UNIVERSIDADE FEDERAL DE VIÇOSA

Genetic parameters estimation and QTL mapping in *Sesamum indicum* L.

Edilson Marques Junior
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2024

EDILSON MARQUES JUNIOR

Genetic parameters estimation and QTL mapping in *Sesamum indicum* L.

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Carlos E. M. dos Santos

Co-advisers: Guilherme da S. Pereira
Kaio O. das Gracas Dias

VIÇOSA - MINAS GERAIS
2024

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

T
M357g
2024

Marques Junior, Edilson, 1993-
Genetic parameters estimation and QTL mapping in *Sesamum indicum* L. / Edilson Marques Junior. – Viçosa, MG, 2024.
1 tese eletrônica (88 f.): il. (algumas color.).

Texto em inglês.
Orientador: Carlos Eduardo Magalhães dos Santos.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2025.074
Modo de acesso: World Wide Web.

1. Gergelim - Melhoramento genético. 2. Marcadores genéticos. I. Santos, Carlos Eduardo Magalhães dos, 1980-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.85

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

EDILSON MARQUES JUNIOR

Genetic parameters estimation and QTL mapping in *Sesamum indicum* L.

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: November 13, 2024.

Assent:

Edilson Marques Junior
Author

Carlos Eduardo Magalhaes dos Santos
Adviser

To God, my parents, brothers, nephews, other family members, and friends who
contributed to this achievement...
I dedicate it.

ACKNOWLEDGMENTS

Firstly, to God, strength and refuge in my life!

To the Universidade Federal de Viçosa and the Postgraduate Program in Genetics and Breeding for the training opportunity.

To CNPq - National Council for Scientific and Technological Development and FAPEMIG - Minas Gerais State Research Support Foundation, for the period of granting the scholarship.

To Sebra Agrícola SA, for this project's structural support and financing.

To Professor Dr. Carlos Eduardo Magalhães dos Santos for his guidance on this journey and for encouraging me always to do my best.

To the Co-Advisors, Dr. Guilherme da Silva Pereira and Dr. Kaio Olímpio das Graças Dias, for their patience, guidance, and the countless pieces of knowledge shared.

To the examining board members for accepting the invitation to participate and contributing to the enrichment of this work.

To my parents, José Renato Marques and Sônia Maria Silva Marques, for their patience, affection, support, prayers, and advice. Without you, I wouldn't be here.

To my brothers Maria Crélia, José Carlos, Ledriana, Claudemir, Lucas, and Vaniele, to my nephews, to my brothers-in-law and sisters-in-law, for support and sincere friendship, you are an example in my life.

To my girlfriend and future wife, Lídia do Nascimento Cavalcante, for the support, advice and affection during this journey.

To my colleagues at Sebra Agrícola, especially Rodrigo Lorenzoni, for their long-standing friendship.

To the Brazilian people, for financing public education and research institutions.

To all the professors and staff at UFV who contributed to this achievement, especially the employees of the Fruit Sector, for always being available when I needed them during my stay in Viçosa. Finally, to all of you and other friends who were not mentioned in this short text due to a memory lapse, thank you very much!

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

“Persistence is the shortest path to success.”
(Charles Chaplin)

ABSTRACT

JUNIOR, Edilson Marques, D.Sc., Universidade Federal de Viçosa, November, 2024. **Genetic parameters estimation and QTL mapping in *Sesamum indicum* L.**. Adviser: Carlos Eduardo Magalhaes dos Santos. Co-advisers: Guilherme da Silva Pereira and Kaio Olimpio das Gracas Dias.

Brazil has experienced an exponential increase in sesame production in recent years. Still, farmers face many challenges in cropping sesame, such as the low number of available cultivars adapted to mechanized harvesting. In Brazil, few advances have been made in crop breeding. There is still a lack of germplasm characterization, and genetic breeding work is being developed to increase productivity and adapt to mechanized harvesting. The first chapter of this work presents a study of estimates of sesame genetic parameters in different environments and the characterization of the diversity available. A total of 28 sesame genotypes were phenotypically evaluated in three environments and genotyped with microsatellite and SNP markers. After statistical analysis, significant genetic variability was observed in the genotypes evaluated, which can be explored in crossings in the breeding program. Most of the phenotypic traits evaluated were highly influenced by the environment, presenting genotype x environment interaction and highlighting the complexity of improvement for this crop. In the second chapter, a QTL mapping study was carried out using SNP markers to identify genomic regions associated with traits of interest in sesame breeding. Two F2 populations (SD and SJ) were developed by hybridization, consisting of 98 and 160 genotypes, respectively. Both populations were genotyped by SNP markers and phenotyped in a greenhouse. After statistical analyses, QTL markers were found for 15 of the 17 phenotypic traits evaluated. QTL markers associated with important agronomic traits in sesame were found, such as productivity-related traits, placental adherence, and growth habit.

Keywords: Sesame breeding; Selection index; Associative mapping; GxE interaction

RESUMO

JUNIOR, Edilson Marques, D.Sc., Universidade Federal de Viçosa, novembro de 2024. **Estimativas de parâmetros genéticos em mapeamento de QTL em *Sesamum indicum* L.**. Orientador: Carlos Eduardo Magalhaes dos Santos. Coorientadores: Guilherme da Silva Pereira e Kaio Olimpio das Gracas Dias.

O Brasil tem registrado nos últimos anos aumento exponencial na produção de gergelim, mas muitos desafios são enfrentados pelos produtores, como o baixo número de cultivares disponíveis e adaptados a colheita mecanizada. No Brasil poucos avanços foram conseguidos no melhoramento da cultura. Ainda faltam trabalhos de caracterização de germoplasma, e o desenvolvimento de trabalhos de melhoramento genético voltados para o aumento de produtividade e adaptação a colheita mecanizada. No primeiro capítulo deste trabalho, foi realizado um estudo de estimativas de parâmetros genéticos de gergelim em diferentes ambientes, bem como a caracterização da diversidade presente. Um total de 28 genótipos de gergelim foram avaliados fenotipicamente em três ambientes, e genotipados com marcadores microssatélites e SNPs. Considerável variabilidade genética foi observada entre os genótipos avaliados, que pode ser explorada por cruzamentos no programa de melhoramento. Significativa influência do ambiente foi observado sobre a maioria dos caracteres fenotípicos avaliados, evidenciando a complexidade do melhoramento para esta cultura. No segundo capítulo, foi realizado um estudo de mapeamento de QTLs usando marcadores SNPs para identificação de regiões genômicas associadas a características de interesse no melhoramento do gergelim. Duas populações F2 (SD e SJ) foram desenvolvidas por hibridação, compostas por 98 e 160 genótipos, respectivamente. As duas populações foram genotipadas por marcadores SNPs e fenotipadas em casa de vegetação. Após as análises estatísticas, foram encontrados marcadores QTLs para 15 das 17 características fenotípicas avaliadas. Marcadores QTLs associados a importantes características agronômicas no gergelim foram encontrados, como número de cápsulas por planta, aderência placentária e hábito de crescimento.

Palavras-chave: Melhoramento de gergelim; Indice de seleção; Mapeamento associativo; Interação GxE

SUMMARY

1. GENERAL INTRODUCTION

Sesamum indicum L., popularly known as sesame, belongs to the Pedaliacea family and to the Sesamum genus, the largest representative genus of the family, with 20 described species and some species yet to be cataloged. This species is cultivated worldwide and used for human consumption (Bedigian, 2015). It is a diploid plant with a basic chromosome number equal to 13 chromosomes and 2n = 26 (Bedigian, 2015; Nayar & Mehra, 1970). The plant is autogamous, although cross-pollination can occur from 5% to 68% (Ashri, 2006). Sesame is one of the oldest oilseed crops in the world, known as the king of oilseeds (Bedigian & Harlan, 1986).

Sesame seeds are rich in protein, carbohydrates, unsaturated fatty acids, fiber, antioxidants, and essential minerals, making them very healthy (Abbas et al., 2022). The oil content of the seed ranges from 41% to 62%, making sesame one of the oilseeds with the highest oil content among the best known (Oboulbiga et al., 2023; Uzun et al., 2008a). Protein content ranges from 20-25% and carbohydrates range from 10% to 17.90% (Onsaard, 2012). The unsaturated fatty acids linoleic and oleic are predominant in sesame oil, corresponding to approximately 46.9% and 37.4%, respectively, and are extremely important because they are not synthesized in the human body and must be ingested in the diet. The unsaturated fatty acids oleic and linoleic reduce LDL (bad cholesterol) and increase HDL (good cholesterol). Other fatty acids are present in smaller proportions, such as palmitic acid (8-10%) and stearic acid (2-5%) (Arslan et al., 2007; Uzun et al., 2008b; Wei et al., 2022).

In line with the growing demand for healthy foods, the global demand for sesame has grown significantly in recent years (Islam et al., 2016). Global sesame production is estimated to have grown by 2.6% per year from 1961 to 2022 (Oboulbiga et al., 2023). From 2010 to 2020, global sesame production jumped from 4.3 million tons to almost 7 million tons (FAO, 2022). This growing demand for sesame grain has allowed the expansion of its cultivation to regions and countries where cultivation is not traditionally practiced on a large scale.

Brazil's share of the global sesame market is still small, and this scenario will likely change in subsequent years. Sesame production in Brazil has grown significantly. The area allocated to sesame production in Brazil grew from the 2022/2023 harvest to the 2023/2024 harvest by 82.7%, going from 361.2 thousand to

659.9 thousand hectares planted in Brazil, and a 107.41% increase in production, going from 174.2 to 361.3 thousand total tons in 2024 (CONAB, 2024).

The growth in Brazilian sesame production is related to the growing crop adoption by producers in the second Brazilian harvest, mainly in the central-west region of Brazil. The importance of the second harvest in Brazil, popularly known as safrinha, is undeniable, especially considering that more than 75% of Brazilian corn is produced in safrinha, generally after the soybean harvest (Cordonnier, 2023). However, the climate risks related to second-crop cultivation and the high investment required in corn cultivation (Nóia Júnior & Sentelhas, 2019) cause producers to seek lower-risk alternatives and optimize property yields. In this scenario, the characteristics of low production cost, lower water demand, relatively short cycle, and profit-generating capacity make sesame a viable option for the second harvest (Arriel et al., 2006; Dossa et al., 2023; Zhang et al., 2021). The expansion of sesame cultivation to the central-western region of Brazil offers an opportunity for Brazil to position itself as a player in the global sesame market.

Large-scale sesame production is especially challenging, given the difficulty in establishing an adequate plant stand at sowing, due to the small size of the seeds, extending to the challenges during mechanized harvesting, which can lead to losses of up to 50% (FIORESE et al., 2023; Langham & Wiemers, 2002). Thus, the challenges for the crop include the development of technologies for efficient crop management, as well as the development of cultivars that are better adapted to growing conditions. Plant size, growth habit, dehiscence, and adherence of grains to the capsule after maturation play an important role in the mechanized management of the crop (Beech & Imrie, 2001; Couch et al., 2017). In Brazil, there are a total of 18 registered sesame cultivars (MAPA, 2024). However, only four are commonly cultivated, two of which were developed by the Brazilian Agricultural Research Corporation (EMBRAPA), BRS Seda (Arriel et al., 2009) and BRS Anahí (Arriel et al., 2015), and the imported cultivars, Trebol and K3. Furthermore, due to its interesting agronomic characteristics for Brazilian conditions, the K3 cultivar corresponds to approximately 95% of the sesame cultivated in Brazil (Elaine Cotta, 2024).

