mean value of 33.40%, but collection 3 presented 34.29% of PTN, with low variability requiring additional efforts to include more favorable allelic forms using genotypes from other collections, and genotypes of good performance that gather characteristics of interest such as MCS and PTN or MCS and OLE were identified. Associated with this, there are collections that can be differentiated and with concentration of alleles favorable for characteristics of diversified industrial quality.

Keywords: Productivity. Quantitative characters. Breeding. Venn diagrams.

SUMÁRIO

1. INTRODUÇÃO GERAL	10
1. REVISÃO BIBLIOGRÁFICA	11
1.1. Soja (<i>Glycine max</i> (L.) Merrill)	11
1.2. Diversidade Genética	13
1.3. Diversidade Genética Aplicada Ao Melhoramento de Soja	13
1.4. Análise classificatória	14
1.5. Diagrama de Venn no melhoramento genético	15
2. REFERÊNCIAS BIBLIOGRÁFIAS	16
CAPÍTULO 1: Potencial de progênies da soja oriundas do sul do Brasil	20
1. INTRODUÇÃO	22
2. MATERIAIS E MÉTODOS	24
2.1. Material genético utilizado	24
2.2. Caracteres Agronômicos avaliados.....	24
2.3. Estatísticas Descritivas.....	25
2.4. Associação entre características.....	25
2.5. Rede de correlações.....	26
3. RESULTADOS E DISCUSSÃO	27
4. CONCLUSÃO	39
5. REFERÊNCIAS BIBLIOGRÁFICAS	39
CAPÍTULO 2: Estudo da diversidade genética entre e dentro de coleções de soja do sul do Brasil	43
1. INTRODUÇÃO	45
2. MATERIAL E MÉTODOS	46
2.1. Descrição do Experimento.....	46
2.2. Caracteres Agronômicos avaliados.....	47
2.3. Procedimentos Biométricos	47
3. RESULTADOS E DISCUSSÃO	50
4. CONCLUSÃO	59
5. REFERÊNCIAS BIBLIOGRÁFICAS	60

1. INTRODUÇÃO GERAL

No Brasil, a taxa de crescimento geométrica de produção da soja (*Glycine max* (L.) Merrill), foi de 6,2%, triplicando a quantidade colhida entre os anos de 2000 e 2018, chegando a produção de 119,3 milhões de toneladas neste último ano (HIRAKURI, 2020). A partir dos anos 2000, com o aumento do poder de compra principalmente nos países emergentes, a demanda por alimentos, principalmente a proteína animal cresceu significativamente (SEIXAS *et al.* 2020). Aliado a isto, a safra de 2022/2023 apresentou previsão de produção em 152,71 milhões de toneladas (CONAB, 2023), sendo utilizada para a produção de grãos, farelo e óleo, atendendo a indústria alimentícia e agropecuária. Ademais, associado à grande extensão territorial brasileira, o melhoramento genético da soja, seja ele feito por empresas públicas e privadas têm buscado cada vez mais as variedades mais produtivas e adaptadas, e começou por volta de 1882, ano de chegada da cultura ao solo brasileiro (SILVA *et al.*, 2017). Segundo Carvalho *et al.* (2017) o sucesso da cadeia produtiva da cultura da soja no Brasil se deve ao desenvolvimento de novas cultivares e a disponibilização destas no mercado, atendendo as necessidades reais ou potenciais do mesmo. A expansão da cultura da região sul do país para as regiões Centro-Oeste, Norte e Nordeste exigiu dos programas de melhoramento o desenvolvimento e a seleção de cultivares com período juvenil mais estendido, que aliado à resistência a doenças e aos eventos transgênicos, resultou no aumento da produtividade da soja (SILVA *et al.*, 2021). Demonstrando assim a importância da diversidade de genótipos, que se adaptam as diferentes condições edafoclimáticas e apresentem uma série de atributos favoráveis simultaneamente, além de satisfazerem as exigências que o consumidor apresentar (CRUZ *et al.*, 2017; VIANNA *et al.*, 2019).

Os estudos de diversidade genética são importantes desde a fase de pré-melhoramento. Segundo Cruz, Ferreira e Pessoni (2011), atualmente eles podem ser realizados com informações fenotípicas e moleculares – tanto dentro quanto entre populações coleções e ecótipos – e são realizados por meio do estudo de caracteres fenotípicos, sendo a maioria deles caracteres quantitativos, apresentando geralmente, distribuição contínua e sendo controlados por poligenes com pequeno efeito e alta influência ambiental. A importância de tais estudos reside na diversidade existente. As análises de diversidade genética identificam genitores adequados para a obtenção de híbridos com alto efeito heterótico e que proporcionem maior segregação, possibilitando assim o surgimento de transgressivos (CRUZ; REGAZZI; CARNEIRO, 2012). Entre e dentro das populações nos programas, a diversidade genética pode ser predita a partir da diferença entre os valores fenotípicos obtidos, e técnicas

multivariadas como análise de componentes principais (PCA), análise discriminante e análises de agrupamento podem ser utilizadas neste tipo de estudo (CRUZ; FERREIRA; PESSONI, 2011). Além disso, os estudos de diversidade genética da soja têm gerado informações importantes para o gerenciamento e a avaliação de germoplasmas. Geralmente eles são feitos baseando-se em informações agronômicas e morfológicas, além de genealógicas (SILVA *et al.*, 2017).

Algumas questões importantes, que permeiam o estudo da diversidade, relacionam-se ao reconhecimento de padrões, a potencialidade de se estabelecer critérios acurados de classificação dos indivíduos dentro de suas coleções e a organização da diversidade para fins de uso recorrente. Características como a produtividade de grãos e qualidade dos grãos de soja tem sido de interesse de diversos grupos de melhoramento, mas algumas vezes grupos de pesquisas priorizam o aumento da produtividade sem tutoramento das consequências sobre as características de qualidade estabelecendo, por esta razão, conjuntos de acessos com maior ou menor riqueza de formas alélicas determinantes das características que não foram privilegiadas na seleção. Além disso, estudos de classificação, fundamentados em análises discriminantes, também se destacam e são muito utilizados em diferentes espécies. Quando se tem um conjunto de acessos que representam a reunião de diferentes coleções é possível postular a hipótese que eles apresentam nítida diferenciação para as características que foram alvos diretos ou indiretos no processo de obtenção, condução e desenvolvimento dos genótipos das referidas coleções. Entretanto, dado a globalização e a conectividade entre grupos de pesquisa, também pode-se formular hipóteses de que coleções de origem diferente conservam grande similaridade em razão de intercâmbios genéticos e combinações gênicas de genótipos de origem diferenciada.

De certa forma estudar a diversidade, reconhecer padrões e apresentar modelos que justifiquem a diferenciação entre coleções e classificação dos indivíduos dentro de suas respectivas coleções são de interesse nos programas de melhoramento. Entretanto, deve-se acrescentar a estes estudos informações sobre a organização da diversidade de forma que o padrão de proximidade entre indivíduos ou coleções possa ser reconhecido permitindo ampliar grupos, permutar acessos dentro de grupos e otimizar futuras reamostragem para fins de enriquecimento da coleção já disponível.

1. REVISÃO BIBLIOGRÁFICA

1.1. Soja (*Glycine max* (L.) Merrill)

A soja (*Glycine max* (L.) Merrill) é uma cultura anual, herbácea, ciclo de 70 a 100 dias e autógama. O centro de origem é no continente asiático, mais especificamente na região da Manchúria. No Brasil, foi introduzida em 1882 no estado da Bahia, e devido as condições edafoclimáticas do Rio Grande do Sul, se adaptou melhor nesta região devido a semelhança ao local de origem (SEDIYAMA; SILVA; BORÉM, 2015). O aumento da demanda pelo grão, principalmente para a produção de óleo e para fins de alimentação animal, a expansão da cultura no país foi acelerada, devido a seleção de genótipos com período juvenil longo e adaptados as condições edafoclimáticas (PELUZIO *et al.*, 2021).

O Brasil é uns dos principais países produtores de soja no mundo, e em relação ao ano anterior apresentou o aumento da área plantada de 3,8% em relação ao ano anterior. Isso é justificado devido ao desenvolvimento de variedade adaptadas a diferentes condições edafoclimática (CONAB, 2022). A soja é considerada uma oleaginosa de suma importância para o agronegócio brasileiro, e a fronteira agrícola do seu cultivo se espalha cada vez mais pelo país, principalmente nas direções do Norte e Nordeste (PELUZIO *et al.*, 2021). Seu cultivo como fonte de proteína e óleo vegetal tornou a soja uma das principais culturas do mundo, mesmo não sendo considerada como um alimento básico, como no caso dos cereais trigo, arroz e milho (SEDIYAMA; SILVA; BORÉM, 2015).

Considerando a extensão territorial, a produção da soja é desuniforme no Brasil, principalmente por causa da diversidade de características de cada região. Elas influenciam na produção da oleaginosa, como por exemplo a escassez hídrica, que pode levar a falhas de germinação, murcha foliar e aparecimento de pragas, o que ocorreu em estados como o Rio Grande do Sul, Paraná Mato Grosso do Sul (CONAB, 2022). Em contraste, o excesso de precipitação e umidade podem levar à perda de produtividade e o aparecimento de pragas, o que ocorreu nos estados de Goiás e Maranhão (CONAB, 2022).

De acordo com as características de cada região, cada programa de melhoramento define diferentes estratégias de seleção. No melhoramento genético de plantas, a seleção dos melhores genótipos é feita com base em suas características comerciais e agronômicas mais importantes, e isso ocorre na condução de programas de melhoramento de plantas autógamas perfeitas, como no caso da soja (LEITE *et al.*, 2020). Além disso, as características fisiológicas da soja apresentam alta correlação com a produtividade da planta, tornando-a alvo de programas de melhoramento que utilizam seleção indireta (TODESCHINI *et al.*, 2019).

Atualmente, os programas de melhoramento buscam a incrementação da estabilidade fenotípica, possíveis resistências e tolerâncias, além de aspectos relacionados ao rendimento (CARVALHO; MAMBRIN, 2022). Para isso, eles utilizam diferentes métodos de melhoramento, a depender do sistema reprodutivo e objetivos do programa. Dentre os métodos mais comuns para o melhoramento de soja pode-se destacar a introdução de germoplasma exógeno, indução de mutação - podendo ser química ou física, seleção direta ou indireta e população de família segregantes, entre outras (CARVALHO; MAMBRIN, 2022).

1.2. Diversidade Genética

A diversidade genética se refere ao grau de diferenciação dos genes em uma espécie/população. Inicialmente, os estudos de diversidade genética fundamentavam-se em informações fenotípicas de características morfológicas, ou ainda em relação a características de desempenho zootécnico e/ou agrônomico (CRUZ; FERREIRA; PESSONI, 2011). Ademais, com o avanço da pesquisa, principalmente na área molecular, novas abordagens foram possíveis, uma vez que a utilização de marcadores moleculares possibilita a detecção de variabilidade a nível de DNA (CRUZ; FERREIRA; PESSONI, 2011).

Os estudos de diversidade genética são extremamente importantes e podem ser utilizados com diversas finalidades. De acordo com os objetivos eles podem ser classificados em estudos de melhoramento genético, associações evolutivas, ou ainda, para conservação e manejo de material genético (CRUZ; FERREIRA; PESSONI, 2011). Nos estudos de melhoramento genético, a diversidade entre grupo de progenitores é realizada com o objetivo de identificar combinações de híbridos com maior heterozigose e efeito heterótico (CRUZ; REGAZZI; CARNEIRO, 2012; CRUZ *et al.*, 2022).

As análises preditivas da diversidade genética também são possíveis. Segundo Cruz, Regazzi e Carneiro (2012) a divergência genética vem sendo estimada através de técnicas multivariadas, sendo elas a análise de componentes principais, variáveis canônicas ou métodos aglomerativos, sendo cada tipo de análise definido de acordo com o banco de dados. Métodos preditivos de forma exploratória são baseados na relação entre a heterose e os desvios de dominância juntamente com a diferença da frequência gênica entre as populações (SILVA *et al.*, 2017).

1.3. Diversidade Genética Aplicada Ao Melhoramento de Soja

Existem diversos estudos de diversidade genética aplicadas ao melhoramento de soja. Geralmente os estudos de natureza preditiva são recomendados quando o pesquisador possui

elevado número de genitores, podendo orientar assim, quantos e quais cruzamentos são mais indicados para a obtenção de híbridos (BHERING; PEIXOTO; CRUZ, 2017). Fato importante na escolha dos genitores é a dissimilaridade entre eles. No caso da soja, a erosão genética causada pelo uso de pouco acessos para obter novos cultivares foi impactante, existindo um número reduzido de acessos que explicam grande parte da variabilidade, levando a uma estagnação na produtividade de grãos (SILVA *et al.*, 2021).

Entretanto a perda de variabilidade genética da soja pode ser contornada com diversas estratégias. A principal estratégia segundo Silva *et al.* (2021) está no cruzamento de cultivares-elite brasileiros com aqueles oriundos do Japão e da China, com a finalidade de ampliar a base genética, possibilitando o aumento da produtividade. Outra estratégia é a criação de pan-genomas, ou seja, a integração de parentes selvagens da cultura afim de formar um gênero com uma diversificação maior de genes, podendo assim recuperar a variabilidade perdida durante a domesticação e melhoramento (TORKAMANEH; LEMAY; BELZILE, 2021).

Os estudos de diversidade genética de soja abrangem diversas características. A Krisnawati e Adie (2019), analisaram a variabilidade genética de soja para a resistência da quebra de vagens. Outros estudos relacionados a produção de biodiesel (OLIVEIRA *et al.* 2019), morfometria reprodutiva (COSTA *et al.*, 2018), tolerância de brotação pré-colheita (KUMAR *et al.*, 2021) e aspetos fitossanitários e nutricionais (LORO *et al.*, 2021).

1.4. Análise classificatória

A análise classificatória tem como objetivo a obtenção de funções que consigam classificar um indivíduo em uma população conhecida com base em um conjunto de dados sobre informações mensuradas, a fim de minimizar a classificação errônea (SANT'ANNA *et al.*, 2018; SKOWRONSKI, 2021; CRUZ *et al.*, 2022). Essa análise leva em consideração o conjunto total de dados, que é dividido em 80% para treinamento e 20% para validação, que é a classificação dos indivíduos (AMARAL *et al.*, 2022). A partir do momento em que indivíduos são classificados corretamente, tais funções podem ser utilizadas para a classificação de indivíduos desconhecidos (SANT'ANNA *et al.*, 2018).

A análise discriminante proposta por Anderson em 1958 gera funções resultantes da combinação linear de características previamente avaliadas a respeito dos indivíduos sabidamente pertencentes a populações diferentes (SANT'ANNA *et al.*, 2018; CRUZ *et al.*, 2022). Amaral *et al.* (2022) encontraram uma Taxa de Erro Aparente (AER) de 58,6% e 50,59% ao utilizar as metodologias de Fisher e Anderson respectivamente, para classificar genótipos

referentes a 11 populações de diferentes grupos de maturidade relativa de soja. Esses autores argumentam que diversos fatores podem levar a classificações errôneas das análises discriminantes. Destaca-se a grande semelhança em genealogia e origem das populações, número insuficiente de variáveis analisadas, baixa qualidade discriminatória, e a utilização de uma abordagem estatística inadequada (AMARAL *et al.*, 2022).

Oliveira *et al.* (2021) encontraram taxa de acerto de 93,33 % ao classificar grupos varietais de *Coffea canephora*, Nogueira *et al.* (2008) encontraram AER de 12,73% e 9,09% e (1,82% e 16,36%) respectivamente para classificação em de indivíduos de soja em 11 cultivares diferentes de acordo com quatro épocas de semeadura. BIANCHINI *et al.* (2020) obtiveram uma taxa 63,5% de classificação correta avaliando o rendimento da cultura da soja sob quatro diferentes tratamentos, 100% de acerto para classificação quanto ao controle de ervas daninhas na cobertura com chicória forrageira, 87,5 % de acerto quanto ao controle de ervas daninhas com cobertura de quinoa, e por fim 75% de acerto para todas as técnicas de manejo utilizadas no estudo. Sá *et al.* (2022) também realizaram um estudo em que utilizaram Redes Neurais Artificiais (RNAs) para classificar genótipos de soja a fim de selecionar os melhores pais e proporem um melhor método de utilização de Mapas Auto-organizáveis de Kohonen, testando sua eficiência utilizando análise classificatória de Anderson e conseguindo um valor maior que 99% de assertividade.

1.5. Diagrama de Venn no melhoramento genético

O uso de metodologias adequadas para a visualização correta de dados apresenta grande importância nas mais diversas áreas e diversos tipos de análises. Venn (1880) propôs uma forma de visualização simples e clara de sobreposições de conjunto de dados. Hoje, com o avanço da bioinformática, principalmente nas áreas de proteômica, genômica e transcriptômica, os diagramas de Venn são utilizados para o descobrimento de genes centrais em grandes conjuntos de dados, na comparação de genes sobrepostos, no compartilhamento de proteínas em diferentes tratamentos e na comparação de conjuntos de dados (JIA; XU; WANG, 2021).

O Diagrama de Venn é um recurso gráfico utilizado para representar associações, interseções e comparações. Inúmeras são suas aplicações, e vem sendo utilizado principalmente em estudos nas áreas de genética molecular e bioinformática, e utilizado para diversas culturas. Silva Filho (2019) utilizou diagramas para mostrar a regulação de fatores de transcrição em *Jatropha curcas* L, Godoy Filho (2020) mostrou o número de SNPs e indels compartilhados por duas cultivares de soja, Solar (2022) verificou quais genótipos de arroz estudados eram

significativamente diferentes da cultivar BRS Pampeira, Pereira (2022) mostrou diversas análises de semelhanças e diferenças entre genótipos e tratamentos para a cultura do algodoeiro.