The main focus of sesame breeding programs has been the development of cultivars adapted to mechanized harvesting, with high production potential, increased seed retention in capsules during and after maturation (placental adhesion), resistance to pests and diseases, maturation uniformity, sweeter grains, greater number of capsules per bud, and adapted to stresses (Islam et al., 2016). Historically, the main breeding methods adopted for sesame were mass selection, inbred lines, hybridization, and mutation breeding. Although sesame breeding dates back a long time, few genetic advances have been achieved compared to other oilseed species (Teklu et al., 2022). This is due to neglect by the international community, low investment in research into this crop, and the narrow genetic base of sesame in some countries, which hinders the progress of some breeding programs (Ju et al., 2021; Tripathy et al., 2019).

The largest sesame germplasm reserves in the world are in a few countries such as India, South Korea, China, and some African countries such as Sudan and Nigeria, but with insufficient variability for some traits of interest (Ju et al., 2021). The lack of variability for important traits in the crop has led some breeders to adopt mutation induction as a breeding method. Lines with traits of great interest in sesame breeding programs, such as greater production potential, oil quality, capsule indehiscence and determinate growth, have been obtained by this method (Begum & Dasgupta, 2015; Ju et al., 2021; Kouighat et al., 2021; Zhang et al., 2018) and approximately 147 mutants have already been registered worldwide (Teklu et al., 2022). With the advent of new technologies, it is essential to associate conventional methods with new technologies for phenotyping and genomics, aiming to discover new genes and assisted improvement (Teklu et al., 2022; Tripathy et al., 2019). With advances in molecular biology, QTL mapping for several important agronomic traits in sesame has been carried out, such as productivity (Mei et al., 2021; Wu et al., 2014), seed coat color (Wang et al., 2016) and disease resistance (Asekova et al., 2021). In new-generation phenotyping, few results are still found for sesame, but efforts in this direction are already being made and are promising, as demonstrated by Sabag et al. (2024).

Sesame breeding in Brazil still lacks genetic characterization and adoption of technologies to advance breeding programs (Arriel et al., 2007). Thus, the present work sought to characterize a panel of sesame genotypes and perform QTL mapping for agronomic characteristics in the crop, which is important for developing crop improvement research in the country.

1.2 References

Abbas, S., Sharif, M. K., Sibt-e-Abbas, M., Fikre Teferra, T., Sultan, M. T., & Anwar, M. J. (2022). Nutritional and therapeutic potential of sesame seeds. *Journal of Food Quality*, *2022*(1), 6163753.

Arriel, N. H. C., ARAUJO, A. E., SOARES, J. J., BELTRÃO, N. E. M., & FIRMINO, P. T. (2006). Cultivo do gergelim. *Sistemas de Produção, Embrapa Algodão*, *06*.

Arriel, N. H. C., Gondim, T. M. de S., Firmino, P. de T., Beltrão, N. E. de M., Vasconcelos, R. A., Costa, I. L. da, Silveira, N. A. da, Sousa, S. L. de, Dantas, E. S. B., & Pereira, J. R. (2009). *Gergelim BRS Seda* (Issue Figura 2).

Arriel, N. H. C., Mauro, A. O. Di, Arriel, E. F., Unêda-Trevisoli, S. H., Costa, M. M., Bárbaro, I. M., & Muniz, F. R. S. (2007). Genetic divergence in sesame based on morphological and agronomic traits. *Crop Breeding and Applied Biotechnology*, *7*, 253–261.

Arriel, N. H. C., Sousa, S. L. de, Heurt, J., Medeiros, A. A. de, Gondim, T. M. de S., Firmino, P. de T., Vasconcelos, R. A., Costa, I. L. da, & Dantas, E. S. B. (2015). *Gergelim BRS Anahí* (Issue 83).

Arslan, Ç., Uzun, B., Ülger, S., & Ilhan Çağirgan, M. (2007). Determination of oil content and fatty acid composition of sesame mutants suited for intensive management conditions. *JAOCS, Journal of the American Oil Chemists' Society*, *84*(10), 917–920. https://doi.org/10.1007/s11746-007-1125-6

Asekova, S., Oh, E., Kulkarni, K. P., Siddique, M. I., Lee, M. H., Kim, J. I., Lee, J.-D., Kim, M., Oh, K.-W., & Ha, T.-J. (2021). An integrated approach of QTL mapping and genome-wide association analysis identifies candidate genes for phytophthora blight resistance in sesame (Sesamum indicum L.). *Frontiers in Plant Science*, *12*, 604709.

Ashri, A. (2006). Sesame (sesamum indicum L.). *Genetic Resources, Chromosome Engineering, and Crop Improvement: Oilseed Crops*, *4*, 231–289.

Bedigian, D. (2015). Systematics and evolution in Sesamum L.(Pedaliaceae), part 1: evidence regarding the origin of sesame and its closest relatives. *Webbia*, *70*(1), 1–42.

Bedigian, D., & Harlan, J. R. (1986). Evidence for cultivation of sesame in the ancient world. *Economic Botany*, *40*, 137–154.

Beech, D. F., & Imrie, B. C. (2001). *Breeding for mechanised sesame production in Australia*.

Begum, T., & Dasgupta, T. (2015). Amelioration of seed yield, oil content and oil quality through induced mutagenesis in sesame (Sesamum indicum L.). *Bangladesh Journal of Botany*, *44*(1), 15–22. https://doi.org/10.3329/bjb.v44i1.22718

CONAB. (2024, October 10). *Boletim da safra de grãos - 12o Levantamento - Safra 2023/24*. Https://Www.Conab.Gov.Br/Info-Agro/Safras/Graos/Boletim-Da-Safra-de-Graos?Limitstart=0.

Cordonnier, M. (2023, September 7). *Conab aumenta produção de milho do Brasil em 1,9 mt para 131,8 milhões*. Https://Www.Soybeansandcorn.Com/Articles/9715/.

Couch, A., Gloaguen, R. M., Langham, D. R., Hochmuth, G. J., Bennett, J. M., & Rowland, D. L. (2017). Non-dehiscent sesame (Sesamum indicum L.): Its unique production potential and expansion into the southeastern USA. *Journal of Crop Improvement*, *31*(2), 101–172.

Dossa, K. F., Enete, A. A., Miassi, Y. E., & Omotayo, A. O. (2023). Economic analysis of sesame (Sesamum indicum L.) production in Northern Benin. *Frontiers in Sustainable Food Systems*, *6*, 1015122.

Elaine Cotta. (2024, January 8). *Gergelim conquista produtores e cresce mais de 250% nas exportações*. Https://Agfeed.Com.Br/Caminhos-Do-Agro/Gergelim-Conquista-Produtores-e-Cresce-Mais-de-250-Nas-Exportacoes/.

FAO. (2022). *FAO - Food and Agriculture Organization of the United Nations*. https://www.fao.org/faostat/en/#data/QCL

FIORESE, D. A., da SILVA, C. J., ARRIEL, N. H. C., SALVADOR, A. L., CAMARGO, G. Z., SIQUEIRA, M. V, MACHADO, D. A., da SILVA, A. C. A., & PAULA, J. de. (2023). *Colheita mecanizada de gergelim: adaptações com foco na redução das perdas.* Campina Grande: Embrapa Algodão, 2023. https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1153747

Islam, F., Gill, R. A., Ali, B., Farooq, M. A., Xu, L., Najeeb, U., & Zhou, W. (2016). Sesame. In *Breeding Oilseed Crops for Sustainable Production* (pp. 135–147). Elsevier.

Ju, M., Miao, H., Wang, H., & Zhang, H. (2021). Mutagenesis for Creation of Genetic Variability in Sesame. In *The Sesame Genome* (pp. 121–129). Springer.

Kouighat, M., Channaoui, S., Labhilili, M., el Fechtali, M., & Nabloussi, A. (2021). Novel genetic variability in sesame induced via ethyl methane sulfonate. *Journal of Crop Improvement*, *35*(5), 654–665.

Langham, D. R., & Wiemers, T. (2002). *Progress in mechanizing sesame in the US through breeding.*

MAPA. (2024, October 10). *Cultivares ou espécies registradas*. Https://Sistemas.Agricultura.Gov.Br/Snpc/Cultivarweb/Cultivares_registradas.Php.

Mei, H., Liu, Y., Cui, C., Hu, C., Xie, F., Zheng, L., Du, Z., Wu, K., Jiang, X., & Zheng, Y. (2021). QTL mapping of yield-related traits in sesame. *Molecular Breeding*, *41*(7), 1–14.

Nayar, N. M., & Mehra, K. L. (1970). Sesame: Its uses, botany, cytogenetics, and origin. *Economic Botany*, 20–31.

Nóia Júnior, R. de S., & Sentelhas, P. C. (2019). Soybean-maize off-season double crop system in Brazil as affected by El Niño Southern Oscillation phases. *Agricultural Systems*, *173*, 254–267.

Oboulbiga, E. B., Douamba, Z., Compaoré-Sérémé, D., Semporé, J. N., Dabo, R., Semde, Z., Tapsoba, F. W.-B., Hama-Ba, F., Songré-Ouattara, L. T., & Parkouda, C. (2023). Physicochemical, potential nutritional, antioxidant and health properties of sesame seed oil: a review. *Frontiers in Nutrition*, *10*, 1127926.

Onsaard, E. (2012). Sesame proteins. *International Food Research Journal*, *19*(4).

Sabag, I., Bi, Y., Sahoo, M. M., Herrmann, I., Morota, G., & Peleg, Z. (2024). Leveraging genomics and temporal high-throughput phenotyping to enhance association mapping and yield prediction in sesame. *The Plant Genome*, e20481.

Teklu, D. H., Shimelis, H., & Abady, S. (2022). *Genetic Improvement in Sesame (Sesamum indicum L.): Progress and Outlook: A Review. Agronomy 2022, 12, 2144* (p. 2144). s Note: MDPI stays neu-tral with regard to jurisdictional claims in ….

Tripathy, S. K., Kar, J., & Sahu, D. (2019). Advances in sesame (Sesamum indicum L.) breeding. *Advances in Plant Breeding Strategies: Industrial and Food Crops: Volume 6*, 577–635.

Uzun, B., Arslan, Ç., & Furat, Ş. (2008a). Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (Sesamum indicum L.). *Journal of the American Oil Chemists' Society*, *85*(12), 1135–1142.

Uzun, B., Arslan, Ç., & Furat, Ş. (2008b). Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (Sesamum indicum L.). *Journal of the American Oil Chemists' Society*, *85*(12), 1135–1142.

Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., Ni, X., Gao, Y., Xiang, H., & Wei, X. (2016). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genomics*, *17*, 1–13.

Wei, P., Zhao, F., Wang, Z., Wang, Q., Chai, X., Hou, G., & Meng, Q. (2022). Sesame (Sesamum indicum L.): A Comprehensive Review of Nutritional Value, Phytochemical Composition, Health Benefits, Development of Food, and Industrial Applications. *Nutrients*, *14*(19), 4079.

Wu, K., Liu, H., Yang, M., Tao, Y., Ma, H., Wu, W., Zuo, Y., & Zhao, Y. (2014). High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (Sesamum indicum L.) based on RAD-Seq techonology. *BMC Plant Biology*, *14*, 1–14.

Zhang, H., Langham, D. R., & Miao, H. (2021). Economic and academic importance of sesame. *The Sesame Genome*, 1–18.

Zhang, H., Miao, H., Wei, L., Li, C., Duan, Y., Xu, F., Qu, W., Zhao, R., Ju, M., & Chang, S. (2018). Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. *BMC Plant Biology*, *18*(1), 1–12.