Escassos são os estudos utilizando tal metodologia para estudos de diversidade genética dentro do melhoramento. Porém, autores como Darkawa *et al.* (2020) utilizaram diagramas de Venn para agrupar em três clusters 173 acessos de inhame branco de Guiné (*Dioscorea rotundata*) utilizando informações fenotípicas e de marcadores moleculares. Fonsêca (2019) mostraram a distribuição de alelos entre bancos de germoplasmas de maracujá (*Passiflora spp.*).

2. REFERÊNCIAS BIBLIOGRÁFIAS

A KRISNAWATI; ADIE, M M. Genetic variability of soybean (*Glycine max* L. Merrill) genotypes for pod shattering resistance. **Iop Conference Series: Earth and Environmental Science**, [S.L.], v. 293, n. 1, p. 012003, 1 jun. 2019. IOP Publishing. <http://dx.doi.org/10.1088/1755-1315/293/1/012003>.

AMARAL, L. O. *et al.* Artificial Neural Network for Discrimination and Classification of Tropical Soybean Genotypes of Different Relative Maturity Groups. **Frontiers In Plant Science**, [S.L.], v. 13, 12 jul. 2022. Frontiers Media SA. <http://dx.doi.org/10.3389/fpls.2022.814046>.

BEHRING, Leonardo Lopes; PEIXOTO, Leonardo de Azevedo; CRUZ, Cosme Damião. Seleção de Genitores. In: SILVA, Felipe *et al.* **Melhoramento da Soja**. Viçosa, MG: Ed. UFV, 2017. Cap. 8. p. 174-195.

CARVALHO, I.R.; NARDINO, M.; SOUZA, V.Q. **Melhoramento e Cultivo da Soja**. Editora. 2017.

CARVALHO, Ivan Ricardo de; MAMBRIN, Riteli Baptista. Estratégias de Melhoramento Genético para a Soja. In: MARTIN, Thomas Newton; PIRES, João Leonardo Fernandes; VEY, Rosana Taschetto. **Tecnologias aplicadas para o manejo rentável e eficiente da cultura da soja**. Santa Maria: Editora Gr, 2022. Cap. 2. p. 105-126.

CONAB. CAMPANHA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira: Grãos. Brasília, DF. V. 9 – safra 21/22 – N. 4. Quarto levantamento, janeiro. 2022.

CONAB. CAMPANHA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira: grãos. Brasília, DF. V10 – safra 2022/23 – N. 4. Quarto levantamento, janeiro. 2023.

COSTA, M. I. A; MELO, L. A; FERREIRA, S. C; MATSUO, E. REPRODUCTIVE MORPHOMETRY AND GENETIC DIVERSITY IN SOYBEAN CULTIVARS. **Nucleus**, [S.L.], v. 15, n. 2, p. 207-215, 30 out. 2018. Fundacao Educational de Ituverava. <http://dx.doi.org/10.3738/1982.2278.2894>.

CRUZ, Cosme Damião *et al.* Biometria aplicada ao melhoramento genético. In: SILVA, Felipe *et al.* **Melhoramento da Soja**. Viçosa, MG: Ed. UFV, 2017. Cap. 11. p. 249-287.

CRUZ, Cosme Damião *et al.* Classificação: abordagens de aprendizado de máquinas e inteligência computacional. In: NASCIMENTO, Moysés *et al.* **ABORDAGENS BIOMÉTRICAS PARA RECONHECIMENTO DE PADRÕES, CLASSIFICAÇÃO E PREDIÇÃO NAS CIÊNCIAS AGRÁRIAS**. Londrina, PR: Mecenas, 2022. Cap. 5. p. 103-166.

CRUZ, Cosme Damião; FERREIRA, Fábio Medeiros; PESSONI, Luiz Alberto. **Biometria Aplicada ao Estudo de Diversidade Genética**. Visconde do Rio Branco, Mg: Suprema, 2011. 620 p.

CRUZ, Cosme Damião; REGAZZI, Adair José; CARNEIRO, Pedro Crescêncio Souza. **Modelos Biométricos Aplicados ao Melhoramento do Genético de Plantas**. 4. ed. Viçosa, Mg: Ed. UFV, 2012. 514 p.

DARKWA, K. *et al.* Comparative assessment of genetic diversity matrices and clustering methods in white Guinea yam (*Dioscorea rotundata*) based on morphological and molecular markers. **Scientific reports**, v. 10, n. 1, p. 13191, 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41598-020-69925-9>

FONSÊCA, Thayse Karollyne dos Santos. **Diversidade e estrutura genética em acessos de maracujazeiros (*Passiflora spp.*) por meio de Regiões Análogas a Genes de Resistência**. 2019. 61 f. Dissertação (Mestrado) - Curso de Ciências Ambientais, Ciências Ambientais, Universidade Estadual do Sudoeste da Bahia, Itapetinga, 2019.

GODOY FILHO, Marcos Antônio de. **Montagem *De novo* de genomas contrastantes de soja para estudos de resistência ao complexo de percevejos**. 2020. 71 f. Dissertação (Mestrado) - Curso de Ciências, Genética e Melhoramento de Plantas, Universidade de São Paulo, Piracicaba, 2020.

HIRAKURI, Marcelo Hiroshi. O contexto econômico da produção de soja. In: SEIXAS, Claudine Dinali Santos *et al.* **Tecnologias de produção de soja**. 17. ed. Londrina, PR: Embrapa Soja, 2020. Cap. 1. p. 15-32. (Sistema de Produção).

JIA, A; XU, L; WANG, Y. Venn diagrams in bioinformatics. **Briefings In Bioinformatics**, [S.L.], v. 22, n. 5, 12 abr. 2021. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bib/bbab108>.

Kumar, Praveen & Pallavi, Mandalapu & Naguri, Swapna & Shahana, Firdoz & Reddy, G & Rakesh, Gojoori. **Genetic Variability and Correlation Studies for Pre-harvest Sprouting Tolerance and Associated Traits in Soybean** [*Glycine max* L. Merrill.]. 1-10. 2021.

LEITE, D. C. *et al.* Non-destructive genotypes classification and oil content prediction using near-infrared spectroscopy and chemometric tools in soybean breeding program. **Journal Of Food Composition And Analysis**, [S.L.], v. 91, p. 103536, ago. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.jfca.2020.103536>.

LORO, M. V. *et al.* ARTIFICIAL INTELLIGENCE AND MULTIPLE MODELS APPLIED TO PHYTOSANITARY AND NUTRITIONAL ASPECTS THAT INTERFER IN THE PHYSIOLOGICAL POTENTIAL OF SOYBEAN SEEDS. **Brazilian Journal Of Agriculture - Revista de Agricultura**, [S.L.], v. 96, n. 1, p. 324-338, 14 abr. 2021. Brazilian Journal of Agriculture. <http://dx.doi.org/10.37856/bja.v96i1.4258>.

NOGUEIRA, A. P. O. *et al.* Novas características para diferenciação de cultivares de soja pela análise discriminante. **Ciência Rural**, [S.L.], v. 38, n. 9, p. 2427-2433, dez. 2008. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-84782008000900003>.

Oliveira, E. C. A. M., Junior, W. P. O., Oliveira, J. D. D., Furnigare, N. S. & Peluzio, J. M. (2019). Divergência genética em linhagens de soja visando a produção de biodiesel no Estado do Tocantins. **Magistra**, 30, 113– 122.

OLIVEIRA, G. F. *et al.* DISCRIMINATION OF VARIETAL GROUPS AND HYBRIDS OF COFFEA CANEPHORA SPECIES USING MULTIVARIATE ANALYSIS. **Revista Brasileira de Biometria**, [S.L.], v. 39, n. 1, p. 194-205, 31 mar. 2021. Revista Brasileira de Biometria. <http://dx.doi.org/10.28951/rbb.v39i1.525>.

PELUZIO, Joanês Mucci; CARVALHO, Edmar Vinícius de; AFFÉRI, Flávio Sérgio; ALMEIDA, Bernardo Coutinho de. Histórico e progressos da cadeia produtiva da soja. In: CASTRO, Leonardo Humberto Silva e *et al.* **Soja: estratégia e sustentabilidade produtiva**. Viçosa, Mg: Ed. UFV, 2021. Cap. 1. p. 14-39. (Científica).

PEREIRA, Erika Thalyta Veras. **Estudos dos mecanismos de resistência de genótipos de algodon submetidos à diferentes condições bióticas e abióticas**. 2022. 198 f. Tese (Doutorado) - Curso de Química Orgânica, Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos, São Carlos, 2022.

SÁ, L.G. *et al.* Kohonen's self-organizing maps for the study of genetic dissimilarity among soybean cultivars and genotypes. **Pesquisa Agropecuária Brasileira**, [S.L.], v. 57, p. 1-8, 2022. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1678-3921.pab2022.v57.02722>.

SANT'ANNA, Isabela de Castro *et al.* RNA: aplicação em estudos classificatórios. In: CRUZ, Cosme Damião; NASCIMENTO, Moysés. **Inteligência Computacional Aplicada ao Melhoramento Genético**. Viçosa, MG: Ed. UFV, 2018. Cap. 7. p. 189-214. (Científica).

SEDIYAMA, Tuneo; SILVA, Felipe; BORÉM, Aluizio. **Soja do plantio a colheita**. Viçosa, MG: Ed. UFV, 2015. 333 p.

SEIXAS, C. D. S. S. *et al.* Tecnologias de Produção de Soja. **Embrapa Soja**. 2020. Londrina, PR. ISSN 2176-2902. n, 17.

SILVA FILHO, Jorge Luís Bandeira da. **GENÔMICA ESTRUTURAL E TRANSCRIPTÔMICA DE GENES CORREGULADORES EM *Jatropha curcas* L. SOB SALINIDADE**. 2019. 103 f. Dissertação (Mestrado) - Curso de Ciências Biológicas, Centro de Biociências, Universidade Federal de Pernambuco, Recife, 2019.

SILVA, Alisson Santos Lopes da *et al.* Contribuições do melhoramento genético da soja. In: CASTRO, Leonardo Humberto Silva e *et al.* **Soja: estratégia e sustentabilidade produtiva**. Viçosa, MG: Ed. UFV, 2021. Cap. 6. p. 228-241. (Científica).

SILVA, Felipe *et al.* **Melhoramento da Soja**. Viçosa, MG: Ed. UFV, 2017. 563 p.

SILVA, Francisco Charles dos Santos *et al.* Importância econômica e evolução do melhoramento. In: SILVA, Felipe *et al.* **Melhoramento da Soja**. Viçosa, MG: Ed. UFV, 2017. Cap. 1. p. 9-29.

SKOWRONSKI, L. *et al.* Supervised learning algorithms in the classification of plant populations with different degrees of kinship. **Brazilian Journal Of Botany**, [S.L.], v. 44, n. 2, p. 371-379, 4 fev. 2021. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s40415-021-00703-1>.

SOLAR, Flávia Lopes. **Caracterização de genótipos mutantes de arroz para tolerância à salinidade e as baixas temperaturas**. 2022. 57 f. Dissertação (Mestrado) - Curso de Agronomia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Pelotas, 2022.

TODESCHINI, M. H. *et al.* Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. **Euphytica**, [S.L.], v. 215, n. 7, 19 jun. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10681-019-2439-9>.

TORKAMANEH, D; LEMAY, M-A; BELZILE, François. The pan-genome of the cultivated soybean (PanSoy) reveals an extraordinarily conserved gene content. **Plant Biotechnology Journal**, [S.L.], v. 19, n. 9, p. 1852-1862, 15 jun. 2021. Wiley. <http://dx.doi.org/10.1111/pbi.13600>.

VENN, J. I. On the diagrammatic and mechanical representation of propositions and reasonings. **The London, Edinburgh, And Dublin Philosophical Magazine And Journal Of Science**, [S.L.], v. 10, n. 59, p. 1-18, jul. 1880. Informa UK Limited. <http://dx.doi.org/10.1080/14786448008626877>

VIANNA, M. S. *et al.* Selection of lineages, genetic parameters, and correlations between soybean characters. **Bioscience Journal**, [S.L.], p. 1300-1314, 2019. EDUFU - Editora da Universidade Federal de Uberlandia. <http://dx.doi.org/10.14393/bj-v35n5a2019-42656>.

CAPÍTULO 1: Potencial de progênies da soja oriundas do sul do Brasil

RESUMO

O melhoramento genético da soja busca genótipos com superioridade agronômica e enfrenta diversos problemas de natureza genética e econômica visto que o Brasil é um país com território muito extenso e requer cultivares adaptadas a cada ambiente. O objetivo do trabalho é caracterizar e estimar o potencial produtivo e de qualidade de genótipos da soja e compreender as associações existentes entre algumas características agronômicas e outras representativas da qualidade industrial e apontar possíveis genótipos que atendam às necessidades do programa de melhoramento. Foram utilizadas quatro coleções totalizando 301 genótipos do programa de melhoramento da UNIJUÍ do estado do Rio Grande do Sul, e analisadas dez características quantitativas, sendo elas massa de cem sementes (MCS), proteína (PTN), óleo (OLE), fibra (FIB), cinza (CIN), ácido palmítico (APA), ácido esteárico (AES), ácido oleico (AOL), ácido linoleico (ALI), ácido linolênico (ALN). Foi realizada uma análise descritiva das características e aplicado o teste de Tukey e a estatística de Lilliefors, depois os dados foram submetidos à uma análise de correlação de Pearson. A rede de correlações foi gerada através da matriz de distância euclidiana, e por fim os melhores genótipos foram analisados de acordo com Diagramas de Venn. As análises mostraram que a MCS esteve inferior, média de 15.66, requerendo esforços do melhorista afim de elevar esta característica. Além disso, dentre os constituintes da semente, apenas o teor de proteína se apresentou baixo, com média no patamar de 33,40% associado a variabilidade de 2,02. PTN e OLE apresentaram curva de distribuição próximo da Normal, indicando possível controle poligênico de natureza aditiva. A correlação mais expressiva foi entre PTN e OLE, valor de -0,7, e já é muito conhecida. MCS se correlacionou positivamente com PTN e negativamente com FIB, valores de 0,18 e 0,31 respectivamente. As redes de correlações individualizadas mostraram que a coleção 4 apresentou maior número de interações com alta intensidade, sejam elas positivas ou negativas e a coleção três foi a que apresentou o menor. Os diagramas apresentados foram concordantes em apontar a dificuldade de evidenciar genótipos que tenham bom desempenho considerando simultaneamente as características MCS, PTN e OLE. Sendo assim conclui-se que os genótipos avaliados apresentaram bom potencial em termos de teor de proteína, atingindo valor médios de 33,40%, porém a coleção 3 apresentou 34.29% de proteína, com baixa variabilidade exigindo esforços adicionais para incluir formas alélicas mais favoráveis recorrendo a genótipos de outras coleções, e foram identificados genótipos de bom desempenho que reúnam características de interesse tais como massa de cem sementes (MCS) e teor de proteína (PTN) ou massa de cem sementes (MCS) e teor de óleo (OLE).

Palavras-Chave: Características comerciais, coleções, variabilidade, *Glycine max*.

ABSTRACT

The genetic improvement of soybean seeks genotypes with agronomic superiority and faces several problems of a genetic and economic nature, since Brazil is a country with a very extensive territory and requires cultivars adapted to each environment. The objective of this work is to characterize and estimate the productive and quality potential of soybean genotypes and to understand the associations between some agronomic characteristics and others that are representative of industrial quality and to point out possible genotypes that meet the needs of the breeding program. Four collections totaling 301 genotypes from the UNIJUÍ breeding program in the state of Rio Grande do Sul were used, and ten quantitative traits were analyzed, namely mass of one hundred seeds (MCS), protein (PTN), oil (OLE), Fiber (FIB), Ash (CIN), Palmitic Acid (APA), Stearic Acid (AES), Oleic Acid (AOL), Linoleic Acid (ALI), Linolenic Acid (ALN). A descriptive analysis of the characteristics was carried out and the Tukey test and Lilliefors statistics were applied, after which the data were submitted to a Pearson correlation analysis. The correlation network was generated through the Euclidean distance matrix, and finally the best genotypes were analyzed according to Venn Diagrams. The analyzes showed that the MCS was lower, an average of 15.66, requiring efforts from the breeder in order to increase this characteristic. In addition, among the constituents of the seed, only the protein content was low, with an average of 33.40% associated with a variability of 2.02. PTN and OLE showed a distribution curve close to Normal, indicating possible polygenic control of an additive nature. The most expressive correlation was between PTN and OLE, value of -0.7, and is already well known. MCS correlated positively with PTN and negatively with FIB, values of 0.18 and 0.31 respectively. The individualized correlation networks showed that collection 4 had the highest number of interactions with high intensity, whether positive or negative, and collection three had the lowest number. The diagrams presented agreed in pointing out the difficulty of showing genotypes that perform well considering the MCS, PTN and OLE characteristics simultaneously. Therefore, it is concluded that the evaluated genotypes showed good potential in terms of protein content, reaching average values of 33.40%, however collection 3 presented 34.29% of protein, with low variability requiring additional efforts to include more favorable allelic forms resorting to genotypes from other collections, and good performance genotypes were identified that gather characteristics of interest such as mass of one hundred seeds (MCS) and protein content (PTN) or mass of one hundred seeds (MCS) and oil content (OLE).