2. CHAPTER I - GENETIC PARAMETERS IN SESAME ACCESSIONS ESTIMATED BY DNA MOLECULAR MARKERS AND PREDICTED GENETIC VALUES

Abstract: This study used data from 64 microsatellite markers and 5895 SNP markers to study the genetic structure and diversity of 28 sesame genotypes. The genetic structure of the genotypes was evaluated using the Bayesian approach on the Structure software, as well as principal component analysis. The 28 sesame genotypes were also evaluated in three environments, and the genetic parameters were estimated. It was possible to observe significant genetic divergence elucidated by the analyses performed. The average observed heterozygosity (Ho) was 0.36, while the expected (He) was 0.38, with an average of 1.42 alleles per SSR locus. All analyses indicated the consistent formation of two main groups among the genotypes studied. Differences in the detection of genetic structure were observed between the Bayesian approach and PCA using SNP data. On the other hand, high concordance was observed between the two methodologies using the SSR data. The mixed model methodology was used to estimate the genetic parameters among the genotypes. For six of the 14 traits, a positive effect of the insertion of the genomic relationship matrix obtained by SNPs in estimating the genetic parameters was observed. A significant genotype x environment interaction (GxE) was observed for nine of the 14 phenotypic traits evaluated, and genetic variance for 10 traits, with heritability values ranging from 0.15 to 0.66. The results obtained in this study show a narrow genetic base for the genotypes studied but with considerable genetic variability to be used in the breeding program. From the FAI-Blup selection index, it was possible to classify the genotypes with greater stability and adaptability in the environments studied. The genotypes YGUU1, IGUU1, KGUU1, and BGUU1 were the most stable and adapted based on the results obtained, which is important for future crossings in the program.

Keywords: Sesame. Molecular markers. Breeding values. Genotype x environment interaction

Resumo: No presente trabalho, dados de 64 marcadores microssatélites e 5895 marcadores SNPs foram utilizados para estudo da estrutura e diversidade genética de 28 genótipos de gergelim. A estrutura genética dos genótipos foi avaliada por meio da abordagem bayesiana, utilizando o software Structure e por meio de análise de componentes principais. Os 28 genótipos de gergelim foram avaliados em três ambientes e os parâmetros os genéticos foram estimados para 14 características fenotípicas. Foi possível observar significativa divergência genética entre os genótipos. A heterozigosidade observada média foi de 0,36, enquanto a esperada (He) foi de 0,38, com média de 1,42 alelos por loco SSR. Todas as análises indicaram a formação consistente de dois principais grupos entres os genótipos estudados. Diferenças na detecção da estrutura genética foi observada entre a abordagem bayesiana e PCA, utilizando os dados SNPs. Por outro lado, alta concordância foi observada entre as duas metodologias utilizando os dados SSR. Para seis, das 14 características, observou-se efeito positivo da inserção da matriz de relacionamento genômico obtida por SNPs na estimativa dos parâmetros genéticos. Foi observado interação genótipo x ambiente (GxE) significativo para nove das 14 características fenotípicas avaliadas, e variância genética para 10 características, com valores de herdabilidade variando de 0,15 a 0,66. Os resultados obtidos no presente trabalho evidenciam uma base genética estreita para os genótipos estudados, mas com considerável variabilidade genética a ser usada no programa de melhoramento. A partir do índice de Seleção FAI-Blup, os genótipos YGUU1, IGUU1, KGUU1 e BGUU1 foram os genótipos com maior estabilidade e adaptabilidade nos ambientes estudados, importantes para serem utilizados em cruzamentos futuros do programa.

Palavras-chave: Gergelim. Marcadores moleculares. Valor genético. Estrutura genética. Interação genótipo x ambiente.

2.1. Introduction

Sesame (Sesamum indicum L) is one of the oldest oilseeds cultivated in the world. It is considered the “king” of oilseeds due to the high content and quality of oil, proteins, and minerals in the grains (Meena et al. 2018; Patil and Lokesha 2018). It consists of an annual plant belonging to the Pedaliaceae family, widely cultivated in tropical and subtropical regions of Africa, Asia, and South America, being the most important species among the 23 species that make up the genus Sesamum (Zhang et al. 2019). The high caloric value, richness of nutrients in the grains, and their nutraceutical and pharmaceutical properties make sesame seeds a “superfood” used in various food dishes (Langyan et al., 2022; Namiki, 2007), in addition to the production of derivatives, such as soaps, lubricants, drugs and cosmetics (Zhang et al. 2021).

The oil content of the seed ranges from 41% to 62%, making sesame one of the oilseeds with the highest oil content among the best known (Uzun et al. 2008a; Oboulbiga et al. 2023). Lignans (sesamin and sesamol), antioxidants abundant in the oil, have important benefits for human health, as they are effective in neutralizing free radicals and protecting cells, in addition to having anti-inflammatory effects, reducing chronic inflammation that can trigger chronic diseases, such as cancer and heart disease. In addition to these benefits, sesame is known for regulating glycemic indexes, which is important in controlling diabetes and hypertension (Abbas et al. 2022; Wei et al. 2022a). Sankar et al. (2006) reported significant effects of sesame oil consumption, which were observed in the reduction of blood pressure, increase of vitamins C and E, ß-carotene and reduced glutathione, decrease of lipid peroxidation and increase of antioxidant levels in hypertensive patients. Antinutritional effects for human health are few reported for sesame, and oxalic and phytic acids would be the main antinutrients reported, with concentrations of 13% and 1.12% of these acids in the sesame husk, respectively (Farran et al. 2000; Wei et al. 2022b).

The interest of producers in sesame cultivation is due to the interesting agronomic characteristics of the species, such as resistance to nematodes, tolerance to drought, and suitability for rotation with main crops since it tends to have a short cycle in relation to other optional crops and a relatively low production cost (Couch et al. 2017). Brazilian sesame production has increased significantly due to the expansion of cultivation to new regions. However, many challenges have been faced in these new cultivation frontiers since large-scale cultivation requires adapted cultivars, especially mechanization. Currently, the cultivars available on the Brazilian market do not meet these needs. Loss during harvest is one of the biggest obstacles, and a problem in most producing countries, which do not yet have indehiscent cultivars. To date, most of the sesame worldwide is dehiscent.

The development of cultivars with greater production potential, early maturation, disease resistance, higher grain quality, and indehiscent and that allow mechanization has been the focus of most sesame breeding programs (Georgiev et al. 2008; Stamatav et al. 2020). Studies have shown that a major challenge is to combine such characteristics in the same genotype, especially for adaptability to mechanization, where characteristics favorable to mechanized management are negatively correlated with production despite advances already having been achieved (Georgiev et al. 2008; Stamatav et al. 2020). New research is needed in Brazil, especially aimed at developing cultivars adapted to new growing regions. Major contributions to the genetic improvement of the crop have been made by Embrapa, which has been conducting research in Brazil for many years aimed at the genetic improvement of the crop. For many years, research has sought to develop sesame cultivars for manual harvesting, practiced on small properties in the Northeast, the predominant system for sesame cultivation in this region (Arriel et al. 2006). However, research to develop cultivars that are better adapted to the new Brazilian climate and cultivation conditions needs to be carried out. Teklu et al. (2022a) reinforce that the increases in sesame production recorded in the world emanate from the expansion of cultivation areas rather than greater yield per area.

Knowledge of genetic structure and diversity and the relationships between genotypes of a population is essential for establishing a breeding program since a high degree of genetic divergence between parents is essential for hybridization. In addition, it allows the breeder to define more assertive strategies, such as avoiding some pre-breeding activities in developing new materials (Rufo et al. 2019). Estimates of genetic diversity can be obtained using several types of genetic markers, such as morphological, biochemical, and molecular markers. Molecular markers play an essential role in breeding programs, from identifying genes responsible for traits of interest to managing backcrossing programs (Schulman 2007). In the study of diversity, these markers provide broad coverage of the population's genetic structure, expressing more realistic genetic variations between genotypes, given the possibility of using many markers and representing neutral traits of simple inheritance (Sensoy et al. 2007; EL-Mansy et al. 2021). Among the various molecular markers, microsatellites (SSR) stand out in studies of genetic diversity due to their characteristics of high reproducibility, codominant nature, high degree of polymorphism, and multiallelism (Nandakumar et al. 2004; Caixeta et al. 2013). On the other hand, SNP markers are more distributed in the genome, providing greater coverage, and are highly efficient in diversity and genetic association studies (Hoffmann et al. 2011).

Reed & Frankham (2001) emphasize that despite the advantages of molecular markers, they may not provide estimates of genetic divergence equivalent to experimental tests for quantitative traits, as they do not provide additive genetic variations, selection differentials, different mutation rates, low statistical power and do not consider environmental effects. In general, molecular and morphological information provide important information for decision-making in plant breeding and should be used together (El-Aziz et al. 2016; EL-Mansy et al. 2021). Thus, the objective of this work was to estimate genetic parameters of 28 sesame genotypes based on morphological and molecular markers and to identify groups of divergent genotypes that can be used in future crosses

2.2. Material and methods

2.2.1. Genetic material and phenotypic characterization

This study was carried out through a partnership between the Universidade Federal of Viçosa (UFV) and the company Sebra Agrícola S.A., located in the municipality of Canarana, state of Mato Grosso (Latitude: 11° 41' 6'' South, Longitude: 41° 46' 10'' West; Altitude: 200 meters above mean sea level). Twenty-eight sesame genotypes of the Germplasm Bank of Sebra Agrícola SA were morphologically characterized to estimate genetic diversity. An experiment was conducted with 20 genotypes in a randomized block design with four replications in the experimental area of Sebra Agrícola SA in 2022. Two more experiments in randomized blocks with three replicates were conducted in the experimental area of Sebra Agrícola S.A in Canarana-MT and Rosário Oeste-MT in 2023, with 28 genotypes, corresponding to the 20 genotypes analyzed in the previous experiment, added to 8 new genotypes. Of the 28 genotypes, 27 were characterized using microsatellite markers, and 21 were genotyped using SNP markers by the DartSeq methodology. The code of each genotype, the type of marker, and the locations where each genotype was evaluated are described in Table 1.1.

Table 1.1. List of genotypes evaluated in the present work, the type of marker in which each genotype was genotyped, and the environments where each genotype was evaluated.

Code number	Genotype	Marker genotyping	Evaluated environment
1	GGUU1	SSR and SNP	1, 2 and 3
2	BGUU1	SSR and SNP	1, 2 and 3
3	HGUU1	SSR and SNP	1, 2 and 3
4	NGUU1	SSR and SNP	1, 2 and 3
5	YGUU1	SSR and SNP	1, 2 and 3
6	WGUU1	SSR and SNP	1, 2 and 3
7	RGUU1	SSR and SNP	1, 2 and 3
8	IGUU1	SSR and SNP	1, 2 and 3
9	CGUU1	SSR and SNP	1, 2 and 3
10	FGUU1	SSR and SNP	1, 2 and 3
11	EGUU1	SSR and SNP	1, 2 and 3
12	SGUU1	SSR and SNP	1, 2 and 3
13	AGUU1	SSR and SNP	1, 2 and 3
14	KGUU1	SSR and SNP	1, 2 and 3
15	TGUU1	SSR and SNP	1, 2 and 3
16	DMDU1	SSR and SNP	1, 2 and 3
17	JMDU1	SSR and SNP	1, 2 and 3
18	PGUU1	SSR and SNP	1, 2 and 3
19	QGUU1	SSR and SNP	1, 2 and 3
20	ZGUU1	SSR and SNP	1, 2 and 3
21	VGUU1	SNP	2 and 3
22	L1	SSR	2 and 3
23	L2	SSR	2 and 3
24	L3	SSR	2 and 3
25	L4	SSR	2 and 3
26	L5	SSR	2 and 3
27	L6	SSR	2 and 3
28	L7	SSR	2 and 3

2.2.2. Molecular characterization and statistical analysis with SSR markers

Fully expanded young leaves of each genotype were used for DNA extraction from the maceration of 200 mg of plant tissue together with 700 µL extraction buffer and 4 µL of 2-mercaptoethanol, following the protocol of Doyle & Doyle (1990). The microplate spectrophotometer (MultiskanTM GO) was used to quantify the concentration and quality of the DNA at wavelengths 230, 260, and 280 nm. An initial set of 125 SSR markers was used for the molecular characterization of the 27 genotypes. A reagent mix was prepared at the following concentrations for polymerase chain reaction (PCR): 50 ng of DNA; 1X buffer (500 mM KCl; 100 mM Tris-HCl pH 8.4; 1% Triton X-100); 1.5 mM MgCl2; 0.2 µM of each dNTP; 0.3 µM of each primer (forward and reverse); 1 U of Taq DNA polymerase Promega. The following conditions were set for amplification of the fragments using a Veriti® thermocycler: denaturation at 94 °C for 2 min, followed by 10 cycles with denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 30 s. Then, another 30 cycles with denaturation at 94 °C for 30 s, annealing at 45 °C for 30 s and extension at 72 °C for 30 s, followed by a final extension cycle at 72 °C for 8 min. The amplified fragments were separated by electrophoresis into 10% polyacrylamide gel and stained with silver nitrate.

From the SSR data, allele frequency, observed (Ho) and expected (He) heterozygosity, and observed and expected polymorphic information content (PIC) for the markers were estimated. The genetic structure of the population was assessed through principal component analysis. The PCA results were plotted using the Factoextra R package (Kassambara and Mundt 2017) in R software (Team 2025) and through a Bayesian approach in the Structure software (Pritchard et al. 2000). The analysis was performed considering the program's default parameters, without a priori information on the number of groups, with the number of groups tested (K) varying from 1 to 6. Twenty-five thousand runs were used for burn-in, followed by 250,000 MCMS (Monte Carlo Markov Chain) repetitions. Twenty iterations were used for each group. Evanno et al. (2005) criterion was used to select the best value of K.

Genetic divergence between genotypes was estimated using a distance matrix generated from the complement of Nei's similarity index in the Poppr package (Kamvar et al. 2014) of R software (Team 2025). Hierarchical clustering among the genotypes was performed using the unweighted pair group method with arithmetic mean

(UPGMA) and plotted using the Factoextra package (Kassambara and Mundt 2017) of R software (Team 2019).

2.2.3. Molecular characterization with SNP markers

The DArT-seq genotyping methodology (DArT Pty Ltd, Canberra, Australia) was used to obtain the SNPs. Two leaf discs of each genotype were collected from young leaves of the 21 genotypes, as described in Table 1.1, dried in ovens at 35 °C for 36 hours, stored in 2 ml tubes, and sent for genotyping, following the company's instructions. The raw data obtained were aligned to the sesame reference genome (Wang et al. 2022) using the Bowtie2 program (Langmead and Salzberg 2012). Variant calling was performed using the FreeBayes program (Garrison and Marth 2012). Biallelic variants with high quality were selected. SNPs were filtered to a minimum call rate of 0.8 and a minimum allele frequency of 5%.

Like the analyses performed with SSR data, diversity indices such as PIC, observed and expected heterozygosity were estimated from the SNPs obtained. The genetic structure among individuals was also evaluated through principal component analysis in the R software (Team 2025), and the Bayesian approach in the Structure software (Pritchard et al. 2000). The parameters used were the same as those previously defined for SSRs. The relationship matrix between genotypes was calculated using the VanRaden method (VanRaden 2008) using the AGHmatrix package in R (Amadeu et al. 2016).

2.2.4. Morphological characterization and statistical analysis

Data was collected on the vegetative, reproductive, and post-harvest phases of the plants. Table 1.2 lists the variables evaluated, the abbreviations used, and the objective for each characteristic to be used in constructing the ideotype for the selection index.

Table 1.2. List of characteristics used in the phenotypic characterization of genotypes in different environments and the program's objectives for such characteristics.

Trait	Code	Program goal
Plant height	PH	Minimize
Height at first capsule	HFC	Minimize
Internode length	IL	Minimize
Primary branch number	PBN	Maximize
Secondary branch number	SBN	Maximize
Number of capsules per plant	CBP	Maximize
Internode number to the first capsule	INFC	Minimize
Seeds per capsule	SBC	Maximize
Capsule length	CL	Maximize
Capsule width	CW	Maximize
Capsule thickness	CT	Maximize
Peeling yield	PY	Maximize
Thousand-grain weight	PMS	Maximize
Yield	Prod	Maximize

The genetic variance components, the genotype x environment interaction, and the genetic values were estimated using the maximum likelihood methodology (REML/BLUP). The significance of the components was verified by the likelihood ratio test (LRT) at 5% probability. The mixed linear models methodology was used to obtain the BLUP estimates of the genetic values, following the general equation of $y_1 = X_i\beta + Z_i b_i + \varepsilon$, where Yi represents the response vector of the observations; $X_i\beta$ are the fixed effects. X is the incidence matrix for the fixed effects, and β is the vector of the parameters of the fixed effects; $Z_i b_i$ are the random effects, where Z is the parameter vector of the random effects, and b is the parameter vector of the random effects, with $b \sim N(0, G)$, where ε is the covariance matrix of the random effects and represents the vector of residuals, assumed to be random. Each trait was analyzed individually.

The genetic relationship matrix obtained from the SNP markers was inserted into the mixed models to evaluate its effect on the estimates of the genetic parameters. The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values evaluated the efficiency of inserting the genomic matrix into the models. The model used without genetic information wasi:

$$Y_{iJk} = \mu + E_j + G_i + GE_{ij} + B_k(E_j) + \epsilon_{ijk}$$

Where,

Y_{iJk}: represents the vector of phenotypic values;

μ: represents the general average;

G_i: random effect of genotype i, where $G_i \sim N(0, \sigma_G^2)$;

E_j: environment fixed effect$j$;

GE_{ij}: genotype by environment interaction, where $GE \sim N(0, \sigma_{GE}^2)$;

$B_k(E_j)$: fixed effect of k-block, nested in ambient j;

ϵ_{ijk}: residual error, where$E_{ijk} \sim N(0, \sigma_{GE}^2)$

To test the effectiveness of molecular information in improving estimates of genetic parameters, a kinship effect was inserted into the model:

$$Y_{iJk} = \mu + E_j + g_i + GE_{ij} + B_k(E_j) + \epsilon_{ijk},$$

Where,

g_i: Genetic effect of genotype i, with $g_i \sim N\left(0, G\sigma_g^2\right)$, where G is the kinship matrix.

The genetic value of each genotype was obtained from the sum of the genotypic effect of each genotype with the overall mean of the experiment. The variance components estimated via REML were phenotypic variance (σ_f^2), genotype x environment interaction variance (σ_{GxE}^2) and heritability (H^2). Mixed model analyses were performed in R software (Team 2019) using the Sommer package (Covarrubias-Pazaran 2016).

For traits with significant GxE interaction, the adjusted blues were obtained by adding the overall mean of the experiment to the blues of the GxE interaction. For subsequent analyses, the adjusted blues of the GxE interaction of the different traits in each environment were considered as an independent variable to study the influence of the environment on the traits.

A Pearson correlation analysis was performed to study the relationship between the variables. Then, a factor analysis with orthogonal rotation using the Varimax method was performed from the adjusted BLUPs, seeking to maximize the relationship between the phenotypic variables and interpret how they group and associate. Finally, seeking to identify the most promising genotypes among the genotypes evaluated, the FAI-BLUP (Harmonic Mean of Relative Performance of Breeding Values - Best Linear Unbiased Prediction) selection index (De Resende 2007) was used to order the genotypes, considering the multiple environments and the stability of the genotypes in relation to the genotype x environment interaction (GxE) in the R package, Metan (Olivoto and Lúcio 2020). The ideotype for selection was based on the objectives of each characteristic evaluated, described in Table 1.2. In addition, the adjusted mean of the genotypes was estimated, considering genotypes as fixed to compute an Euclidean Distance matrix, from which a hierarchical grouping was performed using the intergroup average linkage method (UPGMA).

2.3. Results

2.3.1. Genetic diversity and structure based on molecular information

A total of 5895 SNP markers were obtained for genotype characterization. Of the 125 SSR markers tested, 64 (supplementary material - Table S1) were polymorphic and could be used for genotype characterization. Ninety-one different alleles were found, and the number of alleles per locus ranged from 2 to 6, with an average of 1.42 alleles per locus. The average observed heterozygosity (Ho) was 0.36, ranging from 0 to 1, and the expected heterozygosity (He) ranged from 0.06 to 0.6, with an average of 0.38. The PIC values of the markers ranged from 0.06 to 0.51, with an average of 0.31. The expected PIC values ranged from 0.38 to 0.7.

The average genetic distance between genotypes was 0.93. The maximum distance observed was 1.59 between the pairs of genotypes QGUU1 and L2, and the minimum distance was 0.37 between the lines L2 and L7. The formation of three groups among the studied genotypes was observed, with two main groups with 12 and 11 genotypes in each, respectively. The third group comprised only four genotypes (CGUU1, IGUU1, DMDU1, and WGUU1) (Figure 1.1). The groups obtained show that of the four commercial genotypes widely cultivated in Brazil, three were grouped in the same group, being K3, BRS Anahí and Trebol (KGUU1, AGUU1, TGUU, respectively), and only the genotype BRS Seda (SGUU1) was grouped separately.

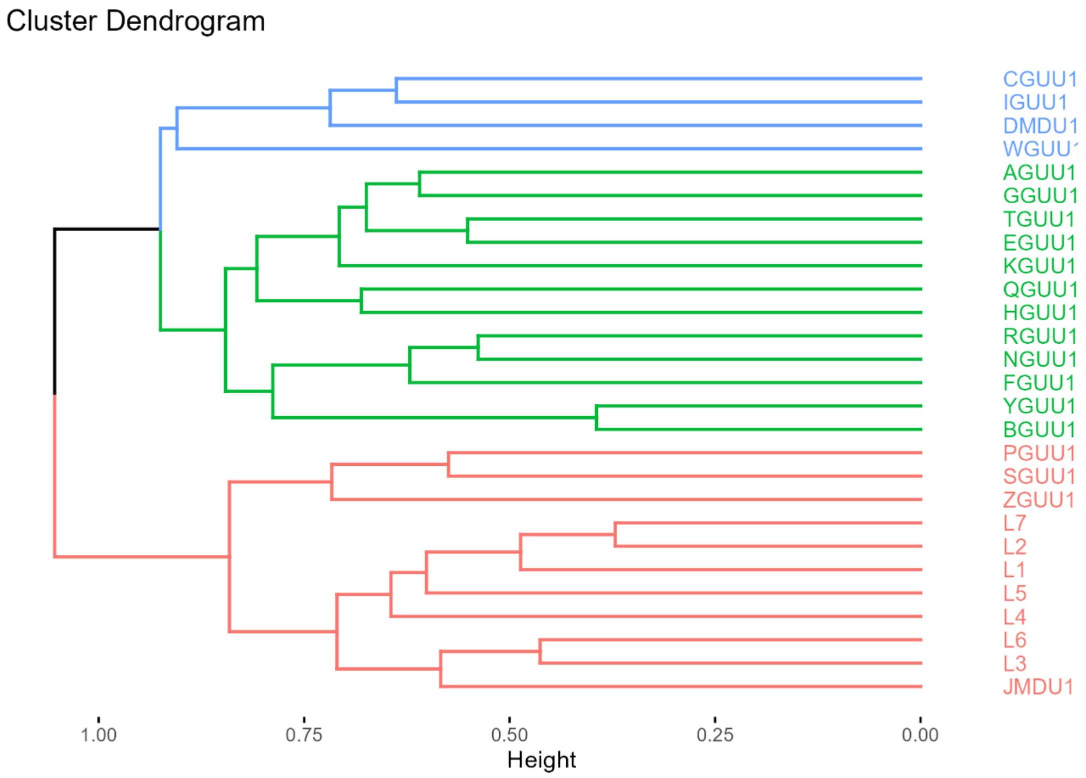

Figure 1.1. Hierarchical clustering of the 27 sesame genotypes obtained by the complete linkage method, constructed from the matrix of genetic distances of the SSR markers obtained by complementing the weighted nei index.