Keywords: Commercial characteristics, collections, variability, *Glycine max*.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) começou a ser cultivada há mais de cinco mil anos e tem seu centro de origem na China. Sendo uma espécie leguminosa, a soja tem alta qualidade proteica, o que torna seus produtos uma ótima fonte de proteínas a base de plantas (QIN; WANG; LUO, 2022). A soja apresenta em média 40% de proteínas, 35% de carboidratos, 20% de óleo e 5% de cinzas (BERHOW *et al.*, 2020). No Brasil, sua utilidade na indústria se refere à produção de grãos, farelo e óleo (CONAB, 2022). No campo da exportação, a soja ainda é considerada o principal produto agrícola, fato que assume importante papel no produto interno bruto do país (SILVA *et al.*, 2021).

Como quinto maior país do mundo em extensão territorial, o Brasil apresenta alta amplitude e diversidade climática, influenciando em uma produção da soja pelo território nacional. Segundo o quarto levantamento 2021/22 da Companhia Nacional de Abastecimento (CONAB, 2022), a estimativa da área plantada da soja cresceu cerca de 3,8%, o que assegura o primeiro lugar como produtor mundial desta oleaginosa. Tal produção é influenciada pelos fatores climáticos de cada região, o que leva a uma necessidade de diferentes tipos de genótipos, e que este sejam adaptadas as necessidades in loci (BATTISTI; SENTELHAS, 2021)

Os programas de melhoramento avançam a cada ano no país, selecionando principalmente genótipos de acordo com suas características comerciais e agronômicas (CARVALHO *et al.*, 2021; CARVALHO *et al.*, 2022). No caso das plantas autógamas perfeitas, como a soja, o programa de melhoramento é conduzido por meio de hibridações artificiais e posteriores avanços das populações segregantes, por autofecundações, visando a obtenção de melhores genótipos (HERRERA *et al.*, 2020; LEITE *et al.*, 2020). Além disso, as características fisiológicas da soja apresentam alta correlação com a produtividade da planta, tornando-a alvo de programas de melhoramento que utilizam seleção indireta (TODESCHINI *et al.*, 2019).

Outros aspectos importantes para os programas de melhoramento da soja dizem respeito às características apresentadas pela cultura e o estudo de diversidade genética. Segundo Vianna *et al.* (2019), os traços selecionados pelos programas de reprodução e pesquisa da soja são o alto rendimento de grãos, ampla estabilidade e adaptabilidade, pais que sejam divergentes e apresentem os traços favoráveis para a o objetivo do programa e superioridade agronômica. Por isso conhecer e possuir uma ampla base genética pode favorecer os programas de melhoramento, uma vez que a identificação de genitores geneticamente diferentes proporciona

maior efeito heterótico no híbrido, além de maior segregação de recombinações (CRUZ; FERREIRA; PESSONI, 2011).

Dados de literatura indicam que a soja apresenta em torno de 40 % de proteína e 20 % de óleo, em base seca. A variação do teor de proteína e óleo é determinada principalmente por fatores genéticos, mas com forte influência ambiental, principalmente àqueles ocorridos no período de enchimento de grãos. Os programas de melhoramento têm enfatizado o aumento do teor de proteínas de modo que a indústria consiga obter o farelo de soja nos patamares desejados (teores mínimos de 46 % e 48 %). Muitas vezes os programas de melhoramento dão maior ênfase ao aumento da produtividade (kg/ha de grãos) e da resistência às doenças, do que à composição química dos grãos e isto pode comprometer a obtenção de linhagens melhoradas agronomicamente, mas com deficiências para uso na indústria com destinação aos usos diversos, tanto do óleo quanto do farelo.

A obtenção de genótipos de soja de alta produtividade e alta qualidade industrial é desejada, porém, a prática encontra uma série de problemas. Um deles, de natureza genética, é associada à existência de relação desfavorável entre a concentração de proteína e a produtividade de grãos, ou entre as concentrações de proteína e óleo, requerendo mais tempo e esforço nos programas melhoramento genético. O outro diz respeito ao aspecto econômico visto que o produtor recebe pela quantidade produzida, e não pelos teores de proteína e óleo dos grãos que são produzidos.

Programas de melhoramento que iniciam seus esforços na busca de materiais genéticos de maior qualidade devem considerar que os aspectos genéticos são determinantes para a variabilidade genética que estará à sua disposição e, portanto, uma avaliação detalhada sobre o material disponível será útil na orientação de melhores estratégias. Por outro lado, também deve ser considerado a origem geográfica da coleção de trabalho pois a influência ambiental na variação do teor de proteína e óleo é fundamental, destacando os efeitos da temperatura do ar no período de enchimento de grãos, que tem papel importante nesta variação, e outros aspectos como variação na disponibilidade de nitrogênio à planta, manejo dentre outros.

Diante do exposto a avaliação do potencial produtivo da soja e sua potencialidade, quanto às características de qualidade industrial, são de grande interesse do melhorista. Segundo Herrera *et al.* (2020), tal potencial produtivo e de qualidade da cultura da soja é expresso levando em consideração os componentes ambientes e genéticos, além da interação resultante entre eles. Além disso, para o cultivo sustentável da cultura é necessário ainda genótipos com

qualidade de grãos e níveis de produtividade superiores (FINOTO *et al.*, 2021). Sendo assim, o presente trabalho tem o objetivo de caracterizar e estimar o potencial produtivo e de qualidade de genótipos da soja (*Glycine max* (L.) Merrill) e compreender as associações existentes entre algumas características agronômicas e outras representativas da qualidade industrial e apontar possíveis genótipos que atendam às necessidades do programa de melhoramento.

2. MATERIAIS E MÉTODOS

2.1. Material genético utilizado

Os genótipos utilizados são provenientes do programa de Melhoramento Genético Linha Grãos e Coberturas da Universidade Regional do Noroeste do Rio Grande do Sul (UNIJUÍ) Experimento Nutraceuticos, realizado no município de Campos Borges, estado do Rio Grande do Sul, com clima definido como clima subtropical úmido. O delineamento utilizado foi de blocos aumentados, em que quatro coleções foram testadas. Na tabela 1 encontra-se as informações das 4 coleções.

Tabela 1. Identificação das progênies e do número de indivíduos correspondentes a cada uma. Sendo F₂, população base; CV, cultivares testemunhas, CB e F₄ e F₇ são segregantes, sendo F₄ e F₇ linhas avançadas.

Coleção	Progênies (Classes)	Número de indivíduos
1	CB	154
2	CV	22
3	Segregantes de F ₄ e F ₇	87
4	F ₂	38
Total		301

Os dados fenotípicos obtidos de dez características quantitativas foram coletados e transformados em uma tabela de médias.

2.2. Caracteres Agronômicos avaliados

As dez características analisadas foram massa de 100 sementes (g) (MCS) que representa um dos componentes primários da produção de grãos e outras características da qualidade industrial que são: Proteína (%) (PTN), Óleo (%) (OLE), Fibra (%) (FIB), Cinza (%) (CIN), Ácido Palmítico (%) (APA), Ácido Esteárico (%) (AES), Ácido Oleico (%) (AOL), Ácido Linoleico (%) (ALI), Ácido Linolênico (%) (ALN).

Inicialmente as amostras de 100 gramas de sementes da soja foram limpas e aferidas. Posteriormente o valor de umidade da foi corrigido para o valor de 13% de umidade. 13% é o teor de umidade indicado para o armazenamento de grãos e evitar danos mecânicos e latentes,

além de pragas, uma vez que sementes colhidas acima de 13% tendem a sofrer danos mecânicos e latentes e abaixo de 13% tendem a sofrer danos imediatos (LORINI *et al.*, 2020).

Após a correção da umidade, foram utilizadas amostras de 55 gramas que avaliaram as seguintes características nos espectrofotômetros: PTN, OLE, FIB, CIN, APA, AES, AOL, ALI, ALN.

2.3. Estatísticas Descritivas

A análise descritiva das características contendo as informações sobre as médias, desvio padrão, variância, coeficiente de variação, valores mínimos e máximos, referentes aos genótipos da soja foram realizadas utilizando o software GENES (CRUZ, 2016).

Para cada variável as estatísticas são calculadas da seguinte forma:

- N: Número total
- Média (\bar{X}): $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- Valor Mínimo
- Valor Máximo
- Coeficiente de variação (CV): $CV = \frac{100\sigma}{\bar{X}}$
- Variância (σ^2) dos dados, dado por: $\sigma^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{1}{n} (\sum_{i=1}^n X_i)^2 \right]$
- Desvio Padrão (σ , dado por: $\sigma = \sqrt{\sigma^2}$)

2.4. Associação entre características

Os dados foram submetidos à uma análise de correlação de Pearson, descrita na equação abaixo:

$$r = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

Em que,

$$Cov(X, Y) = \frac{1}{n-1} \left[\sum_{i=1}^n X_i Y_i - \frac{1}{n} \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right) \right]$$

O coeficiente de correlação é adimensional e seu valor absoluto encontra-se entre os valores de 1 e -1, ou seja, $-1 \leq r \leq 1$. Um índice de correlação nulo, igual a 0, não indica falta de correlação entre as variáveis, somente que não existe uma correlação linear entre as mesmas.

A significância da correlação é avaliada pelo teste t descrita na equação abaixo, em que a hipótese de que o coeficiente de correlação seja igual a 0:

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

em que está associado a $n-2$ graus de liberdade e ao nível de significância α . Quando se tem interesse no teste, a 5% ou 1% de probabilidade.

2.5. Rede de correlações

A matriz de correlação foi analisada através da rede de correlação, que utilizou a matriz de distância euclidiana, em que Y_{ij} representa a observação no i -ésimo genótipo (clone, genótipo, cultivar, linhagem etc.) para a j -ésima característica e , definindo a distância entre o par de genótipos i e i' por meio da expressão:

$$d_{ii'} = \sqrt{\sum_j (Y_{ij} - Y_{i'j})^2}$$

A matriz de distância euclidiana foi gerada através da correlação de Pearson, que mede a relação linear entre duas variáveis sendo definida por:

$$r_{y_1y_2} = \frac{s_{12}}{\sqrt{s_1^2 \times s_2^2}}$$

Em que:

$r_{y_1y_2}$ = correlação linear de Pearson entre Y1 e Y2;

s_{12} = covariância amostral entre as variáveis dependentes Y1 e Y2;

s_1^2 = variância amostral da variável dependente Y1;

s_2^2 = variância amostral da variável dependente Y2.

A espessura e a intensidade da cor das bordas foram controladas por um valor de corte igual a 0.3, que significa que apenas $|r_{ij}| \geq 3$ terão suas linhas destacadas. Além disso, as correlações positivas serão destacadas pela cor verde e as negativas pela cor vermelha. Ademais, proporcionalmente, quanto maior a espessura da borda, maior é a correlação entre as variáveis (EPSKAMP *et al.*, 2012). As características que apresentam correlação positiva tendem a aumentar ou diminuir quando uma aumenta ou diminui respectivamente. A correlação negativa tende a aumentar ou diminuir quando uma diminui ou aumenta respectivamente.

3. RESULTADOS E DISCUSSÃO

Potencial dos acessos de soja para característica de produtividade e qualidade industrial

Na Tabela 2 encontram-se as estimativas referentes ao conjunto total de genótipos disponíveis, independentemente de sua origem, designada neste trabalho de coleção de trabalho.

Tabela 2. Análise descritiva de dez características quantitativas mensuradas em 301 genótipos da soja. Sendo, CV, coeficiente de variação; DP: desvio padrão; MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

Característica	Média	Mínimo	Máximo	CV	Variância	DP
MCS	15.64	12.33	21.48	11.13	3.03	1.74
PTN	33.40	29.2	36.41	4.26	2.02	1.42
OLE	19.07	16.59	21.37	4.28	0.66	0.81
FIB	5.88	5.15	6.54	3.52	0.04	0.21
CIN	5.16	4.99	5.51	1.47	0.01	0.08
APA	10.07	7.10	14.06	11.33	1.30	1.14
AES	4.15	3.65	4.93	3.90	0.03	0.16
AOL	22.88	17.11	32.39	10.86	6.18	2.48
ALI	59.24	46.9	64.08	3.68	4.74	2.18
ALN	2.48	0.00	11.77	66.04	2.68	1.64

Destaca-se que o presente estudo envolve uma característica diretamente relacionada à produção de grãos que é a massa de 100 sementes e as demais referem-se à qualidade dos grãos. Assim, verifica-se que a MCS(g) (Tabela2) variou de 12.33g a 21.48g com média de 15.66g tal fato revela que esforços do melhorista poderá elevar este valor médio em níveis mais elevados. Alguns estudos têm demonstrado que o valor médio se aproxima dos nossos resultados (BRANDT *et al.*, 2006; BARROS *et al.*, 2021; FARIA NETO *et al.*, 2022)

Programas de melhoramento com ênfase na elevação da concentração de proteína ou de óleo no grão também devem priorizar o aumento da produtividade, ou seja, quantos quilos de proteína ou litros de óleo são produzidos por hectare. Uma produtividade de 1000 kg ha⁻¹ de soja com 30 % de proteína resultará em 300 kg de proteína ha⁻¹. Este mesmo patamar de proteína, com genótipos que tenham teor em torno de 40 % será alcançado com produtividade de 750 kg ha⁻¹. Assim, é necessária uma avaliação conjunta destes atributos.

Pela análise descritiva é possível verificar o coeficiente de variação abaixo de 30% para todas as características, com exceção apenas do ALN que apresentou uma alta heterogeneidade,

visto que o coeficiente de variação apresentado foi de 66.04%. Além disso, dentre os constituintes da semente, apenas o PTN se apresentou baixo, com média no patamar de 33,40% associado a variabilidade de 2,02.

O teor de proteína (%) variou de 29.2% a 36.41% com média de 33.41%. Kurasch *et al.*, (2017) encontraram o teor de proteína variando entre 39.15% e 39.70%, com média de 39.59% em 3 ensaios experimentais diferentes e uma série entre os ensaios, Zhou *et al.* (2019) verificaram o teor de proteína médio com valor de 41.60% em uma população selvagem, Conte *et al.* (2020) observaram estimativa média geral de 39.1%. Assim, no presente trabalho verifica-se que valores de PTN estão abaixo do esperado para a qualidade tecnológica da soja que é referenciado como sendo de 38% (LORINI *et al.*, 2020), com o agravante de que o valor máximo manifestado por um dos genótipos foi de 36,41. Assim, o programa de melhoramento deverá investir esforços no estabelecimento de combinações entre genitores com formas alélicas favoráveis para que, nas populações segregantes, possa encontrar transgressivos que venha melhor atender as exigências industriais.

Em relação a característica teor de óleo, a variação constituiu-se entre 16.59 e 21.37 com estimativa de média geral de 19,07%. Valor médio dentro do esperado que é próximo de 19% (LORINI *et al.* 2020). Resultado este corroborado por estudos como de Kurasch *et al.* (2017) que encontraram o teor de óleo variando entre 18.55% e 18.73% e média de 18.075% em 3 ensaios experimentais diferentes e uma série entre os ensaios, Zhou *et al.* (2019) encontraram o teor de óleo de com média de 15.26%, Conte *et al.*, (2020) encontraram 19.1% de média.

A fibra proteica de soja, ou fibra de soja, é uma fibra têxtil extraída do resíduo da semente após a extração do óleo. Essas fibras têm efeitos benéficos à saúde, pois são responsáveis pela diminuição do colesterol e dos níveis de glicose no sangue, e influenciam na regulação do intestino. No presente trabalho, a FIB (%) variou de 5.15% a 6.54% com média de 5.8798%. Estes resultados estão próximos aos valores encontrados por Fachi *et al.* (2020) sendo 5.57% e dos valores disponibilizados por Valadares Filho *et al.* (2023) que é 5.30% para o teor médio de fibra para grãos da soja.

Outra característica em estudo foi o teor de cinzas (%), que variou de 4.99% a 5.51% com média de 5.1562%. Ciabotti *et al.* (2019) verificaram o teor de cinzas entre 5.30% e 5.59%. LORINI *et al.* 2020, argumenta que o desvio da estimativa em 5%. Isso justifica o nosso resultado encontrado.

Por fim, o teor de ácidos graxos (%) também se manteve na faixa esperada. Os ácidos graxos palmítico, esteárico, oleico, linoleico, linolênico, tiveram concentrações médias de 10.07%, 4.15%, 22.89%, 59.25% e 2.48% respectivamente. Ciabotti *et al.* (2019) encontraram o perfil de ácidos graxos de APA, AES, AOL, ALI, ALN, em médias para 11 cultivares e linhagens com estimativas entre 10,37% e 12.30%; 3.58% e 4.13%; 17.89% e 31.87%; 47.36% e 58.31% e ALM variando entre 6.17% e 7.80%, respectivamente.

Abdelghany *et al.* (2020) sugerem que a composição dos ácidos graxos das sementes da soja pode estar associada ao processo de domesticação da soja. Além disso, Tamagno *et al.* (2020) explicam que a variação dos diferentes ácidos graxos em sementes da soja está relacionada principalmente ao período de depósito efetivo mais longo durante o período de enchimento das sementes. Neste período a planta apresenta maior suprimento de nitrogênio, prorrogando a senescência foliar e mantendo o suprimento de carbono para as regiões sumidouros reprodutivos. As diferenças das concentrações dos ácidos graxos podem ser explicadas ainda pela existência de diferentes vias bioquímicas para atingirem o perfil de ácido graxos ou ainda pela influência individual que cada via bioquímica sofre pelo ambiente (TAMAGNO *et al.* 2020).

O perfil de ácidos graxos da semente da soja apresentou discrepância quanto a suas estimativas de coeficiente de variação. Dentre esses caracteres, a maior estimativa de coeficiente de variação foi para o ácido linoleico, que corresponde a 66,04%. Esse resultado é consistente com os relatados por trabalhos anteriores como no caso de Abdelghany *et al.* (2020) com valor de 14.6 % e Tamagno *et al.* (2020) com 6.562%.

Outra informação importante além das estatísticas sumarizadas na Tabela 3 é a distribuição dos valores de cada característica (Figura 1). O padrão de distribuição característico da normalidade e a simetria dos dados é informativo, uma vez que é próprio de características poligênicas, o que corresponde a ação dos genes controladores prioritariamente de natureza aditiva. Por outro lado, o padrão assimétrico pode indicar a ação de números menores de genes e, principalmente, forte influência de efeitos atribuídos às dominâncias no controle das características.

Como é possível observar na Tabela 3 e Figura 1, as características MCS e teor de fibras, Figuras 1a e 1d, respectivamente, apresentaram curvas de distribuição que não se aproximam da distribuição normal, por apresentarem p-valor que reflete a significância da estatística D do teste de Lilliefors. O padrão acentuado de assimetria para característica MCS é indicativo de

algum controle genético em que efeitos gênicos não aditivo possam estar atuando. A característica Fibras foi a que mais se distanciou do padrão de normalidade demonstrando a rejeição do teste de normalidade, por meio da estatística D de Lilliefors, certa assimetria e curva com padrão de curtose elevado.

Tabela 3. Resultados do teste de Lilliefors (estatística D) para quatro características de interesse agrônomo, sendo: MCS: massa de cem sementes; PTN: proteína; OLE: óleo; e FIB: fibra.

Característica	Média	Variância	Simetria ¹	Curtose ²	D ³
MCS	15.64	3.03	0.75*	3.35ns	0.1182*
PTN	33.40	2.02	-0.33*	2.82ns	0.0481ns
OLE	19.01	0.67	-0.15ns	3.15ns	0.0251ns
FIB	5.88	0.04	-0.36*	3.85*	0.0603 *

(1 e 2) *, ns : significativo e não significativo, respectivamente, pelo teste t a 5% de probabilidade

(3) *, ns : significativo e não significativo, respectivamente, pelo teste de Lilliefors a 5% de probabilidade. Ho: é razoável estudar os dados por meio da distribuição normal.

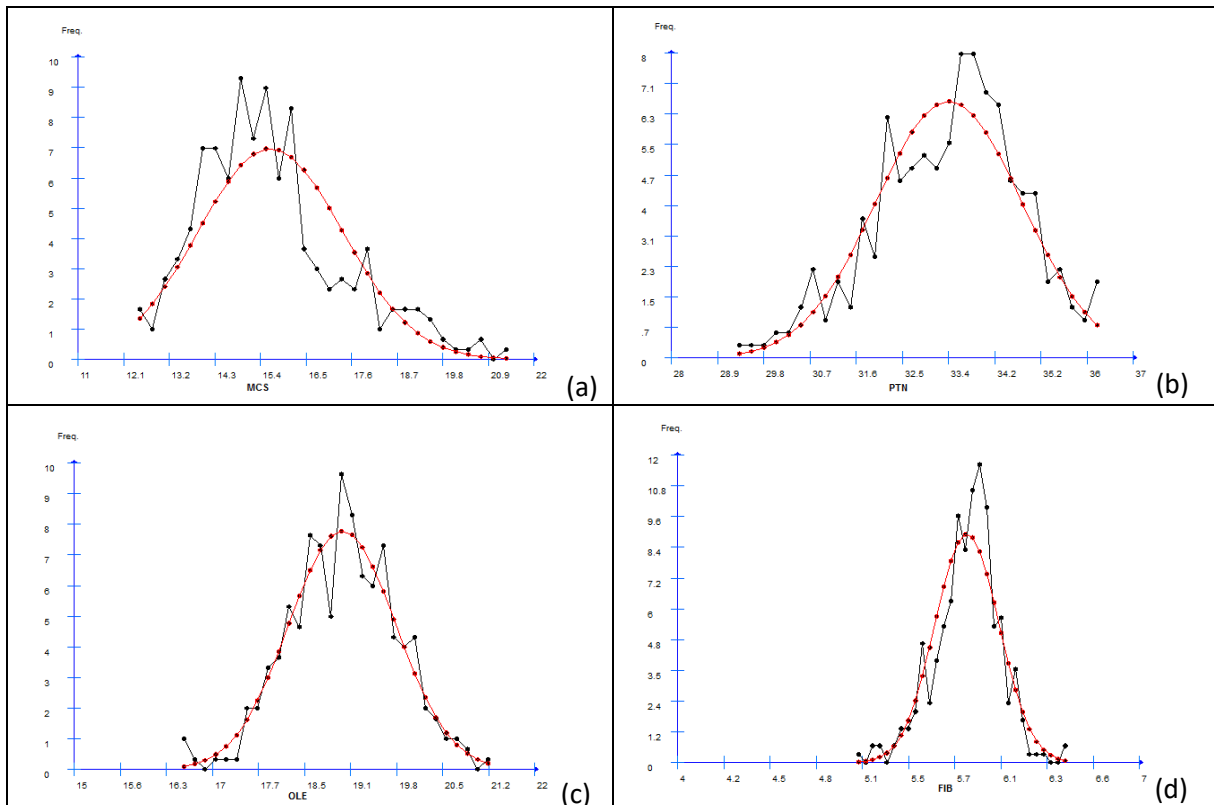


Figura 1. Representação do Teste de Lilliefors, sendo: (a) massa de cem sementes (b) proteína; (c) óleo; e (d) fibra. Linha preta representa a curva da frequência observada; e a linha vermelha representa a curva da frequência esperada; Freq, a frequência; MCS, massa de cem sementes; PTN, teor de proteína; OLE, teor de óleo; FIB, teor de fibras.

Estatisticamente quando o valor da curtose é maior que 3, como manifestado na característica Fibras, a distribuição dos dados torna-se mais alta, com padrão afunilado e com valores concentrados próximos ao valor médio. Diz-se que esta função probabilidade é leptocúrtica, ou que a distribuição tem caudas pesadas que o padrão de normalidade. No contexto do melhoramento genético, tem-se uma variabilidade restrita e maior dificuldade de se encontrar, nas populações segregantes do avanço de combinações híbridas dos genótipos testados, algum tipo de transgressivo que possa superar os limites estabelecidos pelos genitores.

As características PTN e OLE, figuras 1b e 1c respectivamente, apresentaram distribuição com padrão típico de normalidade. Tal fato indica a existência de controle poligênico predominantemente aditivo. Um pequeno grau de assimetria foi detectado para a característica proteína indicado a presença de algum gene de efeito maior, com manifestação de certo grau de dominância.

As informações apresentadas podem ser úteis em exercícios de predições refletindo a dificuldade da atividade em programas de melhoramento. Assim, como ilustração, considerando a existência do padrão de normalidade e os valores de média e variância iguais a 33,40 e 2,02, respectivamente, quantifica-se que a probabilidade de se obter valores na população atual dentro de um limite mais ambicioso, estabelecido entre 36 e 40%, será de 3,37%. Uma maior margem de êxito deverá ser perseguida por meio da implementação de uma população base enriquecida com alelos favoráveis de forma que estes limites sejam mais facilmente obtidos por seleção.

Potencial dos acessos, por coleção individualizada, de soja para característica de produtividade e qualidade industrial

Nesta abordagem foi considerado o interesse particular de utilizar genótipos da coleção 3, representada por linhagens segregantes em F4 e F7 conduzidos pelo programa de melhoramento Universidade Regional do Noroeste do Rio Grande do Sul (UNIJUÍ). Torna-se fundamental uma boa caracterização desta coleção e uma comparação do desempenho relativo em relação a outras 3 coleções com genótipos utilizados nesta experimentação.

A linhagens segregantes em F4 e F7, ou seja, linhagens mais avançadas no programa de melhoramento, apresentaram uma média maior do que a média geral para MCS, tendo a média de 16.58g, variando de 13.61g a 20.11g. Pelo teste Tukey, foi possível diferenciar estatisticamente a 5% de probabilidade a coleção três das demais coleções, indicando que a

característica em questão está em bom patamar, mas é superada, estatisticamente, pela coleção 2 (CV).

A soja é um alimento calórico-proteico importante para diminuir a desnutrição. É uma alternativa proteica de boa qualidade para vegetarianos, possui uma fração lipídica rica em ácidos graxos poli-insaturados e carboidratos com atividade prebiótica. Há interesse em que as cultivares de soja apresentem um teor médio entre 36% e 40% de proteína. Entretanto, há relatos de atingir conteúdos superiores a 45%, no caso de cultivares especiais utilizadas em cruzamentos genéticos como fonte para alto teor de proteínas. PTN apresentou média geral de 33.40% (Tabela 2), mas superior na coleção de interesse (Coleção 3) em relação à média geral das quatro coleções, visto que a média foi de 34.29%, variando entre 31.05% e 36.41%. Este valor se diferenciou das demais coleções pelo teste Tukey. Kurasch *et al.*, (2017) encontraram o teor de proteína máximo de 39.15% em 3 ensaios diferentes e uma série entre os ensaios, Zhou *et al.* (2019) encontraram o teor de proteína médio com valor de 41.60% em uma população selvagem, Conte *et al.* (2020) encontraram 39.1% de média. O aumento deste teor de proteína para patamar de 38% é desejável, conforme salienta (LORINI *et al.*, 2020).

O óleo é utilizado como matéria-prima pela indústria para produção de óleo refinado, gorduras hidrogenadas, margarinas, maionese dentre outros produtos (MANDARINO; ROESSING, 2001). Também tem sido utilizado em produtos industriais como tintas, lubrificantes, solventes, plásticos e resinas (ERHAN, 2005). Mais recentemente, tem sido a principal matéria-prima para produção de biodiesel.

A coleção de interesse apresentou desempenho próximo da média geral das coleções quanto ao teor de óleo (média geral de 19,07%, conforme Tabela 2) apresentando média de 19%, variando entre 17.16% e 20.12%. Valor mais elevados foi encontrado na coleção 2 atingindo teor de óleo de 19,44.

Das quatro coleções avaliadas, a coleção 3 apresentou o teor de fibra (%) mais baixo. Destaca que ela apresentou valor próximo ao encontrado por Fachi *et al.* (2020) sendo 5.57% e dos valores disponibilizados por Valadares Filhos *et al.* (2023) que é 5.30% para o teor médio de fibra para grãos da soja.

CIN, APA e AES estiveram nos patamares mais elevados em comparação com as demais coleções avaliadas. Para Cinzas o valor médio encontrado foi 5,16 dentro do esperado da composição centesimal das sementes da soja, que varia em 5% (LORINI *et al.* 2020).

O teor de ácidos graxos (%) também se manteve na faixa esperada. Os ácidos graxos oleico, linoleico e linolênico, tiveram concentrações médias relativamente baixas, na coleção 3, com valores, respectivamente, de 23,01%, 59,90% e 1,89% respectivamente.

Tabela 4. Análise descritiva do potencial individual de dez características quantitativas mensuradas em genótipos da soja em quatro coleções diferentes. CB, coleção segregante que passou por Campos Borges; CV, coleção cultivar; SEG, coleção segregantes de F₄ e F₇; F₂, coleção de população base; Min, valor mínimo; Max, valor máximo; MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

Característica	CB			CV			SEG			F2		
	Média	Min	Max	Média	Min	Max	Média	Min	Max	Média	Min	Max
MCS	15.11c	12.44	19.33	17.86a	14.97	21.48	16.58b	13.61	20.11	14.32d	12.33	19.32
PTN	33.01b	29.50	36.28	32.97b	29.20	35.78	34.29a	31.05	36.41	33.22b	30.77	36.00
OLE	19.18ab	16.71	21.37	19.44a	17.87	20.96	19.00b	17.16	20.12	18.59c	16.59	19.94
FIB	5.94a	5.28	6.54	5.81b	5.48	6.16	5.80b	5.28	6.20	5.91a	5.15	6.51
CIN	5.15a	4.99	5.47	5.12b	5.00	5.29	5.16a	5.08	5.45	5.17a	4.99	5.51
APA	10.19a	7.13	14.06	10.23a	8.71	11.51	10.20a	8.43	11.87	9.17b	7.10	13.25
AES	4.15a	3.65	4.56	4.12a	3.97	4.36	4.15a	3.74	4.51	4.18a	3.80	4.93
AOL	22.17b	17.11	30.79	22.45b	18.50	26.38	23.01b	19.11	29.89	25.77a	18.16	32.39
ALI	59.30b	53.75	63.98	60.42a	56.68	64.08	59.90ab	56.22	63.98	56.85c	46.90	60.98
ALN	2.94a	0.00	10.72	1.50c	0.00	3.87	1.89bc	0.00	4.87	2.54ab	0.00	11.77

(*) Médias seguidas pelas mesmas letras, na horizontal, não diferem entre si pelo teste Tukey a 5% de probabilidade.

Na Tabela 5 são apresentadas informações sobre a variabilidade existente dentro de cada coleção. Comparativamente, a coleção 3 não manifestou valores comparativamente mais elevado para nenhuma característica mensurada. A baixa variabilidade manifestada para as características proteínas e óleo é um alerta para que se inclua contribuições de genitores. Uma vez que possam ser fonte de alelos favoráveis à estas características complementares aqueles já existentes na coleção de interesse.

Tabela 5. Análise descritiva da variabilidade manifestada em dez características quantitativas mensuradas em genótipos da soja de diferentes coleções. CB, coleção segregante que passou por Campos Borges; CV, coleção cultivar; SEG, coleção segregantes de F₄ e F₇; F₂, coleção de população base; Min, valor mínimo; Max, valor máximo; Var, valor da variância; DP, valor do desvio padrão; MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

Característica	CB		CV		SEG		F2	
	Var	DP	Var	DP	Var	DP	Var	DP
MCS	1.78	1.34	3.36	1.83	2.41	1.55	1.75	1.32
PTN	1.97	1.40	2.71	1.65	0.91	0.96	1.94	1.39
OLE	0.70	0.84	0.69	0.83	0.45	0.67	0.67	0.82
FIB	0.04	0.19	0.04	0.21	0.03	0.17	0.08	0.28
CIN	0.01	0.08	0.00	0.07	0.00	0.06	0.01	0.10
APA	1.49	1.22	0.63	0.80	0.70	0.84	1.47	1.21
AES	0.03	0.16	0.01	0.10	0.02	0.15	0.04	0.21
AOL	4.93	2.22	4.63	2.15	3.24	1.80	8.72	2.95
ALI	3.88	1.97	2.80	1.67	2.19	1.48	7.98	2.82
ALN	2.46	1.57	1.29	1.13	1.56	1.25	4.98	2.23

A quantificação do potencial agrônomo e da variabilidade de genótipos em coleções recorrentes é fundamental para orientar os programas na busca de novos genótipos para fins de complementação gênica.

Relações entre característica de produtividade e qualidade industrial de genótipos de soja

Na Tabela 6 são apresentados resultados importantes em relação às relações das características de soja estudadas. A primeira análise destaca as associações entre a característica MCS, que é um componente primário da produção de grãos, com as demais representativas da qualidade do grão. Assim, verifica-se que MCS apresentou correlação positiva com PTN com estimativa de 0.18, e negativa com o FIB, -0.31, isso indica que o processo de seleção indireta, ao selecionar a característica de MCS, acarretará seleção também de PTN, e a característica teor de FIB será reduzida. He *et al.* (2021) encontraram correlação forte e negativa de -0.99 entre FIB e MCS e negativa entre MCS e PRO de -0.24. Essas estimativas são discrepantes dos encontrados neste estudo.

Tabela 6. Estimativas de coeficientes de correlação de Pearson entre dez características quantitativas da soja. Os índices em negrito são significativos a 1% de probabilidade pelo teste t. MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

	MCS	PTN	OLE	FIB	CIN	APA	AES	AOL	ALI	ALN
MCS	1	0.18	0.05	-0.31	-0.07	-0.01	-0.10	0.03	0.10	-0.09
PTN		1	-0.70	-0.28	0.39	-0.01	0.03	0.24	-0.06	-0.09
OLE			1	-0.02	-0.14	0.13	0.15	-0.16	0.03	0.00
FIB				1	-0.39	0.08	-0.05	-0.28	0.19	-0.01
CIN					1	-0.06	-0.08	0.44	-0.51	0.19
APA						1	0.20	-0.68	-0.03	0.43
AES							1	0.12	-0.24	-0.02
AOL								1	-0.52	-0.29
ALI									1	-0.58
ALN										1

É importante as informações de associação entre as características de qualidade. Neste trabalho, foi estimada correlação alta e negativa entre os teores de proteína e óleo, com estimativa de -0.70. Vários autores corroboram nossos resultados, encontrando valores similares para esta estimativa de correlação (KURASCH *et al.* 2017; JIANG *et al.* 2018; CONTE *et al.* 2020; SOBKO *et al.* 2020. As estimativas de correlações negativas podem ser relacionadas devido a diferenças em vias biossintéticas de ácidos graxos (PATEL *et al.*, 2004).

Apesar de ser conhecido a muitas décadas a correlação negativa em questão, os mecanismos bioquímicos específicos ainda não são entendidos corretamente (KURASCH *et al.*, 2017). Em partes esta correlação negativa pode ser explicada por uma competição existente por carbono, uma vez que ambas as vias biossintéticas de óleos e de proteínas compartilham algumas etapas bioquímicas do metabolismo do carbono (SUGIMOTO *et al.* 1989). Além disso, Diers *et al.* (1992) sugerem que ambas características são controladas pelo mesmo gene ou grupo de genes, o que foi comprovado posteriormente por Lesteri *et al.* (2013) que mapearam tais genes.