Figure 1.2 illustrates the genetic structure of the studied genotypes in the Structure software (Bayesian approach). For both markers, the greatest K (ΔK) variation was observed from 2 to 3, indicating K = 2, suggesting that the genetic structure of the studied genotypes is better described when divided into two main groups. It is possible to observe that a better genetic structure is observed with the SSR data, with a clearer distinction of the groups (Figure 2a). In fact, Figure 2b shows that some individuals clearly belong to one of the two groups (100% in red or 100% in green). At the same time, several genotypes have a mixed composition, suggesting gene flows or the sharing of a common ancestor between some genotypes of the groups, which is common in genetic breeding programs, where crossings and selection can lead to the mixing of lineages of different origins.

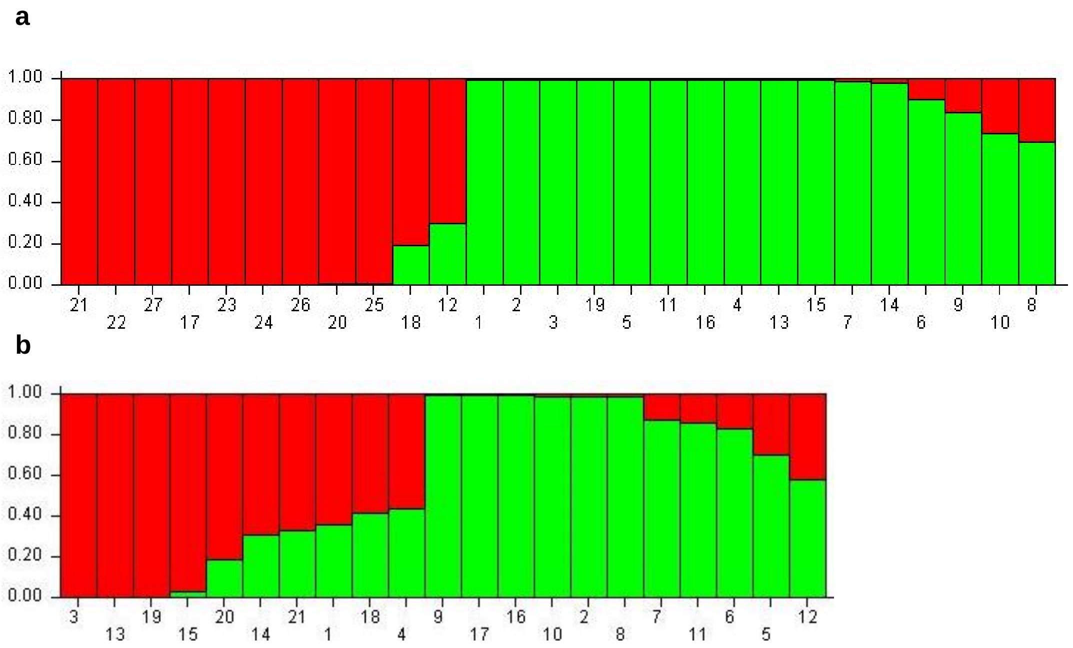

Figure 1.2. Analysis of the genetic structure and barplot of the genetic profile of the genotypes obtained by molecular markers SSR (a) and SNPs (b) distributed into two groups based on the hypothetical populations obtained. The genotypes are distributed on the horizontal and vertical axes; the bars represent the probability of each genotype belonging to each group, ranging from 0 to 1. Individuals: 1: GGUU1, 2: BGUU1, 3: HGUU1, 4: NGUU1, 5: YGUU1, 6: WGUU1, 7: RGUU1, 8: IGUU1, 9: CGUU1, 10: FGUU1, 11: EGUU1, 12: SGUU1, 13: AGUU1, 14: KGUU1, 15: TGUU1, 16: DMDU1, 17: JMDU1, 18: PGUU1, 19: QGUU1, 20: ZGUU1, 21: VGUU1, 22: L1, 23: L2, 24: L3, 25: L4, 26: L5, 27: L6, 28: L7.

Like the results observed in the Bayesian approach, two distinct groups are observed in the principal component analyses performed on the SSR data (Figure 1.3a), with high concordance between the groups formed. However, with the SNP markers, the groups formed in the Bayesian approach (Figure 1.2b) were completely different from those observed in the PCA analysis (Figure 1.3b). Differences in the distribution of genotypes can be seen in the graphs of the two approaches, which is expected since the nature of the markers is different, in addition to the number of genotypes analyzed with microsatellite markers being greater.

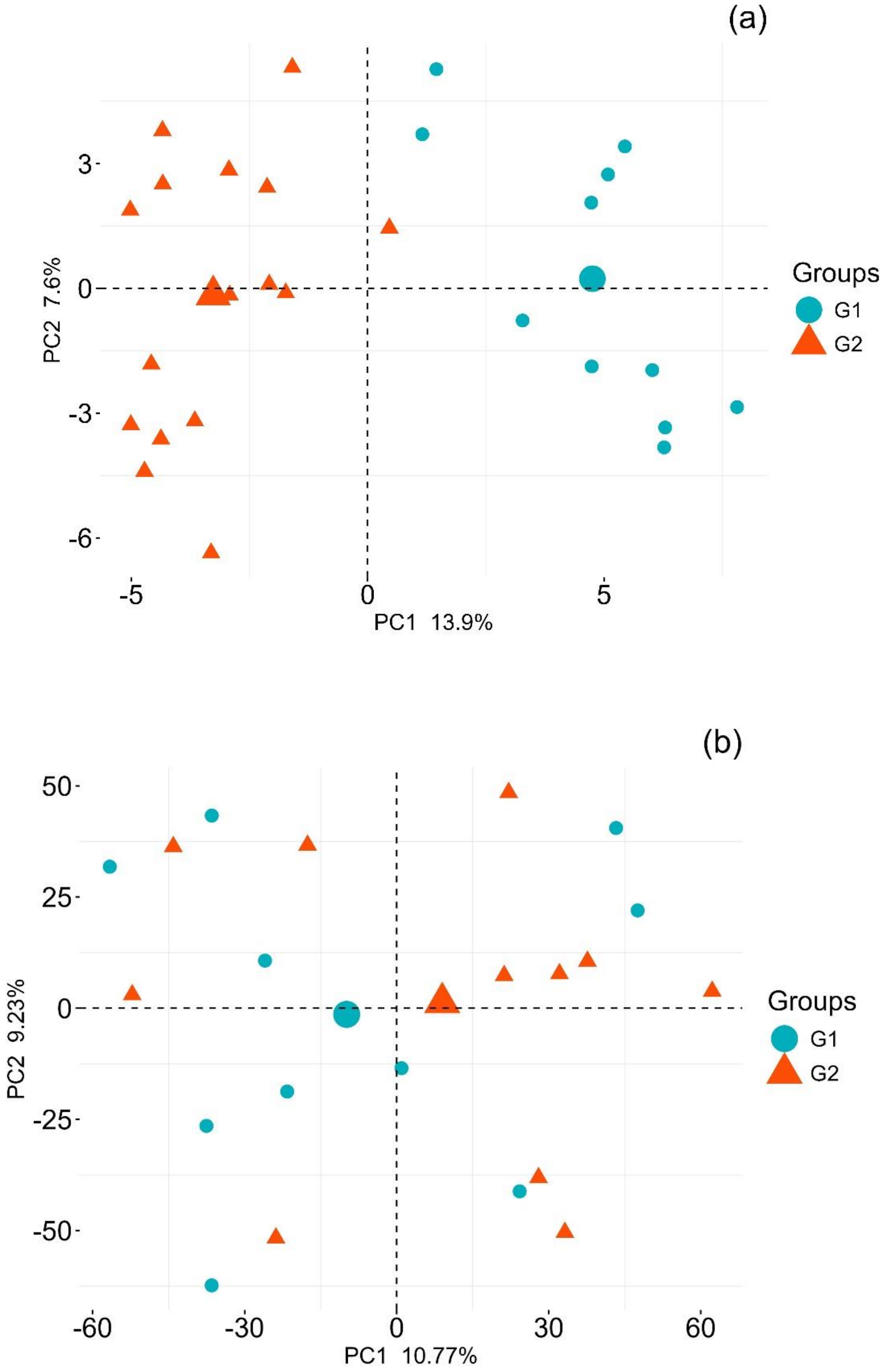

Figure 1.3. Dispersion by principal components obtained from SSR (a) and SNPs (b) data showing the distribution of genotype groups, colored according to the hypothetical groups obtained in the Bayesian approach.

Despite the discrepancy in the results of structuring the genotypes using SNPs, it was observed that there was a tendency for the genotypes to be divided into two groups in all analyses performed, which highlights the consistency of the results. The distribution of the genotypes in hierarchical clustering was studied in relation to the dispersion in the PCA analyses (graphics not shown). High agreement was observed between the hierarchical clustering of the genotypes and the PCA obtained with the markers, where the distribution of the genotypes in the hierarchical clusters followed the same pattern in the PCA analysis with the SSR markers. In the PCA obtained from the SNPs, the distribution pattern was inconsistent between this and the dendrogram obtained.

2.3.2. Morphological analysis

When evaluating the efficiency of inserting genetic similarity matrices in the estimation of genetic parameters for the 14 characteristics evaluated, it is observed that the insertion of genetic information is useful for six traits, improving the quality of the model (Table 1.3).

Table 1.3. Comparative table of AIC and BIC values for models without (M1) and with the insertion of the genomic matrix (M2) in the models for estimating genetic parameters.

	AIC		**BIC**	
Traits	**M1**	**M2**	**M1**	**M2**
Primary branch number	115.0462	112.822	149.3396	147.1156
Secondary branch number	177.63352	176.114	211.92697	210.40748
Height to the first capsule	-3.42943	1.70676	30.90779	36.04398
Plant height	-26.03331	-12.12797	8.30391	22.20925
Internode length	127.31068	127.31068	161.56018	161.56018
Internode number to the first capsule	88.48406	93.88176	122.73356	128.13126
Capsule by plant	144.75766	149.31222	179.09488	183.64944
Capsule length	160.62157	161.41383	194.87107	195.66333
Capsule width	73.97918	70.32836	108.22868	104.57786
Capsule thickness	55.43506	51.19145	89.68456	85.44095
Seeds by capsule	152.60915	156.17548	184.08409	187.65042
Peeling yield	172.58591	172.62552	204.00254	204.04216
Yield	190.92617	189.54719	223.75821	222.37923
Thousand-grain weight	90.87605	90.23491	124.10615	123.46501

Significant genetic variance was observed for 10 of the 14 traits evaluated, using the likelihood ratio test (LRT) at 5% probability (Table 1.4). The heritability values for these traits ranged from 0.15 to 0.67. The highest heritability values were observed for plant height and height to the first capsule, with heritabilities of 0.67 and 0.36, respectively. The lowest heritability values obtained were for number of capsules per plant and number of seeds per capsule (0.15 and 0.17, respectively).

Table 1.4. Analysis of genetic variance for quantitative traits evaluated for morphological characterization among sesame genotypes.