O teor de proteína ainda apresentou correlação negativa de -0.28 com fibras, e positiva com Cinza e Ácido Oleico, valores de 0.39 e 0.24 respectivamente. Carvalho *et al.* (2021) encontrou correlação de 0.38 entre PTN e FIB, valor inverso ao encontrado no nosso estudo.

FIB se correlacionou negativamente com o CIN e AOL, 0.39 e 0.28 e positivamente com ALI 0.19. He *et al.* (2021) encontraram correlação positiva de 0.251 entre FIB e CIN, já Santanna *et al.* (2023) encontraram correlação uma correlação negativa e fraca entre essas características. Nosso estudo também encontrou uma correlação negativa de -0.28 entre PRO e

FIB, que significa à medida que a característica proteína é selecionada, FIB não o é. Fachi *et al.* (2020) encontraram 0.38, valor discrepante do encontrado neste estudo.

O ácido linoleico se correlacionou negativamente com o ácido linolênico no valor de -0.58. Indicando que o favorecimento de ALI influencia na diminuição da frequência de alelos que favoreçam ALN. Índice de correlação positiva também pôde ser observado entre o ácido palmítico e o ácido linolênico, com valor de 0.43. Abdelghany *et al.* (2020) encontrou uma correlação também positiva, no valor de 0.2, Zhou *et al.* (2019) encontraram 0.336.

Observou-se também uma alta correlação negativa entre o APA e AOL com estimativa de -0.68. Zhou *et al.* (2019) encontraram -0.388, Abdelghany *et al.* (2020) encontraram -0.16. Esta associação negativa pode ser explicada pela associação direta dos ácidos graxos com sua biossíntese nas vias lipídicas (WOYANN *et al.*, 2019).

Uma correlação negativa também entre o ácido linoleico e o teor de cinzas foi verificada, o ácido oleico e o ácido linolênico, com valores de -0.51, -0.52 e -0.58 respectivamente. Tais correlações negativas AOL e ALI também foram encontradas variando de -0.47 a -0.54 em três populações diferentes em estudos como de Cardinal *et al.* (2007), e ainda mais expressivo no estudo de Abdelghany *et al.* (2020), com valor de -0.85, e em Zhou *et al.* (2019) -0.830. Entre ALI e ALN Zhou *et al.* (2019) encontraram 0.297. Entre AOL e ALN Zhou *et al.* (2019) encontraram -0.750. Segundo Zhou *et al.* (2019) a forte correlação negativa encontrada entre os AOL, ALI e ALN podem ser explicadas em parte, pela localização de suas enzimas (FAD2 e FAD3) dentro do retículo endoplasmático, enzimas estas responsáveis por suas conversões.

Correlações também foram encontradas entre o ácido palmítico e o ácido esteárico e entre o ácido esteárico e o ácido oleico, com valor de 0.20 e 0.12 respectivamente. Tais valores podem ser explicados pela presença de possíveis enzimas e suas competições por substratos dentro da via biossintéticas dos ácidos graxos (ZHOU *et al.* 2019).

Além da rede de correlação geral, redes de correlações individuais (Figura2) para cada coleção foram analisadas, no âmbito de verificar se as correlações encontradas separadamente diferem ou não. Um fato interessante que a coleção 4 apresentou maior número de interações com alta intensidade, sejam elas positivas ou negativas, já a coleção três foi a que apresentou o menor número de interações com alta intensidade. Tais interações são esperadas, uma vez que a população base, coleção 4, do programa de melhoramento (figura 2d) não foi submetida a ciclos de seleção, mantendo assim grande número de correlações entre suas características. O contrário também pôde ser observado, pois a coleção 3 (figura 2c), sendo esta as linhagens

avanzadas em F4 e F7, já foram submetidas a ciclos de seleção, diminuindo assim as correlações evidentes entre as características, uma vez que com o passar dos ciclos, algumas características são selecionadas em detrimento de outras.

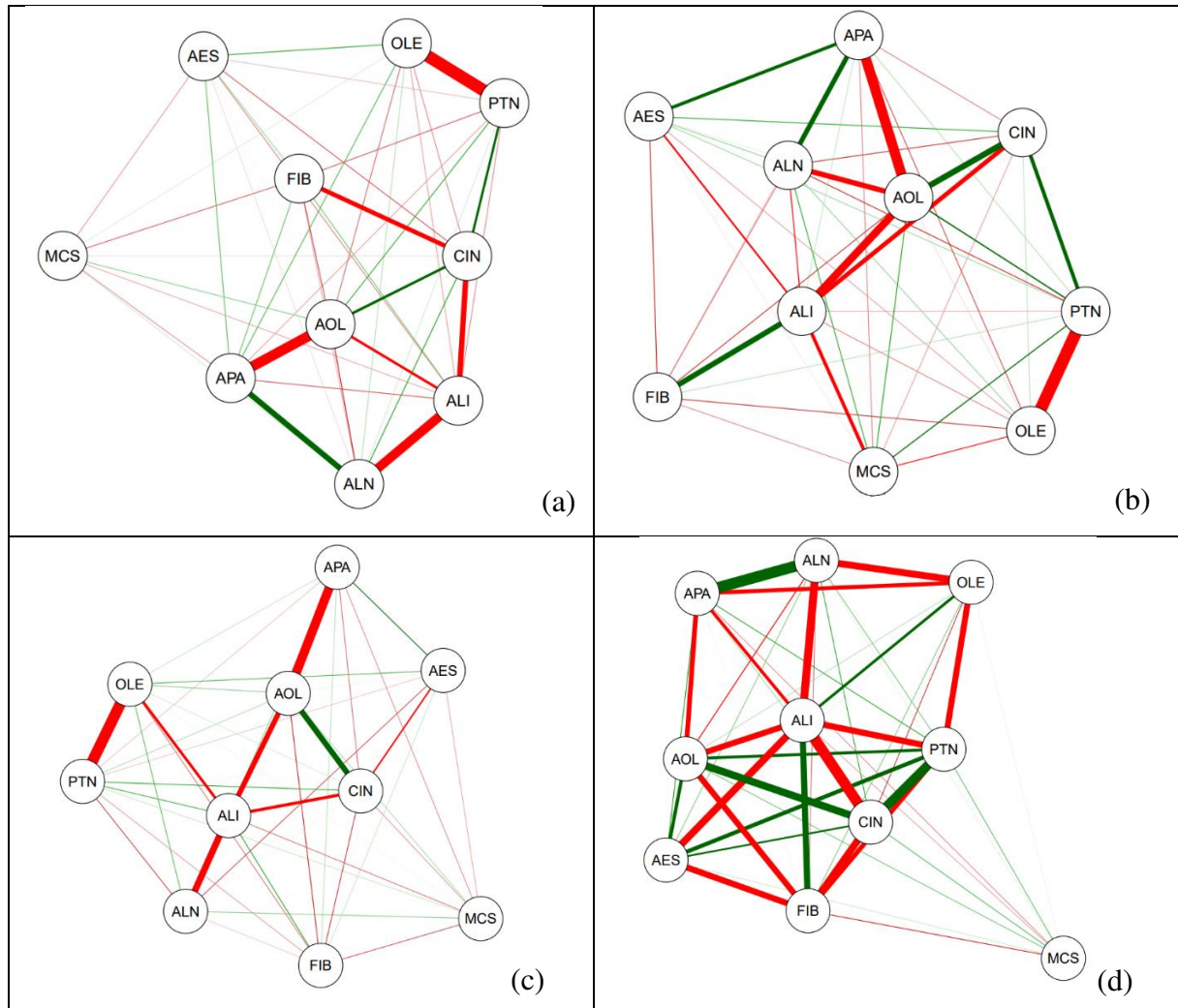


Figura 2. Redes de correlações representando cada coleção individualmente. Linhas verdes e vermelhas representam correlações positivas e negativas, respectivamente. a: a coleção CB; b, a coleção de CV; c, a coleção de segregantes; e d, a coleção de F2. MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

Na Figura 2, com base na informação das quatro coleções, a característica MCS se posiciona de forma distanciada demonstrando sua baixa associação com a demais características estudadas e que refletem a qualidade industrial da soja. Uma exceção é manifestada na coleção 2 (Figura 2 b) envolvendo a MCS e o componente ALI com associação positiva e expressiva.

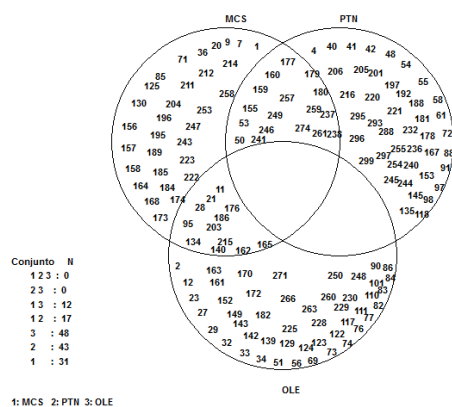
Vale ressaltar ainda, que na literatura a correlação negativa entre proteína e óleo em sementes de soja é bastante conhecida. Em todas as redes de correlação é evidenciada a associação negativa entre estas duas características. Na coleção 2 (Figura 2 b) esta associação

foi de maior magnitude, como evidenciado na espessura da ligação entre estas variáveis na reta verde usada na ligação da rede. Entretanto, como evidência na figura, esta associação negativa pode se manifestar com diferentes intensidades nas diferentes coleções.

Destaca-se os blocos de correlações envolvendo as características AOL, ALI, ALN cuja manifestação é consistentemente positiva. A este bloco é incluído a característica APA com associação negativa consistente com ALN.

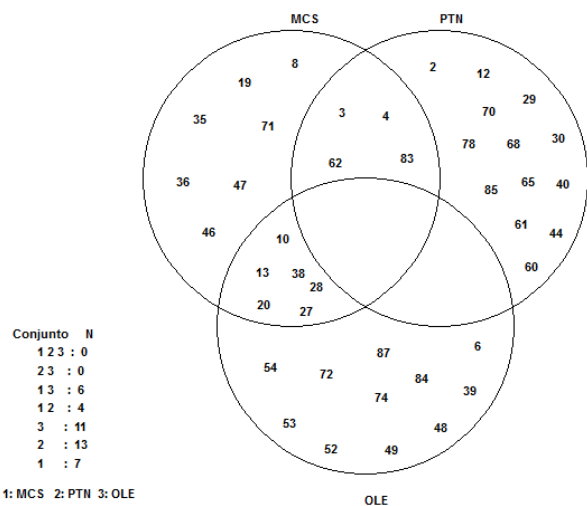
Desempenho de acessos de soja para características alvos em programa de melhoramento da soja

A análise gráfica apresentada pelo Diagrama de Venn (Figura 3) demonstra os melhores genótipos selecionados para três características de interesse, sendo elas a MCS, PTN e OLE.



(a)

Todas as coleções (20 % de indivíduos superiores)



Coleção 3 (20 % de indivíduos superiores)

(b)

Figura 3. Diagrama de Venn representando os melhores genótipos para MCS (massa de cem sementes), PTN (proteína) e OLE (Óleo). Para as quatro coleções estudadas (a); para a coleção três individualmente (b).

Os diagramas apresentados, envolvendo todas as quatro coleções ou particularizando a coleção 3 que envolve segregantes, são concordantes em apontar a dificuldade de evidenciar genótipos que tenham o melhor desempenho considerando simultaneamente as características MCS, PTN e OLE. Nestes diagramas a interseção destes três conjuntos resulta em um conjunto vazio (Figura 3). A correlação negativa entre os caracteres PTN e OLE (igual a -0,70 em todas as coleções e -0,73 na coleção 3) proporcionou conjunto vazio na interseção entre conjuntos dos 20% melhores desempenho nestas características de interesse.

Considerando o enfoque, em termos de concentração de esforços do programa de melhoramento, na coleção 3, recomenda-se investir no potencial dos genótipos 3, 4, 62 e 83 de melhor desempenho para as características MCS e PTN. Outros genótipos de com desempenho destacável também estão disponíveis nas demais coleções avaliadas podendo ser incorporados ao programa de melhoramento genético.

Se o interesse for o aumento do teor de óleo associado a maior MCS, deve-se ter atenção aos genótipos 10, 13, 20, 27, 28 e 38 da coleção 3 de bom desempenho para ambas as características.

4. CONCLUSÃO

Os genótipos avaliados apresentaram potencial destacável em termos de teor de proteína, atingindo valor médios de 33,40% com máximo de 36,41%. Das quatro coleções avaliada, a envolvendo genótipos de populações segregantes apresentou 34.29% de proteína, com baixa variabilidade exigindo esforços adicionais para incluir formas alélicas mais favoráveis recorrendo a genótipos de outras coleções.

A associação entre proteína e óleo foi ratificada como negativa. A correlação encontrada apresentou estimativa de -0.70 considerando todas as coleções.

Foram identificados genótipos de bom desempenho que reúnam características de interesse tais como massa de cem sementes (MCS) e teor de proteína (PTN) ou massa de cem sementes (MCS) e teor de óleo (OLE). Entretanto, a interseção de genótipos com melhores desempenhos simultâneos para estas três características são vazias, determinada principalmente pela associação negativa entre OLE e PTN.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ABDELGHANY, A. M. *et al.* Profiling of seed fatty acid composition in 1025 Chinese soybean accessions from diverse ecoregions. **The Crop Journal**, [S.L.], v. 8, n. 4, p. 635-644, ago. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.cj.2019.11.002>.

BARROS, L. D. *et al.* AVALIAÇÃO DAS CARACTERÍSTICAS FÍSICAS E FÍSICO-QUÍMICAS DE DIFERENTES VARIEDADES CULTIVADAS DE SOJA. **Avanços em Ciência e Tecnologia de Alimentos - Volume 3**, [S.L.], p. 289-302, 2021. Editora Científica Digital. <http://dx.doi.org/10.37885/210203054>.

BERHOW, M. A. *et al.* Quantitative NIR determination of isoflavone and saponin content of ground soybeans. **Food Chemistry**, [S.L.], v. 317, p. 126373, jul. 2020. Elsevier BV.

BRANDT, E. A. *et al.* Desempenho agrônomico de soja em função da sucessão de culturas em sistema de plantio direto. **Ciência e Agrotecnologia**, [S.L.], v. 30, n. 5, p. 869-874, out. 2006. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1413-70542006000500007>.

- CARDINAL, A. J. *et al.* Correlations between Palmitate Content and Agronomic Traits in Soybean Populations Segregating for the *fap1*, *fap nc*, and *fan* Alleles. **Crop Science**, [S.L.], v. 47, n. 5, p. 1804-1812, set. 2007. Wiley. <http://dx.doi.org/10.2135/cropsci2006.09.0577>.
- CIABOTTI, S. *et al.* Chemical composition and lipoxygenase activity of soybean (*Glycine max* L. Merrill.) genotypes, specific for human consumption, with different tegument colours. **Brazilian Journal Of Food Technology**, [S.L.], v. 22, 2019. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1981-6723.00318>.
- CONAB. CAMPANHA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira: Grãos. Brasília, DF. V. 9 – safra 21/22 – N. 4. Quarto levantamento, janeiro. 2022.
- CONTE, M. V. *et al.* Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. **Plos One**, [S.L.], v. 15, n. 5, 22 maio 2020. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0233290>.
- CRUZ, C. D. Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum Agronomy**, v. 38, n. 4, p. 547–552, 2016.
- CRUZ, Cosme Damião; FERREIRA, Fábio Medeiros; PESSONI, Luiz Alberto. **Biometria Aplicada ao Estudo de Diversidade Genética**. Visconde do Rio Branco, Mg: Suprema, 2011. 620 p.
- DIERS, Brian W. *et al.* RFLP analysis of soybean seed protein and oil content. **Theoretical and Applied Genetics**, v. 83, n. 5, p. 608-612, 1992.
- ERHAN, S.V. Industrial uses of vegetable oils. Champaign: AOCS Press, 2005.
- EPSKAMP, S. *et al.* qgraph : Network Visualizations of Relationships in Psychometric Data. **Journal of Statistical Software**, v. 48, n. 4, 2012.
- FACHI, S.M. *et al.* Research Article Multivariate selection of nutritional aspects of soybean in an F5 segregating family. **Genetics And Molecular Research**, [S.L.], v. 19, n. 2, 2020. Genetics and Molecular Research. <http://dx.doi.org/10.4238/gmr18423>.
- FARIA NETO, A. L. de *et al.* EFEITO DO RETARDO NA COLHEITA NA QUALIDADE DE GRÃOS DE CULTIVARES DE SOJA. **Nativa**, [S.L.], v. 10, n. 1, p. 47-53, 15 mar. 2022. Nativa. <http://dx.doi.org/10.31413/nativa.v10i1.12265>.
- HE, Y. *et al.* Aquafaba from Korean Soybean II: physicochemical properties and composition characterized by nmr analysis. **Foods**, [S.L.], v. 10, n. 11, p. 2589, 26 out. 2021. MDPI AG. <http://dx.doi.org/10.3390/foods10112589>.
- JIANG, Guo-Liang *et al.* Genetic Analysis of Sugar Composition and Its Relationship with Protein, Oil, and Fiber in Soybean. **Crop Science**, [S.L.], v. 58, n. 6, p. 2413-2421, 30 ago. 2018. Wiley. <http://dx.doi.org/10.2135/cropsci2018.03.0173>.
- KURASCH, A. K. *et al.* Phenotypic Analysis of Major Agronomic Traits in 1008 RILs from a Diallel of Early European Soybean Varieties. **Crop Science**, [S.L.], v. 57, n. 2, p. 726-738, 31 jan. 2017. Wiley. <http://dx.doi.org/10.2135/cropsci2016.05.0318>.
- LEITE, D. C. *et al.* Non-destructive genotypes classification and oil content prediction using near-infrared spectroscopy and chemometric tools in soybean breeding program. **Journal Of Food Composition And Analysis**, [S.L.], v. 91, p. 103536, ago. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.jfca.2020.103536>.

LESTARI, P. *et al.* Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. **Frontiers In Plant Science**, [S.L.], v. 4, 2013. Frontiers Media SA. <http://dx.doi.org/10.3389/fpls.2013.00176>.

LORINI, Irineu *et al.* Colheita e pós-colheita de grãos. In: SEIXAS, Claudine Dinali Santos *et al.* **Tecnologia de produção de soja**. Londrina, Pr: Embrapa Soja, 2020. Cap. 14. p. 317-345. ISSN 2176-2902; n 17.

MANDARINO, J.M.G.; ROESSING, A.C. Tecnologia para a produção de óleo de soja: descrição das etapas equipamentos, produtos e subprodutos. Londrina: Embrapa Soja, 2001. 36 p. (Embrapa Soja. Documentos,171).

PATEL, M. *et al.* High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. **Theoretical And Applied Genetics**, [S.L.], v. 108, n. 8, p. 1492-1502, 14 fev. 2004. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00122-004-1590-3>.

QIN, P. WANG, T. LUO, Y. A review on plant-based proteins from soybean: health benefits and soy product development. **Journal Of Agriculture And Food Research**, [S.L.], v. 7, p. 100265, mar. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.jafr.2021.100265>.

SANTANA, D. C. *et al.* Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. **Remote Sensing Applications: Society and Environment**, [S.L.], v. 29, p. 100919, jan. 2023. Elsevier BV. <http://dx.doi.org/10.1016/j.rsase.2023.100919>.

SILVA, Alisson Santos Lopes da *et al.* Contribuições do melhoramento genético da soja. In: CASTRO, Leonardo Humberto Silva e *et al.* **Soja: estratégia e sustentabilidade produtiva**. Viçosa, MG: Ed. UFV, 2021. Cap. 6. p. 228-241. (Científica).

SOBKO, O. *et al.* Environmental Effects on Soybean (*Glycine Max (L.) Merr*) Production in Central and South Germany. **Agronomy**, [S.L.], v. 10, n. 12, p. 1847, 24 nov. 2020. MDPI AG. <http://dx.doi.org/10.3390/agronomy10121847>.

SUGIMOTO, T. *et al.* Phosphoenolpyruvate Carboxylase Level in Soybean Seed Highly Correlates to Its Contents of Protein and Lipid. **Agricultural And Biological Chemistry**, [S.L.], v. 53, n. 3, p. 885-887, mar. 1989. Oxford University Press (OUP). <http://dx.doi.org/10.1080/00021369.1989.10869369>.

TAMAGNO, S. *et al.* Dynamics of oil and fatty acid accumulation during seed development in historical soybean varieties. **Field Crops Research**, [S.L.], v. 248, p. 107719, mar. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.fcr.2020.107719>.

TODESCHINI, M. H. *et al.* Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. **Euphytica**, [S.L.], v. 215, n. 7, 19 jun. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10681-019-2439-9>.

VALADARES FILHO, S.C. *et al.*, **CQBAL 4.0**. Tabelas Brasileiras de Composição de Alimentos para Ruminantes. 2018. Disponível em: www.cqbal.com.br. Acesso em: 13 jan 2023.

VIANNA, M. S. *et al.* Selection of lineages, genetic parameters, and correlations between soybean characters. **Bioscience Journal**, [S.L.], p. 1300-1314, 2019. EDUFU - Editora da Universidade Federal de Uberlândia. <http://dx.doi.org/10.14393/bj-v35n5a2019-42656>.

WOYANN, L. G. *et al.* Multiple-trait selection of soybean for biodiesel production in Brazil. **Industrial Crops And Products**, [S.L.], v. 140, p. 111721, nov. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.indcrop.2019.111721>.

ZHOU, Z. *et al.* Assessment of Phenotypic Variations and Correlation among Seed Composition Traits in Mutagenized Soybean Populations. **Genes**, [S.L.], v. 10, n. 12, p. 975, 27 nov. 2019. MDPI AG. <http://dx.doi.org/10.3390/genes10120975>.

CARVALHO, I. R. *et al.* Soybean nutraceutical canonic interrelations and its reflections on breeding. **Agropecuária Catarinense**, v. 34, n. 3, p. 67-75, 2021. <https://doi.org/10.52945/rac.v34i3.1155>.

BATTISTI, Rafael; SENTELHAS, Paulo César. Variabilidade e mudanças climáticas: impactos na cadeia produtiva da soja. In: CASTRO, Leonardo Humberto Silva e *et al.* **Soja: estratégia e sustentabilidade produtiva**. Viçosa, Mg: Ed. UFV, 2021. Cap. 3. p. 67-113. (Científica).

CARVALHO, I. R. *et al.* Methods for estimation of genetic parameters in soybeans: an alternative to adjust residual variability. **Acta Scientiarum. Agronomy**, v. 45, 2022. <https://doi.org/10.4025/actasciagron.v45i1.56156>.

CARVALHO, I. R. *et al.* Soybean nutraceutical canonic inter-relations and its reflections on breeding. **Agropecuária Catarinense**, v. 34, n. 3, p. 67-75, 2021. <http://dx.doi.org/10.52945/rac.v34i3.1155>.

HERRERA, Gustavo Capato et al. Adaptabilidade e estabilidade de linhagens de soja na região sul do Brasil por meio de modelagem mista. **Journal of Agronomic Sciences**, v. 9, p. 185-202, 2020.

FINOTO, Everton Luis *et al.* SOWING TIMES IN ADAPTATION, STABILITY, PRODUCTIVITY, AND OIL AND PROTEIN CONTENTS OF SOYBEAN GENOTYPES. **Revista Caatinga**, v. 34, p. 799-812, 2021.

CAPÍTULO 2: Estudo da diversidade genética entre e dentro de coleções de soja do sul do Brasil

RESUMO

Diariamente novas cultivares de soja são desenvolvidas com a finalidade de atender ao mercado quanto à produtividade de grãos e superioridade agrônômica. Aliado a isso os estudos de diversidade genética auxiliam os programas de melhoramento em sua quantificação e caracterização de genótipos entre e dentro das coleções. O objetivo do trabalho foi de caracterizar e estimar a diversidade genética de genótipos com base em caracteres quantitativos, enfatizando também características relativas à qualidade industrial. Foram utilizadas quatro coleções totalizando 301 genótipos do programa de melhoramento da UNIJUÍ do estado do Rio Grande do Sul. Foi realizado o estudo da diversidade fundamentado na análise discriminante e agrupamento pelo método de Tocher. A distribuição da variabilidade entre e dentro das coleções foi avaliada pelo método da AMOVA e a discriminação dos genótipos foi realizada pelas funções linear e quadrática de Anderson, e por fim os melhores genótipos divergentes foram observados por diagramas de Venn. A análise de componentes principais mostrou dois componentes foram suficientes para explicar 88.38% da variância disponível nas características analisadas e a dispersão gráfica evidenciou que as coleções 3 e 4 são dissimilares. Nenhum grupo foi estabelecido com exclusividade ou prevalência de uma determinada coleção pelo agrupamento de Tocher. A AMOVA evidenciou que a coleção 1 é mais heterogênea e a coleção 3 é mais homogênea, além disso, 80.45% da variação encontrada está dentro das coleções. As análises discriminantes de Anderson tiveram taxa de erro de 26.44% e 19.96% para as análises Linear e quadrática, respectivamente, evidenciando que determinados fatores podem dificultar a classificação correta das coleções. A coleção de interesse, coleção 3 apresentou genótipos semelhantes entre si, 3 e 4, e 20 e 38, e ambos se mostraram divergentes do genótipo 38, e a população base apresentou bons genótipos para teores de proteína e óleo e massa de sementes e óleo. Desse modo, conclui-se que existem coleções que podem ser diferenciadas e com concentração de alelos favoráveis para características de qualidade industrial diversificada.

Palavras-Chave: Melhoramento genético, características quantitativas, Diagramas de Venn.

ABSTRACT

New soybean cultivars are developed every day in order to meet the market's needs in terms of grain productivity and agronomic superiority. Allied to this, genetic diversity studies help breeding programs in their quantification and characterization of genotypes between and within collections. The objective of this work was to characterize and estimate the genetic diversity of genotypes based on quantitative characters, also emphasizing characteristics related to industrial quality. Four collections totaling 301 genotypes from the UNIJUÍ breeding program in the state of Rio Grande do Sul were used. A diversity study was carried out based on discriminant analysis and grouping using the Tocher method. The distribution of variability between and within the collections was evaluated by the AMOVA method and the discrimination of genotypes was performed by Anderson's linear and quadratic functions, and finally the best divergent genotypes were observed by Venn diagrams. The principal component analysis showed that two components were sufficient to explain 88.38% of the variance available in the analyzed characteristics and the graphical dispersion showed that collections 3 and 4 are dissimilar. No group has been established with exclusivity or prevalence of a given collection by Tocher's grouping. AMOVA showed that collection 1 is more heterogeneous and collection 3 is more homogeneous, in addition, 80.45% of the variation found is within the collections. Anderson's discriminant analyzes had an error rate of 26.44% and 19.96% for the linear and quadratic analyses, respectively, showing that certain factors can hinder the correct classification of collections. The collection of interest, collection 3, presented genotypes similar to each other, 3 and 4, and 20 and 38, and both were divergent from genotype 38, and the base population presented good genotypes for protein and oil contents and mass of one hundred seeds and oil. Thus, it is concluded that there are collections that can be differentiated and with a concentration of favorable alleles for characteristics of diversified industrial quality.

Keywords: Genetic improvement, quantitative traits, Venn diagrams.

1. INTRODUÇÃO

A soja possui valor comercial e grande importância mundialmente. A espécie mais cultivada de soja atualmente é *Glycine max* (L.) Merrill, uma cultivar melhorada a partir da soja silvestre (*Glycine soja* Siebold e Zucc.), e domesticada na China (WEN *et al.*, 2009; VIANA *et al.*, 2022). Desde então, o crescimento do cultivo está diretamente relacionado com os programas de melhoramento, que desenvolvem diversas cultivares com alta produtividade, resistência a estresse biótico e abiótico, e entre outros (AMARAL *et al.*, 2022).

A busca por variedades produtivas e com caracteres agrônômicos superiores é desejável. Segundo Vianna *et al.* (2019) os programas de melhoramento investem principalmente no desenvolvimento em cultivares com alta produtividade de grãos, superioridade agrônômica, ampla adaptabilidade e estabilidade, além da divergência que deve existir entre os genitores, a detenção de caracteres favoráveis nestes e a alta variabilidade genética. Tal variabilidade genética é escassa frente a outras culturas. Segundo Carter *et al.* (2016) somente 17 cultivares de soja correspondem a cerca de 86% do parentesco na população de soja dos Estados Unidos lançadas entre 1947 e 1988, e esses cultivares são utilizados em programas de melhoramento em todo o mundo.

Devido ao tipo de reprodução da soja, a perda da diversidade é um aspecto muito importante. Essa diversidade tem se mostrado imprescindível para precaver programas de melhoramento a respeito da perda do ganho genético, além de ser uma preocupação significativa para o futuro e o desenvolvimento de cultivares da soja. Entretanto, um ponto negativo é que o processo de melhoramento geralmente não envolve etapas que avaliem os impactos da seleção na diversidade genética (VIANA *et al.*, 2022). Por isso a importância dos estudos de diversidade, principalmente na etapa de seleção, uma vez que a heterose será expressa de melhor forma quando pais contrastantes são cruzados durante o processo de hibridação (CRUZ; REGAZZI; CARNEIRO, 2012).

Existem diversos estudos e técnicas para analisar a diversidade, e graças as estimativas de dissimilaridade o melhorista consegue inferir a respeito da semelhança e a diferença entre os genótipos. A análise de dispersão gráfica, por exemplo, permite avaliar a diversidade entre os genótipos, geralmente bi ou tridimensionalmente, evitando a perda de informações do grau de similaridade de indivíduos do mesmo grupo quando acontece a utilização de métodos de agrupamento com matrizes de grau elevado, se mostrando assim, uma ótima alternativa para o estudo da diversidade (CRUZ; FERREIRA; PESSONI, 2011).

Outros tipos de análises também são utilizados para o estudo da divergência genética. Geralmente ela tem sido quantificada através de métodos multivariados como por exemplo a análise de componentes principais e métodos aglomerativos, sendo a escolha do método ideal definida pela precisão que o pesquisador almeja, a facilidade da análise e a forma de obtenção dos dados originais (CRUZ, REGAZZI; CARNEIRO, 2012). Conseqüentemente, métodos classificatórios como análise discriminante permitem identificar variáveis que separam grupos e permitem a predição de genótipos quanto ao pertencimento ao grupo mais adequado individualmente (BIANCHINI *et al.*, 2020).

Um enfoque importante é quantificar e caracterizar a diversidade genética entre genótipos de uma coleção de trabalho em relação a outras coleções disponíveis no banco de germoplasma para ter uma medida comparativa da variabilidade disponível e de materiais recorrentes caso seja necessário a sua ampliação. Estudos da diversidade entre e dentro de coleções permitem estabelecer um histórico de possíveis causas em que a variabilidade foi preservada, ampliada ou restringida tendo em vista os objetivos e as ações diretas, ou indiretas, do melhorista para a formação das coleções base.

Há tendência do mercado consumidor, em especial quando se trata de exportações, de se ter um padrão de cultivares com maior teor de óleo e proteína para usos diversos. Entretanto, em função da remuneração do produto da soja se fundamentar na quantidade produzida, pouca ênfase foi dada ao aumento dos componentes determinantes da qualidade industrial e, assim, as coleções estabelecidas, quando avaliadas no contexto da variabilidade das características de qualidade, podem manifestar diferentes padrões uma vez que a seleção e a fixação de formas alélicas favoráveis podem ter ocorrido de forma totalmente aleatória uma vez que foram involuntariamente negligenciadas em detrimento do alcance da maior produtividade de grãos.

Além disso, os estudos de diversidade genética da soja têm gerado informações importantes para o gerenciamento e a avaliação de germoplasmas. Geralmente, eles são realizados com base em informações agronômicas e morfológicas, e genealógicas (SILVA *et al.*, 2017). Sendo assim, o presente trabalho tem o objetivo de caracterizar e estimar a diversidade genética de genótipos da soja (*Glycine max* (L.) Merrill), envolvendo diversas coleções, com base em caracteres agronômicos quantitativos, enfatizando também características relativas à qualidade industrial da soja.

2. MATERIAL E MÉTODOS

2.1. Descrição do Experimento

O material genético utilizado é proveniente do programa de Melhoramento Genético Linha Grãos e Coberturas da Universidade Regional do Noroeste do Rio Grande do Sul (UNIJUÍ). Experimento Nutraceuticos realizado no município de Campos Borges, estado do Rio Grande do Sul, com clima definido como clima subtropical úmido. O delineamento utilizado foi de blocos aumentados, em que quatro coleções foram testadas.

Tabela 1. Identificação das progênies e do número de indivíduos correspondentes a cada uma. Sendo F₂, população base; CV, cultivares testemunhas, CB e F₄ e F₇ são segregantes, sendo F₄ e F₇ linhas avançadas

Progênies (Classes)	Genótipos
CB	154
CV	22
Segregantes de F ₄ e F ₇	87
F ₂	38
Total	301

Os dados referentes a dez características quantitativas e de qualidade industrial foram coletados e transformados em uma tabela de médias.

2.2. Caracteres Agronômicos avaliados

As dez características analisadas foram massa de 100 sementes (g) (MCS) que representa um dos componentes primários da produção de grãos e outras representativas da qualidade industrial que são: Proteína (%) (PTN), Óleo (%) (OLE), Fibra (%) (FIB), Cinza (%) (CIN), Ácido Palmítico (%) (APA), Ácido Esteárico (%) (AES), Ácido Oleico (%) (AOL), Ácido Linoleico (%) (ALI), Ácido Linolênico (%) (ALN).

Inicialmente, as amostras de 100 gramas de sementes da soja foram limpas e aferidas. Posteriormente, o valor de umidade da foi corrigido para 13% de umidade. Essa umidade é indicada para o armazenamento de grãos e evitar danos mecânicos e latentes, além de pragas, uma vez que sementes colhidas superior a ótima tende a sofrer danos mecânicos e latentes e inferior a ótima tende a sofrer danos imediatos (LORINI *et al.*, 2020).

Após a correção da umidade relativa foram estabelecidas amostras de 55 gramas foram levadas ao espectrofotômetro para a análise das características PTN, OLE, FIB, CIN, APA, AES, AOL, ALI, ALN.

2.3. Procedimentos Biométricos

Para fins de reconhecimento de padrões

Inicialmente, foi realizado estudo da diversidade genética, fundamentado na análise discriminante, em que a ênfase é realçar a diversidade dos centroides representativos das quatro coleções estudadas. Assim, a análise de componentes principais foi realizada depois que os dados médios foram padronizados. Os padrões de dispersão permitiram o reconhecimento da variabilidade entre centroides e entre genótipos que constituíam cada coleção.

Nestes estudos, procurou-se enfatizar a diversidade relativa das coleções 3 (Segregantes de F4 e F7) e 4 (população F2) em relação as demais tendo em vista a concentração de esforços no melhoramento de tais coleções.