Trait	**Best model**	**Genetic variance (σg)**	**GxE Variance**	**Residual**	**Heritability**	**LRT (σ Genetica)**	**LRT (σ GE)**
Primary branch number	M2	0.4003	0.2223	0.5943	0.3289	7e-05 ***	0.00036 **
Secondary branch number	M2	0.0631	0.1492	0.8616	0.0588	0.1996	0.03736 *
Height of the first capsule	M1	333.3151	373.9874	209.475	0.3636	0.00194 **	9.24e-17 ***
Plant height	M1	1554.1649	148.1245	632.6167	0.6656	2.51e-12 ***	0.01266 *
Internode length	M1	0	0.1416	0.6824	0	0.9992	0.01723 *
Internode number to the first capsule	M1	0.9922	0.5743	1.9303	0.2837	0.00132 **	0.00608 ***
Capsule by plant	M1	42.6968	2.3967	237.3528	0.1512	0.01662 *	0.9008
Capsule length	M1	0.0239	0.0061	0.0398	0.343	1e-05 ***	0.08162
Capsule width	M2	0.0015	0.0009	0.003	0.2819	0.00017 ***	0.00225 **
Capsule thickness	M2	0.0018	0.0021	0.0018	0.3196	0.00069 ***	3.33e-14 ***
Seeds by capsule	M1	16.6528	0	80.93	0.1707	0.00244 **	0.999
Peeling yield	M1	-0.5975	0	80.309	-0.0075	0.8422	0.8422
Yield	M2	1350688.795	0	32902838.4	0.0394	0.2022	0.9962
thousand-grain weight	M2	0.0492	0.0697	0.0585	0.2772	0.0160*	4.38e-11 ***

* traits with significant genetic variance or genotype x environment interaction by the maximum likelihood ratio test (LRT) at 5% probability. ^{ns} features with non-significant variance by the maximum likelihood ratio test (LRT) at 5% probability

Among the 14 traits evaluated, nine showed significant genotype x environment interaction, evidencing the influence of the environment on trait expression and the differential response of genotypes between environments. The number of secondary branches and internode length variables did not show significant genetic variance but showed significant genotype x environment interaction. Subsequent analyses were performed using only the variables with genetic variance and significant GxE interaction.

The mean blues of the variables that did not present GxE interaction were combined with the adjusted blues of the traits with GxE interaction. The blues for each trait in the different environments were considered an independent variable, and a correlation analysis was performed. A correlation graph demonstrated the intensity of the relationship between the variables (Figure 1.4).

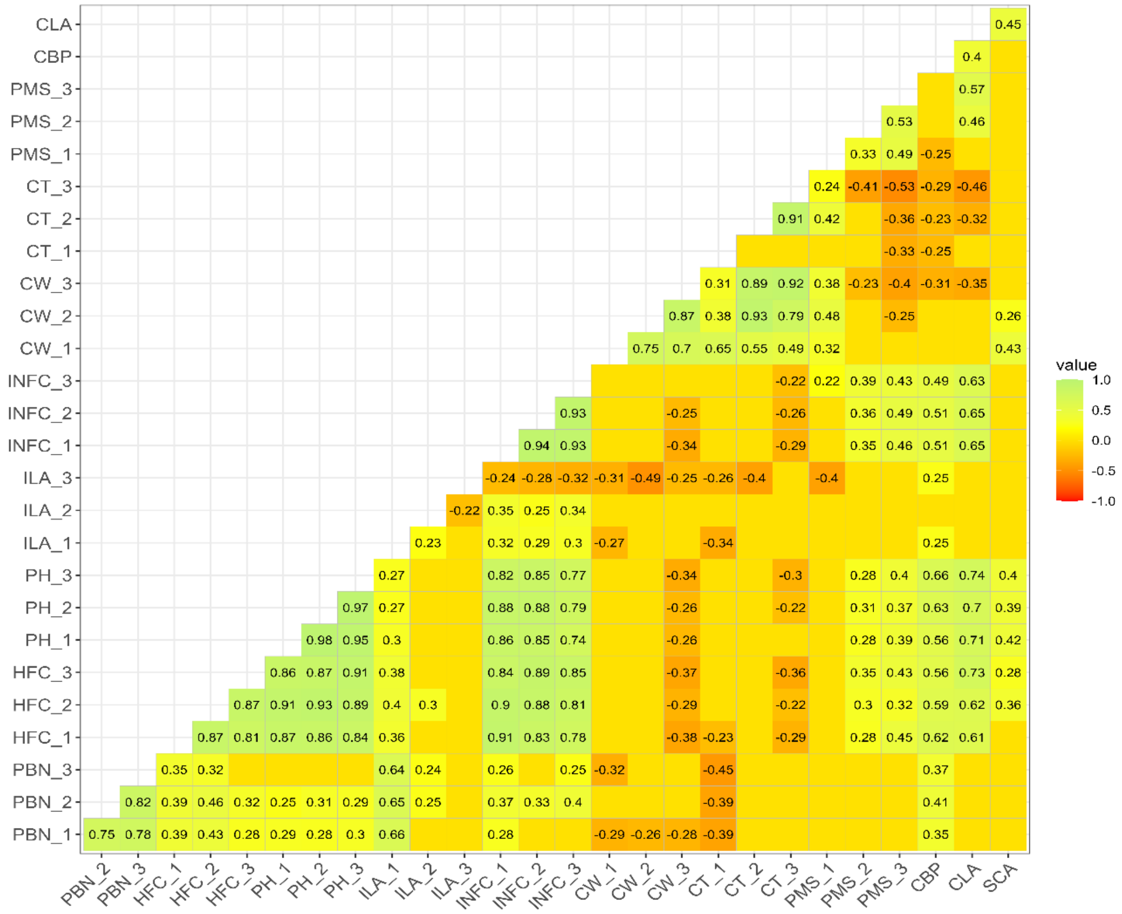

Figure 1.4. Pearson correlation analysis between variables across environments. Variables that presented significant GxE were treated as independent variables based on the variables' association with the first two factors.

Some relationships between the variables are interesting from the point of view of sesame breeding. It is observed that HFC (height to first capsule) in the different environments presents a high positive correlation with PH (plant height) and INFC (number of internodes to first capsule). In addition, HFC presents a medium positive correlation with PBN (number of primary branches), ILA (internode length), PMS in environments 2 and 3, CBP (number of capsules per plant), CLA (capsule length) and SCA (number of seeds per capsule) (only HFC_2 and HFC_3). The variable ILA in environment 3 (ILA_3) presents a negative correlation with INFC, CW (capsule width), CT (capsule thickness), and PMS in environment 3 (PMS_3). It is observed that CLA (capsule length) presents a positive average correlation with SCA and with PMS in environments 2 and 3 (0.46 and 0.57, respectively).

For most variables, the correlation with each other in the different environments is positive and high, such as PBN, HFC, and PH. However, some variables show no correlation, weak correlation, or negative correlation. The variable ILA in environment 1 (ILA_1) shows no correlation with ILA_2 and a negative correlation with ILA_3. Similarly, CT in environment 1 shows no correlation with CT in the other environments and indicates the influence of the environment on the phenotypic performance of the genotypes. To deepen the study of the relationship between the variables, factor analysis with orthogonal rotation was performed to maximize the relationship between the variables (Figure 1.5).

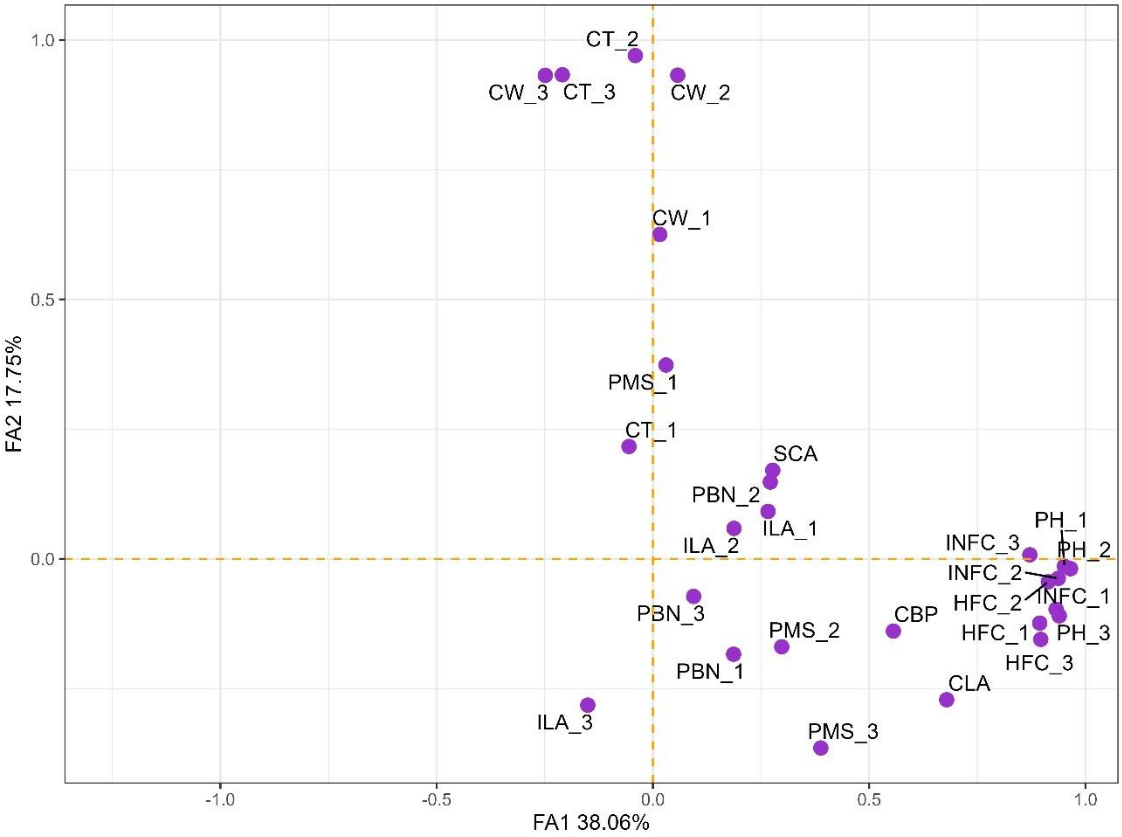

Figure 1.5. Maximized relationship between the variables in the first two rotated factors (FA1 and FA2) of a rotated factor analysis.

Two groups are formed, with variables that are extremely related to each other. The first group is positively related to the FA1 axis. It is formed by the variables PH, HFC, and INFC of the three environments and shows a high correlation between these variables, responding similarly to the different environments. It is observed that although these variables present a significant correlation with the variables ILA, PBN, PMS, CBP, CLA, and SCA in the Pearson correlation analysis (Figure 4), in this graph, these variables are more separated, indicating a lower direct relationship. The high correlation of the variables CW and CT is evidenced in the FA2 axis. The variable CT in environment 1 (CT_1) is very close to the origin of the RC2 axis, showing a low relationship of this variable with this axis and reinforcing the different response of this variable in environment 1 in relation to the others, as previously discussed. The same is observed for the variables PBN, ILA, and PMS, which show a low relationship in their response across the different environments. The trait PMS along the environments (PMS_1, PMS_2, and PMS_3) are distributed along the FA2 axis. On the other hand, the traits PBN (PBN_1, PBN_2, and PBN_3) and ILA (ILA_1, ILA_2, and ILA_3) are dispersed in the graph, indicating that they are more affected by the genotype x environment interaction. The traits CLA and CBP are closer and indicate a similar response in the environments.

When computing the Euclidean distance matrix using the adjusted mean for the genotypes, the greatest divergence was observed for AGUU x L7 (D = 11616.14). In contrast, the closest genotypes were the L2 x L6 lines (D = 20.7). The mean distance between the genotypes was D = 1328.17. Cluster analysis based on the Euclidean distance matrix divided the 28 genotypes into four groups (Figure 1.6). The commercial genotype BRS Anahí (AGUU1) was the most divergent, clustering in isolation. In addition, a reorganization of the genotypes between the groups was observed in relation to the clustering with molecular data. The BRS Seda (SGUU1) genotype was the only one that had grouped separately from the other commercial genotypes in the molecular cluster; however, in the present analysis, only SGUU1 and Trebol (TGUU1) were grouped in the same group, while the other genotypes were distributed among the groups.

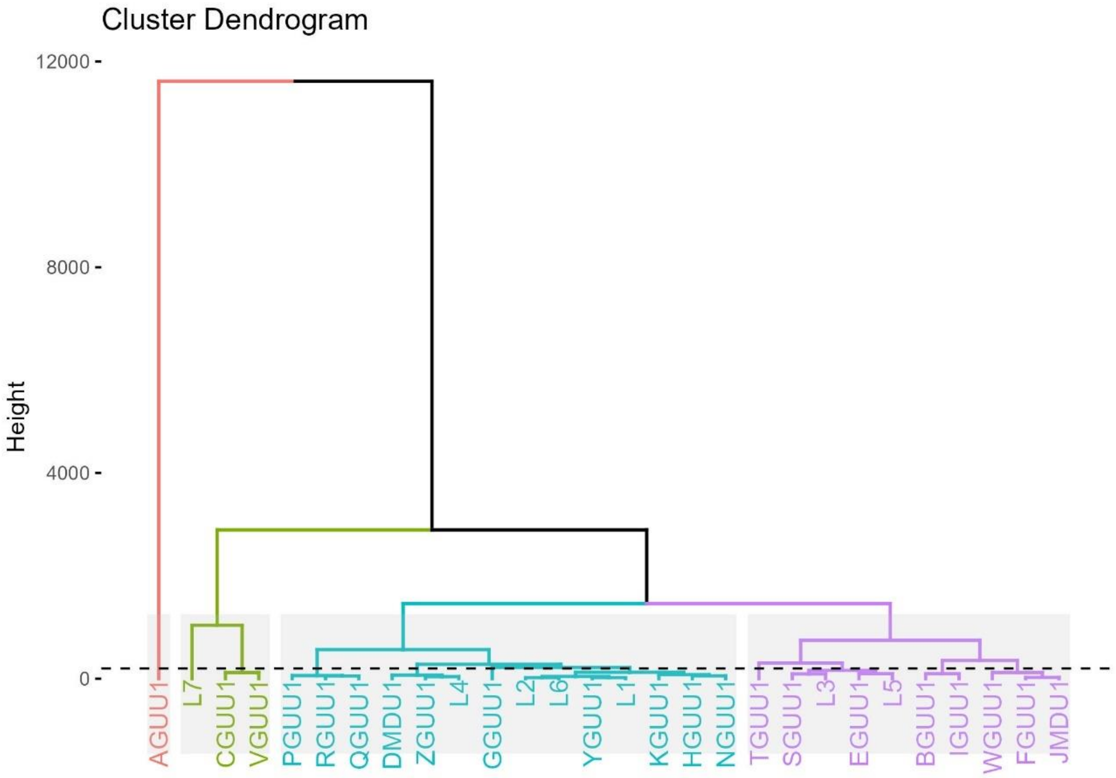

Figure 1.6. Cluster analysis for the 28 sesame genotypes performed from a Euclidean distance matrix estimated from the genotype blups. Two main groups are formed with the horizontal cut at the height of the dashed line. The shaded rectangles observe four subgroups.

By establishing the ideotype of the evaluated characteristics and classifying the genotypes based on the FAI-BLUP index, it was possible to observe which genotypes were the most promising, with the best performance in terms of adaptability and stability among the evaluated environments (Figure 1.7). The genotype with the best performance was YGUU1, followed by genotypes IGUU1, KGUU1, and BGUU1. It is possible to observe that among the most promising genotypes (red dots), there are genotypes belonging to different groups in the hierarchical analysis, with greater divergence. The genotypes YGUU1 and KGUU1 were grouped together in group 3, and the genotypes BGUU1 and IGUU1 were grouped in group 4. In addition, if we consider the possibility of increasing the number of selected genotypes, other genotypes, such as HGUU1, L3, and L6, appear to be good alternatives for directing crosses with genotypes from other groups, being viable options for generating segregating populations with greater adaptability and stability.

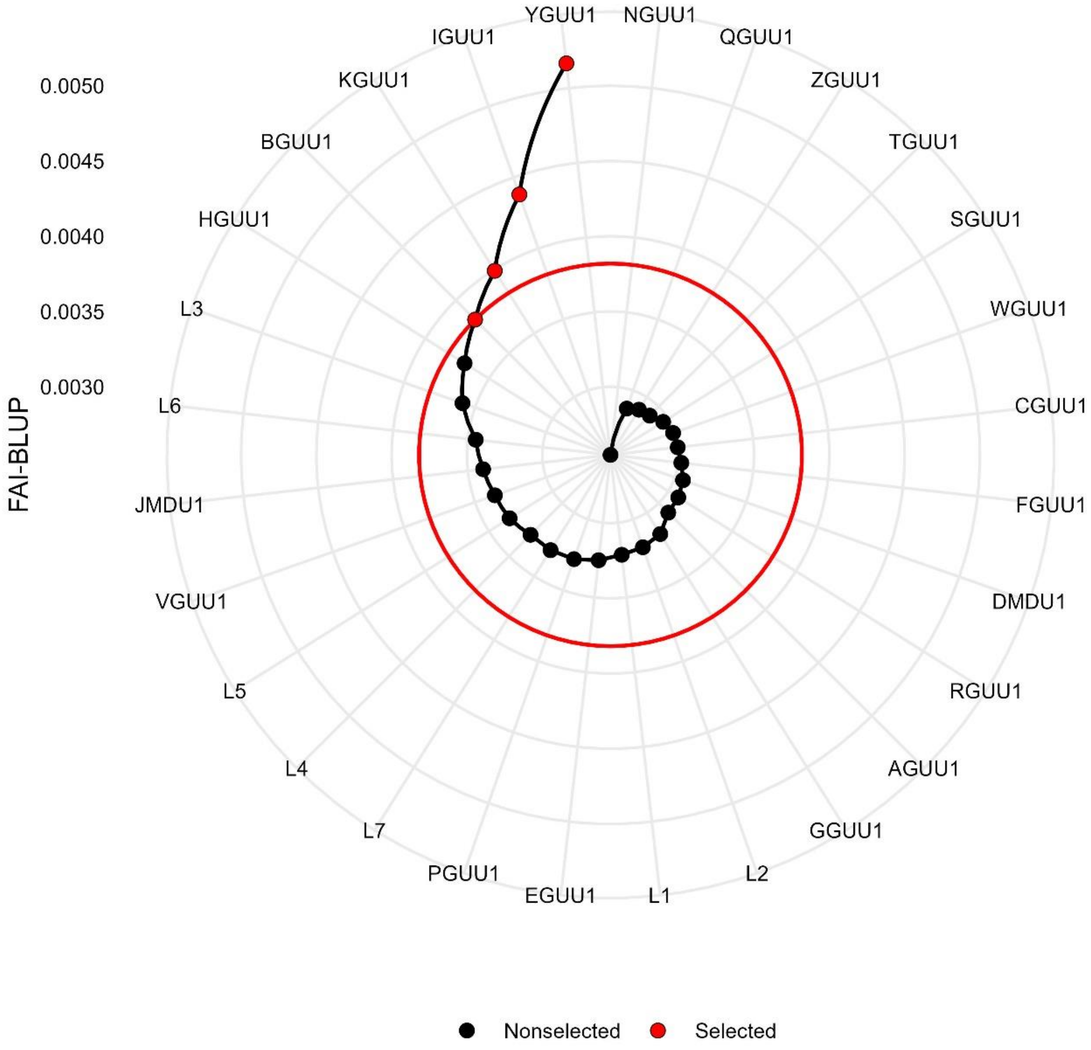

Figure 1.7. Classification of genotypes evaluated by the FAI-BLUP index, based on the estimated genetic value for the characteristics under study.

2.4. Discussion

Studies of genetic structure and diversity are important tools for characterizing species populations, whether for conservation purposes, exploitation of natural resources, or establishment of breeding programs. This study provided important information regarding the genetic relationships among sesame genotypes that can facilitate the definition of strategies for the crop breeding program in Brazil. Most of the genotypes evaluated in this study have unknown origin, except for the Brazilian commercial cultivars KGUU1 (K3), AGUU1 (BRS Anahí), SGUU1 (BRS Seda), and

TGUU1 (Trebol), and the lines L1 to L7, developed in the breeding program of Sebra Agrícola SA, and this study allowed us to understand the genetic relationships among these genotypes.

Sesame is an autogamous species, and heterozygosity levels are expected to be low (Kinman and Martin 1954). However, higher rates of observed heterozygosity were found for some markers evaluated, reaching 1. It is important to note that heterozygous loci were observed even in commercial genotypes, such as SGUU1 (BRS Seda) and AGUU1 (BRS Anahí). A possible explanation for the frequency of heterozygotes observed in the genotypes studied may be a low rate of cross-fertilization in the species, which can reach 9.6% (0.5-9.6%) (LAKGHAM 1944). Excess of heterozygosity in sesame has also been reported by Cho et al. (2011), and the same author discusses the possibility that the frequency of heterozygotes in sesame is related to errors during the selection of more productive genotypes by producers and the mixing of populations.

The mean value of observed heterozygosity obtained in the present work (0.36) agrees with the results of a sesame diversity study found in the literature. Teklu et al. (2021) and (2022b) reported a Ho of 0.43 in both studies, characterizing the genetic diversity of 100 sesame genotypes; Gebremichael et al. (2011) reported a Ho of 0.23 when characterizing the diversity of 50 sesame genotypes. However, Araújo et al. (2019) and Asekova et al. (2018) reported a lower frequency of heterozygotes than that found in the present study (0.12 and 0.01, respectively). The average number of alleles per locus of 1.42 found can be considered low when compared with other studies in the literature that report higher averages than this, such as 3.37 (Pandey et al. 2015), 6,9 (Asekova et al. 2018), 7,6 (Cho et al. 2011) and 19,5 (Adu-Gyamfi et al. 2019), which indicates the need to insert new genotypes with different alleles to increase the allelic variability of the population.

Despite the low number of alleles per locus, it was possible to observe genetic divergence among the genotypes studied, highlighted by the results of the clustering and genetic structure analysis obtained. The results from the Structure software with the SSR markers showed the formation of two main groups, with 11 and 16 genotypes, respectively. The same pattern is observed in the hierarchical clustering, where two main groups are formed. Forming two main divergent groups shows the possibility of directing crosses between genotypes from different groups for hybridization.

The insertion of the genomic matrix obtained from the SNPs for 21 of the 28 genotypes evaluated in the models to estimate the genetic parameters proved efficient for six evaluated characteristics. Using models that allow the incorporation of genetic similarity matrices can result in significant gains in the accuracy of selection and, consequently, in the improvement of genetic progress, especially in multi-environment experiments, where estimates of the GxE interaction become fundamental (Dekkers 2007).

Mixed models allow capturing both genetic and GxE interaction variations more accurately, in addition to enabling the calculation of BLUPs (Best Linear Unbiased Predictions), which can be used to improve genetic selection, among other genetic parameters (Piepho et al. 2008; Resende et al. 2012). GxE interaction determines how different genotypes respond to environmental variations (Ramalho et al. 2012). Most of the traits evaluated in this study showed significant GxE interaction, as shown in Table 1.4, and reveals that the performance of the genotypes for these traits is inconsistent across the environments evaluated. Strategies that seek to perform selection in multiple environments and identify more stable genotypes with broad adaptability and specific adaptability for certain environments are important in genetic breeding programs since genotypes with lower GxE interaction are considered more stable. In contrast, genotypes with high specific adaptability are more appropriate for specific environments (Gauch & Zobel 1996; Colombari Filho et al. 2013). Identifying sesame genotypes with greater environmental stability is particularly important since the breeding of this crop in Brazil is in its initial stages, where the set of subprograms for the development of genotypes adapted to specific environments is not viable. The number of capsules per plant and capsule length traits did not present significant GxE, showing high stability across environments. It is important to emphasize that environments 1 and 2 are environments whose experiments were repeated over time and here are considered as different environments, which may lead to bias in the estimates of the GxE interaction due to the confounding of the effects of location and year that may be affecting the genotypes evaluated and was not considered in the present study.