O reconhecimento dos padrões e seu grau de diversidade foi avaliado por meio da técnica de agrupamento de otimização de Tocher, evidenciando a similaridade entre genótipos provenientes da mesma coleção ou de coleções diferentes. Para este estudo, utilizou-se a matriz de distância euclidiana média padronizada cuja distância é dada por:

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

Em que:

v = número de características estudadas

Y_{ij} = valor observado (padronizado) do i -ésimo genótipo, para a j -ésima característica

Para fins de quantificação da variabilidade entre e dentro de coleções

A distribuição da variabilidade genética entre e dentro das coleções foi avaliada pela metodologia AMOVA (análise de variância molecular), descrita por Excoffier *et al.* (1992). Nesse método, a matriz de distâncias entre todos os pares de genótipos é utilizada em um esquema de análise de variância hierarquizada, produzindo estimativas de componentes de variância que mede a diferenciação entre elas.

O esquema da análise de variância molecular (AMOVA), com dados agrupados em dois níveis hierárquicos, é apresentado a seguir:

Tabela 2. Esquema da AMOVA com dados em dois níveis hierárquicos. GL, graus de liberdade; SQ, soma de quadrados; QM, quadrado médio; E(QM) esperança do quadrado médio; g, número de coleções; SQE, soma de quadrados entre; QME, quadrado médio entre; N, número de genótipos; SQD, soma de quadrado dentro, WMD, quadrado médio dentro; SQ, soma de quadrados totais.

Fonte de Variação	GL	SQ	QM	E(QM)
Entre Coleções	g-1	SQE	QME	$\hat{\sigma}_i^2 + \tilde{N}\sigma_p^2$
Dentro Coleções	N-g	SQD	QMD	$\hat{\sigma}_i^2$
Total	N-1	SQT	-	σ_T^2

em que N representa o número médio de genótipos amostrados por coleção. Com amostras de tamanhos desiguais, \tilde{N} é obtido por:

$$\tilde{N} = \frac{N - \sum_i \frac{N_i^2}{N}}{g-1}, \text{ sendo } N_i \text{ o número de genótipos da } i\text{-ésima coleção. E } g \text{ igual ao}$$

número de coleções.

Os estimadores dos componentes de variância que expressam as diferenciações entre e dentro de subpopulações são obtidos por meio de:

$$\hat{\sigma}_i^2 = QMD \text{ e } \hat{\sigma}_i^2 = \frac{QME - QMS}{\tilde{N}}$$

O coeficiente de correlação entre genótipos, amostrados aleatoriamente entre coleções, é dado por:

$$r = \phi_{ST} = \frac{Cov(Y_{ij}, Y_{ij'})}{\sqrt{V(Y_{ij})V(Y_{ij'})}}$$

logo, $\phi_{ST} = \frac{\sigma_p^2}{\sigma_T^2}$, é uma medida da diversidade relativa entre as coleções avaliadas.

Um estudo das diferenças médias entre as coleções foi apresentado com apresentação dos valores médios em gráficos em forma de teia.

Para fins de discriminação das coleções

Neste estudo, foi utilizada a análise discriminante proposta por Anderson (1958), linear e quadrática, em que se consideram as informações dos genótipos sabidamente pertencentes a diferentes coleções. A partir dessas informações são geradas funções, que são combinações lineares das características avaliadas e têm por finalidade promover a melhor discriminação entre os indivíduos, alocando-os em suas devidas coleções.

Uma vez obtidas as funções discriminantes, foi avaliado a sua eficácia, que é dependente do grau de dissimilaridade entre as coleções analisadas e, principalmente, da quantidade e qualidade das variáveis consideradas na discriminação. Como as funções discriminantes são obtidas a partir de análises prévias de observações que se supõe serem, de fato, pertencentes às coleções pode-se calcular a probabilidade de má classificação, reclassificando toda observação até então disponível. A classificação de uma observação pertencente a uma coleção π_j em outra é indicativo de menor eficiência da função discriminante estimada, contribuindo para o acréscimo na taxa de erro aparente.

Para fins de reconhecimento de genótipos divergentes e de bom desempenho

Uma avaliação simultânea do potencial e da diversidade de genótipos de coleções de interesse (coleções 3 e 4) foi realizado. As características utilizadas foram massa de cem sementes, teor de proteína e teor de óleo, e destacado o desempenho dos 20% melhores genótipos em diagrama de Venn. Esta análise foi complementada com estudo multivariado por meio da técnica de componentes principais aplicados unicamente aos genótipos de cada coleção referida (coleções 3 e 4).

Todas as análises estatísticas foram realizadas utilizando o programa Genes (CRUZ, 2016).

3. RESULTADOS E DISCUSSÃO

Estudo da diversidade genética e reconhecimento de padrões de coleções

É reconhecido a importância da ação de programas de melhoramento genético para a obtenção de materiais superiores que atendam aos interesses do produtor e consumidor da soja. Entretanto, tendo em vista alguns aspectos genéticos relativos às correlações desfavoráveis entre caracteres importantes de qualidade industrial e destes em relação a caracteres de produtividade a variabilidade, quando estudada no contexto das características que refletem a qualidade industrial, não tem um padrão definido. Pela remuneração direcionada à quantidade produzida, espera-se que programas diferentes de melhoramento (que geram diferentes

coleções) possam ter buscado a fixação de formas alélicas favoráveis para estas características. Entretanto, como as características de qualidade foram negligências, o padrão de variabilidade pode, ou não, ser completamente discordante de uma coleção para outra levando à necessidade de ser conhecido. A análise discriminante, por componentes principais, apresentada na Tabela 1 e Figura 1 permite inferir sobre o padrão de quatro coleções estudadas com ênfase em características de qualidade industrial.

A análise de componentes principais (Tabela 1) mostrou que dois componentes foram suficientes para explicar o 88.38% da variância disponível nas dez características principais originárias considerando os quatro centroides representativos das coleções. Estatisticamente, seriam necessários apenas 3 componentes (menor entre os valores $v - \text{número de variáveis} + c - 1$, sendo c o número de coleções). Nesse caso, uma vez que o primeiro e o segundo componentes conseguiram reter acima de 80% da variabilidade existente de maneira satisfatória, a dispersão gráfica bidimensional é recomendada (CRUZ; REGAZZI; CARNEIRO 2012). A dispersão gráfica é demonstrada na Figura 1.

Ao observar a dispersão gráfica na Figura 1, num primeiro momento é dito que as coleções são consideradas similares, visto que as quatro coleções apresentam padrões de dispersão próximos e aleatório entre si (Figura 1a) com bastante sobreposição de pontos representativos dos genótipos das quatro coleções. Entretanto, ao analisar separadamente as coleções 3 (segregantes avançadas em F4 e F7) e 4 (população base F2) na Figura 1b, é possível perceber uma diferenciação entre elas, uma vez que a coleção dois encontra-se majoritariamente situada na região superior e a coleção quatro na região inferior do gráfico, dito isto, é possível dizer que as coleções 3 e 4 são consideradas dissimilares.

Percebe-se, que a coleção 1, CB, apresenta o maior grau de dispersão pelo gráfico, não se diferenciando das demais coleções. Isso indica que a variabilidade dentro da coleção é alta, porém ao se comparar a variabilidade desta entre as outras coleções essa variabilidade é pequena.

Tabela 1. Diversidade genética entre acessos representativos de quatro coleções de soja, por meio da técnica de componentes principais. Var, estimativa de variância; Acum, variância acumulada; MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.

Var	Acum	MCS	PTN	OLE	FIB	CIN	APA	AES	AOL	ALI	ALN
6.87	68.70	0.3622	0.0095	0.3569	-0.2671	-0.3261	0.3374	-0.3721	-0.3193	0.3702	-0.2691
1.97	88.38	0.1689	0.6155	-0.2511	-0.5008	0.2135	-0.0764	0.1462	0.2278	0.0159	-0.3923
1.16	100,00	-0.192	0.4674	0.0243	0.1184	0.3933	0.4215	0.0773	-0.4121	0.2236	0.4145

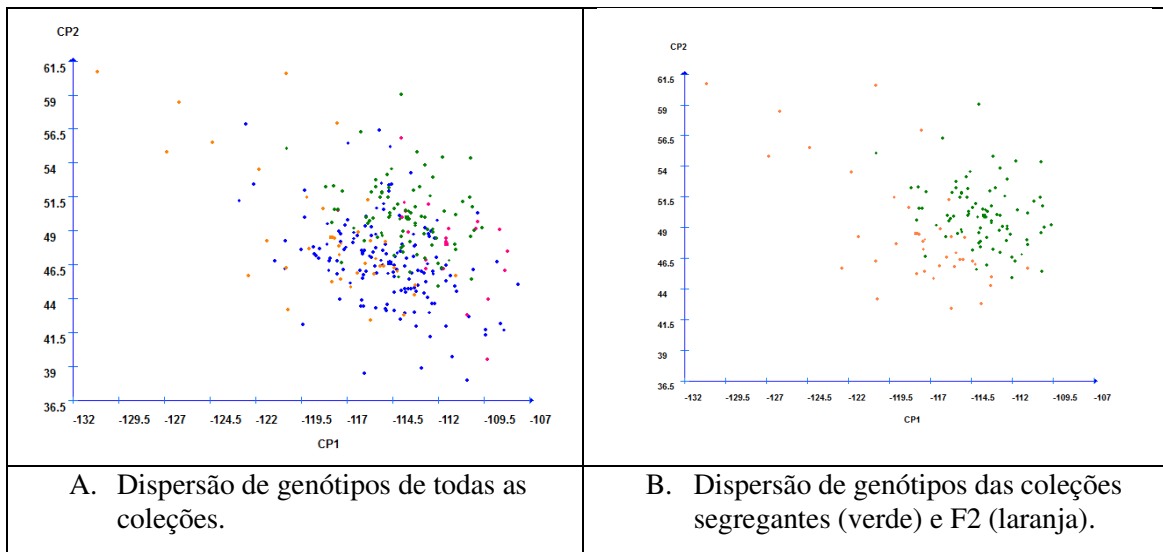
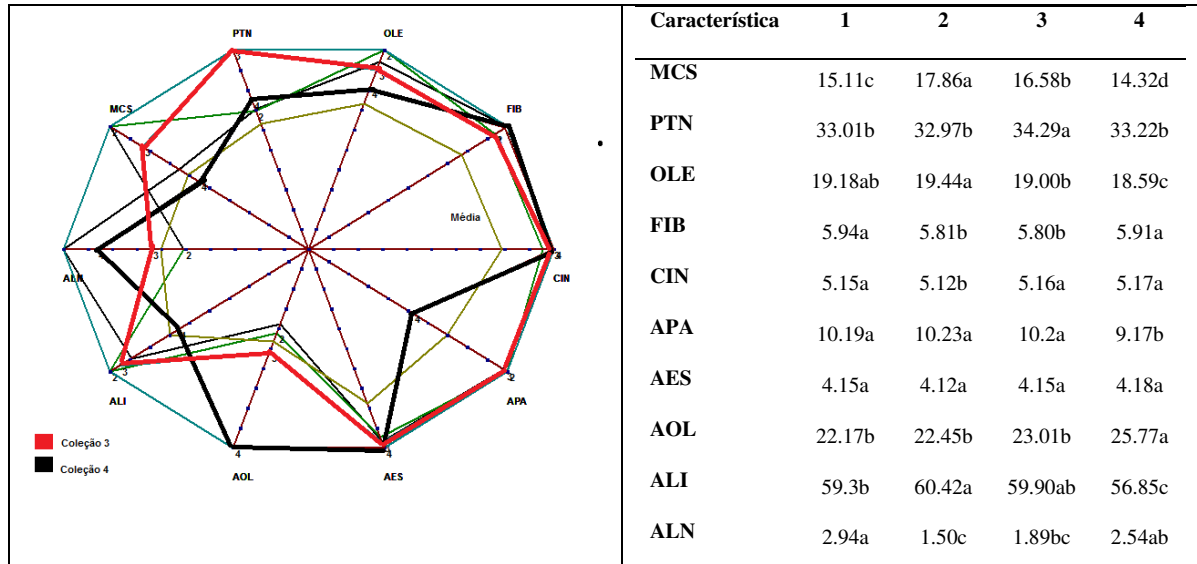


Figura 1. Dispersão gráfica de escores de genótipos de soja representativos de quatro coleções, destacando 1 (azul) origem CB, 2 (vermelho (CV), 3 segregantes (verde) e 4 população F2 (laranja). CP1, componente 1; CP2, componente 2.

Os resultados revelam que as coleções estão sendo estabelecidas com enfoque diferente de forma que em algumas priorizaram, direta ou indiretamente, o enriquecimento de formas alélicas favorável para um componente de qualidade e outros contemplaram outros componentes. O reconhecimento deste fato, além de importante per se, orientará o melhorista a estabelecer combinações híbridas para que a deficiência de desempenho em genótipos de uma coleção possa ser superada pela inclusão de fontes favoráveis oriundas de outra coleção.

Um estudo mais pormenorizado, com base nos valores médios obtidos pelas coleções, pode ser visualizado na Tabela 2. Pelo teste Tukey, foi possível diferenciar estatisticamente a 5% de probabilidade a coleção três das demais coleções, indicando que a característica MCS apresenta alto desempenho, mas é superada, estatisticamente, pela coleção 2 (CV). Além dessa característica, foi possível diferenciar a coleção alvo estatisticamente para as características PTN, sendo dela a maior média entre as quatro coleções. Para as demais oito características, sendo elas, OLE, FIB, CIN, APA, AES, AOL, ALI, ALI, ALN, a coleção três se diferenciou de algumas coleções, porém não se sobressaiu em relação as demais.

Tabela 2. Desempenho médio das coleções com destaque para os genótipos representativos da população segregante (vermelho) e F2 (preto) para dez características agrônômicas. MCS: massa de cem sementes; PTN: Proteína; OLE: Óleo; FIB: Fibra; CIN: Cinza; APA: ácido palmítico; AES: ácido esteárico; AOL: ácido oleico; ALI: ácido linoleico; ALN: ácido linolênico.



(*) Médias seguidas pelas mesmas letras, na horizontal, não diferem entre si pelo teste Tukey a 5% de probabilidade.

A Tabela 2 é acompanhada de um gráfico, na forma de teia, em que se evidencia o grande contraste entre as coleções 3 e 4. Pelo teste Tukey, apenas as médias das características CIN, AES e ALN não diferiram para estas duas coleções, porém existe uma clara diferenciação para as demais características. A coleção 3 tem estimativas superiores em relação à coleção 4 para MCS, PTN e OLE, mas estimativas inferiores em relação a AOL. Assim, apesar de serem duas coleções em uso pelo programa de melhoramento elas requerem atenção diferenciada. A escolha de como tais coleções são conduzidas, e as estratégias de seleção, direta ou indireta, podem conduzir a um produto muito diferenciado ou distanciado dos objetivos de manter a qualidade industrial.

Uma forma auxiliar de constatar o padrão de diversidade dos genótipos das coleções é apresentado na Tabela 3. O agrupamento de Tocher revela que nenhum grupo (não unitário) foi estabelecido com exclusividade ou prevalência de uma determinada coleção. Adicionalmente informa que algum tipo de diferenciação possa existir quando se realça coleções particulares, como destacado na Figura 1b.

Tabela 3. Padrão de agrupamento de genótipos de soja representativos de quatro coleções (1 = CB, 2 = CV, 3= segregantes e 4 populações F2), obtido a partir da distância euclidiana média padronizada e uso do método de otimização de Tocher.

Grupos	Acessos (Coleções)
I	1 1 3 3 1 3 1 3 3 3 3 3 1 1 1 3 3 3 3 1 3 3 1 3 1 3 3 3 1 1 3 3 3 1 1 3 3 1 1 1 1 1 1 3 1 3 1 1 3 4 1 3 3 2 1 3 1 2 1 3 1 1 1 3 3 3 1 1 1 1 1 1 3 2 3 2 3 1 1 1 1 1 4 1 4 1 1 1 1 3 1 3 1 1 3 4 1 4 1 1 3 1 3 1 3 1 3 3 1 3 3 3 2 3 3 3 3 1 1 1 3 1 1 3 1 1 1 3 3 3 1 1 3 2 4 1 1 4 1 2 4 1 1 1 4 1 1 4 3 1 1 4 1 1 3 3 1 1 3 1 2 1 1 1 1 4 2 3 4 3 1 1 1 3 3 1 3 3 1 3 1 1 1 2 3 3 1 1 1 1 1 1 1 4 3 1 1 4 1 1 4 4 1 3 4 3 2 1 1 3 4 4 1 1 3 1 1 2 1 1 2 3 3 3 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 4 3 4 1 2 1 3 1 4 1 1 4 1 1 4 3 4 1 4 1 1 3 1 4 4 1 1 1 1 2 3 4 1 3 2 2 1 1 3 2 4
II	4 4 1 4 4 4 3 3 1 1 1 1
III	1 1 1 4
IV	1 2
V	2
VI	4

Variabilidade entre e dentro coleções

A Tabela 4, apresenta a variância média de cada população e a análise de variância molecular (AMOVA), que foi utilizada com a finalidade de particionar a variação entre e dentro das coleções. Resultado demonstra que a coleção que apresentou maior distância média, ou seja, maior dispersão, foi a coleção 1, cerca de 34.93, seguida das coleções 3, 4 e 2, com valores de 13.49, 11.88 e 4.66%. É interessante notar que a coleção 2 apresenta indivíduos mais homogêneos, eles se apresentam menor distância média, e a coleção 1 constitui-se de indivíduos mais heterogêneos.

Tabela 4. Resultado da diversidade entre e dentro de coleções obtido pelo método da AMOVA de Excoffier *et al.* aplicado à matriz de dissimilaridade representada pelo quadrado da distância euclidiana padronizada. Pop, indica as populações; Ni, número de indivíduos; DistMédia, distância média; FV, Fonte de Variação; GL, graus de liberdade; SQ, soma de quadrado, QM, quadrado médio.

Pop	Ni	DistMédia	FV	GL	SQ	QM	$\hat{\sigma}^2$	$\hat{\sigma}^2(\%)$
1	154	34.93	Entre	3	10.7921	3.5974	0.0532	19.554
2	22	4.66	Dentro	297	64.9582	0.2187	0.2187	80.446
3	87	13.49	Total	300	75.7503	0.2525	0.2719	100
4	38	11.88	$\emptyset_{ST} = 0.1955$					

A maior parte da variação foi particionada entre e dentro das coleções, que corresponde a aproximadamente 80.45% dentro das coleções. A parcela inferior da variação, aproximadamente 19.55%, foi encontrada entre as coleções. Isso indica que as coleções apresentam estrutura genética considerável. Além disso, o \emptyset_{ST} , parâmetro de diferenciação genética entre as populações foi de aproximadamente 0.20. Tal resultado reforça as análises anteriores, indicando uma alta variabilidade dentro das coleções, mas dificultando a distinção entre as quatro coleções. Por isso, é importante utilizar outras metodologias complementares para compreender e inferir a respeito das diferenças existentes, principalmente, entre a coleção três e as demais. Wen *et al.* (2009) encontraram uma variação particionada em subpopulações geográficas da soja anual silvestre na China de 2.70% e 9.30% entre e dentro respectivamente. Tiwari *et al.* (2019) encontraram a variância entre e dentro em valores de 31.39% e 61.69% comparando três grupos com 148 acessos indianos de soja, e obtiveram estimativa média aproximadamente de \emptyset_{ST} 0.84 dos três grupos. Viana *et al.* (2022) encontraram estimativas de 11.33 e 88.66 de variação entre e dentro das populações de soja estudada, e \emptyset_{ST} de 0.11.

Discriminação entre as populações

Como as coleções são diferentes, as análises linear e quadrática de Anderson apresentam a proposta de classificar os indivíduos em suas respectivas coleções de maneira correta, minimizando a probabilidade da classificação total equivocada (CRUZ *et al.*, 2022). A análise linear classificou os indivíduos com uma Taxa de Erro Aparente (TEA) de 26.24%. Destaca-se, o fato de que esta análise linear não classificou erroneamente nenhum indivíduo das coleções 2 e 3 dentro da coleção 4, e esta última teve uma taxa de acerto de 89.47%, não classificando nenhum indivíduo dentro da coleção 2. Já a análise quadrática, apresentou uma TEA de 18.93%, repetindo a classificação em que nenhum indivíduo das coleções 2 e 3 são erroneamente

classificados dentro da coleção 4, sendo o inverso verdadeiro, pois esta última não teve nenhum indivíduo classificado dentro das coleções 2 e 3.

Quando a taxa de erro aparente é alta deve-se concluir que as coleções analisadas não são suficientemente diferenciadas para que possam ser distinguidas por meio das funções discriminantes. Ou, as coleções analisadas, apesar de serem diferenciadas, não puderam ser distinguidas em razão da quantidade e qualidade das variáveis consideradas na discriminação. Na literatura, estudos utilizando a análise discriminante de Anderson são diversos e algumas análises são bem-sucedidas e outras nem tanto. Amaral *et al.* (2022) obtiveram uma taxa de erro aparente de 50.59% ao classificar 1517 genótipos de soja em 11 populações diferentes. Bianchini *et al.* (2020) verificaram TEA de 36.5% ao classificar diferentes técnicas de manejo de plantas daninhas na cultura da soja,

As classificações errôneas dos genótipos podem se dar por quatro motivos principais. Primeiramente, as populações/coleções podem ser muito homogêneas em sua genealogia, ainda pode ocorrer que o número de variáveis analisadas é insuficiente para a classificação correta dos genótipos, pode acontecer ainda que tais características apresentam baixa qualidade discriminatória, e por fim, a abordagem estatística é inadequada (CRUZ; CARNEIRO; REGAZZI, 2014).

Tabela 5. Resultado da análise discriminante, linear e quadrática, proposta por Anderson considerando informações de genótipos de soja representativos de quatro coleções. TEA, taxa de erro aparente.

Coleção*	Linear				Quadrática			
	1	2	3	4	1	2	3	4
1	114	11	15	14	119	9	18	8
2	2	15	5	0	0	19	3	0
3	12	16	59	0	9	7	71	0
4	3	0	1	34	3	0	0	35
TEA	26.24	Erros	79/301		18.93	Erros	57/301	

(*)1 = CB; 2 = CV; 3 = segregantes; 4 = população F2.

Genótipos divergentes e de melhor potencial das coleções 3 e 4

Considerado o interesse particular de utilizar genótipos da coleção 3, representada por linhagens segregantes em F4 e F7 conduzidos pelo programa de melhoramento Universidade Regional do Noroeste do Rio Grande do Sul (UNIJUÍ), torna-se fundamental uma boa caracterização e o entendimento da diversidade genética desta coleção e uma comparação do

desempenho relativo em relação a outras três coleções com genótipos utilizados nesta experimentação.

Ao observar os escores dos melhores 20 genótipos presentes dentro da coleção (Figura 2), percebe-se que os genótipos 3, 4, 61, 62 e 83 são genótipos melhores para duas características em questão, que são a massa de cem sementes e o teor de proteína. Ademais, os genótipos 10, 13, 20, 27, 28 e 38 são considerados bons para as características massa de cem sementes e teor de óleo. Ainda é possível observar no Diagrama de Venn que não foram encontrados genótipos que apresentem melhores estimativas para o teor de óleo e de proteína e que também para massa de cem sementes, e teores de proteína e óleo.

A dificuldade em encontrar genótipos que sejam destacáveis para os teores de proteína e óleo se dá principalmente devido a correlação negativa já encontrada entre estas duas características em estudos anteriores, estimativa de -0.73. Porém, é possível perceber dentro dos genótipos selecionados, a dispersão destes no gráfico de componentes principais mostra a existência de genótipos mais semelhantes e divergentes entre si. Os genótipos 3 e 4, assim como os genótipos 20 e 38, são semelhantes entre si, que por sua vez, são muito divergentes do genótipo 28. A distribuição dos genótipos auxilia o melhorista no planejamento de cruzamentos, visando explorar a variabilidade entre os genótipos, o que justifica a relação dos genótipos entre si.

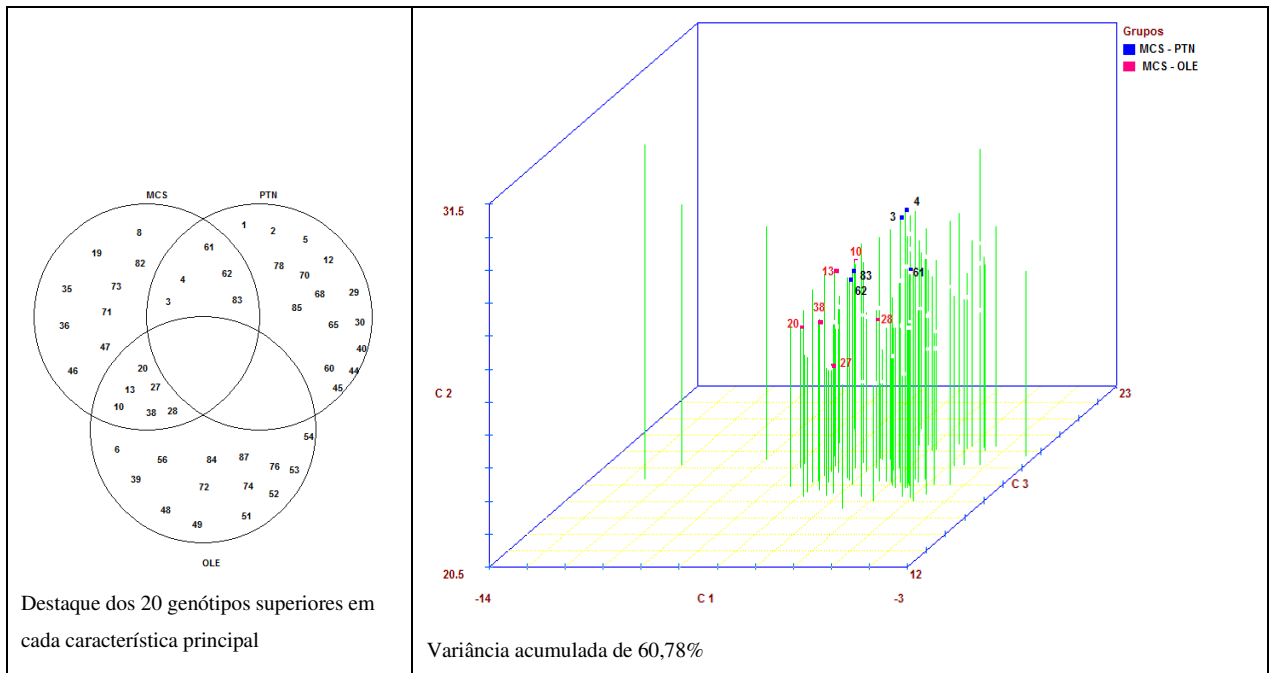
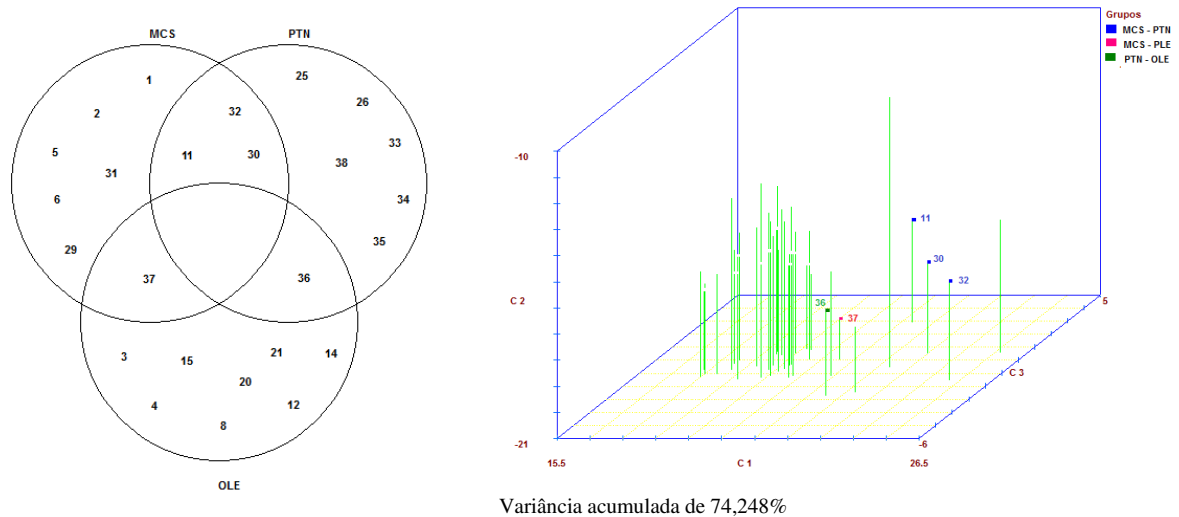


Figura 2. Dispersão gráfica de escores de genótipos de soja representativos da coleção caracterizada como *população segregante*, em relação a eixos estabelecidos por componentes principais. Destaque em relação aos genótipos superiores considerando os conjuntos de interseção de superioridade de desempenho conforme ilustrado no diagrama de Venn. MCS, Massa de cem sementes; PTN, teor de proteína; OLE, teor de óleo; C1, componente 1; C2, componente 2, C3, componente 3.

Ao considerar a população base (F_2) pode-se perceber que existem melhores genótipos que atendem a interseção de serem melhores para MCS e PTN, como no caso dos genótipos 3, 11 e 30. Outros genótipos foram destacáveis para o OLE e MCS, como no caso do genótipo 37. Verificou-se também, genótipos que foram melhores para OLE e PTN, como no caso do genótipo 36. Sendo assim, o programa de melhoramento deve investir nestes genótipos, em especial no genótipo 36, uma vez que ele apresenta potencial simultâneo para as características de PTN e OLE.



Destaque dos 10 genótipos superiores em cada característica principal

Figura 3. Dispersão gráfica de escores de genótipos de soja representativos da coleção caracterizada como *população F2*, em relação a eixos estabelecidos por componentes principais. Destaque em relação aos genótipos superiores considerando os conjuntos de interseção de superioridade de desempenho conforme ilustrado no diagrama de Venn. MCS, Massa de cem sementes; PTN, teor de proteína; OLE, teor de óleo; C1, componente 1; C2, componente 2, C3, componente 3.

4. CONCLUSÃO

A análise de reconhecimento de padrões evidenciou a existência de coleções diferenciadas e com concentração de alelos favoráveis para características de qualidade industrial diversificada;

A análise discriminante quadrática apresentou taxa de erro aparente elevada (estimativa de 18.33%) mas revelou que a discriminação entre duas coleções de interesse é elevada indicando que as características de qualidade foram, direta ou indiretamente, alteradas de forma divergente e contribuíram para esta diferenciação.

Foi identificado em duas coleções de interesse, os genótipos de melhor desempenho e que reúnem desempenho destacável para as características como massa de cem sementes (MCS) e teor de proteína (PTN) ou massa de cem sementes (MCS) e teor de óleo (OLE). Entretanto, a interseção de genótipos de melhor desempenho simultâneo para estas três características promovem, como ilustrado em diagrama de Venn, conjunto vazio, determinado principalmente pela associação negativa entre OLE e PTN.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, L. O. *et al.* Artificial Neural Network for Discrimination and Classification of Tropical Soybean Genotypes of Different Relative Maturity Groups. **Frontiers In Plant Science**, [S.L.], v. 13, 12 jul. 2022. Frontiers Media SA. <http://dx.doi.org/10.3389/fpls.2022.814046>.
- ANDERSON, T. W. An introduction to multivariate statistical analysis. New York, John Wiley & Sons. 242p. 1958.
- BIANCHINI, A. *et al.* Multivariate analysis using a discriminant method for evaluating the techniques of weed management in soybean crop. **Planta Daninha**, [S.L.], v. 38, p. 12345-123456, 2020. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-83582020380100077>.
- CARTER, T. E. *et al.* Genetic Diversity in Soybean. **Agronomy Monographs**, [S.L.], p. 303-416, 12 dez. 2016. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. <http://dx.doi.org/10.2134/agronmonogr16.3ed.c8>.
- CRUZ, Cosme Damião *et al.* Classificação: abordagens de aprendizado de máquinas e inteligência computacional. In: NASCIMENTO, Moysés *et al.* **ABORDAGENS BIOMÉTRICAS PARA RECONHECIMENTO DE PADRÕES, CLASSIFICAÇÃO E PREDIÇÃO NAS CIÊNCIAS AGRÁRIAS**. Londrina, PR: Mecenasa, 2022. Cap. 5. p. 103-166.
- CRUZ, Cosme Damião. Genes Software-extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum. Agronomy**, v. 38, p. 547-552, 2016.
- CRUZ, Cosme Damião; FERREIRA, Fábio Medeiros; PESSONI, Luiz Alberto. **Biometria Aplicada ao Estudo de Diversidade Genética**. Visconde do Rio Branco, Mg: Suprema, 2011. 620 p.
- CRUZ, Cosme Damião; REGAZZI, Adair José; CARNEIRO, Pedro Crescêncio Souza. **Modelos Biométricos Aplicados ao Melhoramento do Genético de Plantas**. 4. ed. Viçosa, MG: Ed. UFV, 2012. 514 p.
- CRUZ, Cosme Damião; CARNEIRO, Pedro Crescêncio Souza; REGAZZI, Adair José. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa, MG: Ed. UFV, 2014. 668 p.
- EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M. **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data**. *Genetics*. v. 131, p. 479-491, 1992.
- LORINI, Irineu *et al.* Colheita e pós-colheita de grãos. In: SEIXAS, Claudine Dinali Santos *et al.* **Tecnologia de produção de soja**. Londrina, Pr: Embrapa Soja, 2020. Cap. 14. p. 317-345. ISSN 2176-2902; n 17.
- SILVA, Francisco Charles dos Santos *et al.* Importância econômica e evolução do melhoramento. In: SILVA, Felipe *et al.* **Melhoramento da Soja**. Viçosa, MG: Ed. UFV, 2017. Cap. 1. p. 9-29.
- TIWARI, S. *et al.* Genetic diversity and population structure of Indian soybean (*Glycine max* (L.) Merr.) as revealed by microsatellite markers. **Physiology And Molecular Biology Of Plants**, [S.L.], v. 25, n. 4, p. 953-964, 4 jun. 2019. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s12298-019-00682-4>.

VIANA, J. P. G. *et al.* Impact of multiple selective breeding programs on genetic diversity in soybean germplasm. **Theoretical And Applied Genetics**, [S.L.], v. 135, n. 5, p. 1591-1602, 27 fev. 2022. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00122-022-04056-5>.

VIANNA, M. S. *et al.* Selection of lineages, genetic parameters, and correlations between soybean characters. **Bioscience Journal**, [S.L.], p. 1300-1314, 2019. EDUFU - Editora da Universidade Federal de Uberlandia. <http://dx.doi.org/10.14393/bj-v35n5a2019-42656>.

WEN, Z. *et al.* Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. **Theoretical And Applied Genetics**, [S.L.], v. 119, n. 2, p. 371-381, 18 Maio 2009. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00122-009-1045-y>.