Table 1.4 shows that the trait's number of secondary branches (PBN) and internode length (ILA) showed a significant GxE interaction but without significant genetic variance, suggesting that, for these traits, the environmental effect has a greater influence on the genotypic performance of the genotypes than the genetic differences between them. This may be related to a greater influence of the environment on the phenotypic expression of the trait or the phenotypic plasticity of the genotypes, which affect the phenotypic expression of the trait in response to the environment (Des Marais et al. 2013). Branch number is a trait with high phenotypic plasticity in sesame. Ramesh et al. (2021) reported that the number of sesame branches was one of the characteristics with the greatest phenotypic plasticity as a function of spacing observed in their experiment. Similar results were reported by Gadri et al. (2020). The authors observed that branching was highly influenced by the management and environmental changes to which the genotypes were subjected.

Currently, the main requirements for sesame cultivars in Brazil are productivity, disease resistance, adaptation to mechanization, reduced losses, indehiscence and determined growth, and higher grain quality (Georgiev et al. 2008; Stamatav et al. 2020). Adapting sesame to mechanized harvesting involves the development of smaller cultivars with lower height at the first capsule since tall plants lead to significant losses during harvesting. The characteristics evaluated in this study can be classified into variables related to plant size, grouping the variables plant height (PH), height to the first capsule (HFC), and internodes number to the first capsule (INFC) since these variables directly influence plant size; and variables related to productivity, comprising the other variables.

Correlation analysis allows for measuring levels of genetic association between two or more characteristics and enables the indirect selection of low-heritability characters (Hallauer and Miranda 1988). Initially, the correlations obtained in the present study show that, in general, plant size variables are directly related to productivity variables, with medium to high correlations (Figure 1.4). These results are supported by the results reported by Haruna et al. (2012) and Aye & Htwe (2019), where plant height positively correlated with plant productivity, which shows that the development of smaller genotypes with high productive potential may not be trivial.

Although some traits evaluated in this study did not present significant genetic variation among genotypes, high heritability values were found for important traits to be explored in the crop improvement program. The productivity measured among genotypes did not present significant genetic variance. Productivity is a complex trait highly influenced by the environment. Studying the nature and level of association of morphological traits linked to productivity can make the selection of genotypes with greater productive potential more effective, especially in this case (Aye and Htwe

2019). In sesame, the traits that most influence the productivity of genotypes are the number of capsules per plant, the number of seeds per capsule, and seed weight (Mei et al., 2021). By maximizing the relationship of variables in the factor analysis, it is observed that the variables related to plant size are not as strongly related to the variables related to productivity as observed in the correlation analysis. In addition, traits such as capsule width (CW), capsule thickness (CT), PMS, and primary branch number (PBN) are highly correlated to the RC2 axis, indicating a low relationship with PH, HFC, and INFC. This shows that changes in plant size may not significantly impact traits related to productivity. Furthermore, identifying other factors and traits that can help maximize productivity is essential. The trait internode length in environment 3 (CLA_3) showed a negative correlation with capsule width, capsule thickness, and PMS in environment 1, showing that reducing this trait may be an important approach for increasing productivity. However, this trait is highly influenced by the GxE interaction, and it may not be easy to manage to maximize productivity. Haruna et al. (2012) report that dry mass per plant, leaf area index, and number of branches showed a positive correlation with grain yield. The authors report that the number of branches was one of the traits that contributed most to grain yield. Aye & Htwe (2019) also reported a significant correlation between several morphological traits and productivity, including days to 50% flowering, number of primary branches, number of capsules per plant, and capsule length, which were also evaluated in the present study. The trait days to 50% flowering, together with the cycle, although not analyzed in the present study, are especially important for sesame since, in Brazil, sesame is grown as a second crop with a planting period limited by the soybean harvest. Late soybean harvest can make sesame planting unfeasible, making short-cycle varieties interesting.

Selection based on selection indices is an important approach in plant breeding, especially when selecting multiple traits of interest, allowing for a combination and balance between the most favorable and interesting traits. (Wellmann 2023). The FAI-BLUP index and the clustering results show that it is possible to carry out crosses between contrasting and promising genotypes. Possible crosses would involve YGUU1 x BGUU1, YGUU1 x IGUU1, KGUU1 x YGUU1, KGUU1 x IGUU1, BGUU1 x HGUU1, and HGUU1 x IGUU1, representing a cross of superior and divergent genotypes. When ranking the genotypes to identify the most promising genotypes, the most planted genotype in Brazil, the K3 genotype (KGUU1) (Elaine Cotta 2024), was the third most promising genotype. This genotype, despite presenting interesting agronomic characteristics, lacks grain quality and low PMS. Among the proposed crosses, the cross involving IGUU1 and KGUU1 is especially interesting since these genotypes present placental adhesion of the grains, an essential characteristic to reduce losses during mechanized harvesting (Stamatov and Deshev 2018; Qureshi et al. 2022).

2.5. Conclusion

This study allowed the characterization of the genetic structure and diversity of sesame genotypes belonging to the germplasm bank maintained by Sebra Agrícola SA. Considerable genetic divergence among the genotypes studied was observed, although microsatellite analysis showed a narrow genetic base with low allelic variation, indicating the need for the insertion of new alleles in the breeding population. The genetic divergence observed will allow the direction of new crosses to be carried out in the breeding program and the obtaining of new genetic combinations with greater potential to generate genotypes of greater agronomic interest. Regarding the GxE interaction, nine traits were shown to be highly influenced by the environment, showing that the behavior of the phenotypic traits of sesame is not consistent in different environments and reinforces the need to select genotypes with greater stability and broad adaptability. These results show the importance of considering environmental variation in sesame breeding, as done in the present work. In the same way, employing statistical methods that consider the effects of GxE interaction on genotypic performance and the incorporation of other available tools, such as the insertion of the genetic relationship matrix, to make genotypic selection more accurate is also important.

2.6. References

Abbas S, Sharif MK, Sibt-e-Abbas M, et al (2022) Nutritional and therapeutic potential of sesame seeds. J Food Qual 2022:6163753

Adu-Gyamfi R, Prempeh R, Zakaria I (2019) Diversity assessment of some sesame (Sesamum indicum L.) genotypes cultivated in northern Ghana using morphological and simple sequence repeat (SSR) markers. Advances in Agriculture 2019:

Amadeu RR, Cellon C, Olmstead JW, et al (2016) AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Genome 9:plantgenome2016-01

Araújo E de S, Arriel NHC, dos Santos RC, de Lima LM (2019) Assessment of genetic variability in sesame accessions using SSR markers and morpho-agronomic traits. Aust J Crop Sci 13:45–54

Arriel NHC, ARAUJO AE, SOARES JJ, et al (2006) Cultivo do gergelim. Sistemas de Produção, Embrapa Algodão

Asekova S, Kulkarni KP, Oh KW, et al (2018) Analysis of molecular variance and population structure of sesame (Sesamum indicum L.) genotypes using simple sequence repeat markers. Plant Breed Biotechnol 6:321–336. https://doi.org/10.9787/PBB.2018.6.4.321

Aye M, Htwe NM (2019) Trait association and path coefficient analysis for yield traits in Myanmar sesame (Sesamum indicum L.) germplasm. Journal of Experimental Agriculture International 41:1–10

Caixeta ET, Ferrão LF v., Maciel-Zambolim E (2013) Marcadores moleculares. In: BORÉM A, FRITSCHE-NETO R (eds) Biotecnologia aplicada ao melhoramento de plantas. Suprema, Visconde do Rio Branco, pp 9–30

Cho Y-I, Park J-H, Lee C-W, et al (2011) Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes Genomics 33:187–195

Colombari Filho JM, de Resende MDV, de Morais OP, et al (2013) Upland rice breeding in Brazil: a simultaneous genotypic evaluation of stability, adaptability and grain yield. Euphytica 192:117–129

Couch A, Gloaguen RM, Langham DR, et al (2017) Non-dehiscent sesame (Sesamum indicum L.): its unique production potential and expansion into the southeastern USA. J Crop Improv 31:101–172

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One 11:e0156744

De Resende MD V (2007) SELEGEN-REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos.

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. Journal of animal breeding and genetics 124:331–341

Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annu Rev Ecol Evol Syst 44:5–29

Elaine Cotta (2024) Gergelim conquista produtores e cresce mais de 250% nas exportações. In: https://agfeed.com.br/caminhos-do-agro/gergelim-conquista-produtores-e-cresce-mais-de-250-nas-exportacoes/

El-Aziz A, Farid S, Elkomey S (2016) Evaluation of molecular and phenotypic diversity in relation to heterosis in some tomato lines under different climatic conditions. Journal of Agricultural Chemistry and Biotechnology 7:141–151

EL-Mansy AB, Abd El-Moneim D, ALshamrani SM, et al (2021) Genetic diversity analysis of tomato (Solanum lycopersicum L.) with morphological, cytological, and molecular markers under heat stress. Horticulturae 7:65

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Farran MT, Uwayjan MG, Miski AMA, et al (2000) Performance of broilers and layers fed graded levels of sesame hull. Journal of Applied Poultry Research 9:453–459

Gadri Y, Williams LE, Peleg Z (2020) Tradeoffs between yield components promote crop stability in sesame. Plant Science 295:110105

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907

Gebremichael DE, Parzies HK (2011) Genetic variability among landraces of sesame in Ethiopia. Afr Crop Sci J 19:

Georgiev S, Stamatov S, Deshev M (2008) Requirements to sesame (Sesamum indicum L.) cultivars breeding for mechanized harvesting. Bulgarian journal of agricultural science 14:616–620

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State Univ. Press, Ames. Quantitative genetics in maize breeding 2nd ed Iowa State Univ Press, Ames

Haruna IM, Aliyu L, Olufajo OO, Odion EC (2012) Contributions of some growth characters to seed yield of sesame (Sesamum indicum L.). Journal of Food and Agriculture Science, Londres 2:9–14

Hoffmann TJ, Kvale MN, Hesselson SE, et al (2011) Next generation genome-wide association tool: Design and coverage of a high-throughput European-optimized SNP array. Genomics 98:79–89. https://doi.org/10.1016/j.ygeno.2011.04.005

Kamvar ZN, Tabima JF, Gřunwald NJ (2014) Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:1–14. https://doi.org/10.7717/peerj.281

Kassambara A, Mundt F (2017) Factoextra: extract and visualize the results of multivariate data analyses. R package version 1:337–354

Kinman ML, Martin JA (1954) Present Status of Sesame Breeding in the United States 1. Agron J 46:24–27

LAKGHAM DG (1944) Natural and controlled pollination in sesame. Journal of Heredity 35:255–256

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Meena R, Singh B, Meena KC, et al (2018) Performance of Front Line Demonstrations on Sesame (Sesamum indicum L.) in Karauli District of Rajasthan, India. Int J Curr Microbiol App Sci 7:1507–1511

Mei H, Liu Y, Cui C, et al (2021) QTL mapping of yield-related traits in sesame. Molecular Breeding 41:1–14

Nandakumar N, Singh AK, Sharma RK, et al (2004) Molecular fingerprinting of hybrids and assessment of genetic purity of hybrid seeds in rice using microsatellite markers. Euphytica 136:257–264

Oboulbiga EB, Douamba Z, Compaoré-Sérémé D, et al (2023) Physicochemical, potential nutritional, antioxidant and health properties of sesame seed oil: a review. Front Nutr 10:1127926

Olivoto T, Lúcio AD (2020) metan: An R package for multi-environment trial analysis. Methods Ecol Evol 11:783–789

Pandey SK, Das A, Rai P, Dasgupta T (2015) Morphological and genetic diversity assessment of sesame (Sesamum indicum L.) accessions differing in origin. Physiology and Molecular Biology of Plants 21:519–529. https://doi.org/10.1007/s12298-015-0322-2

Patil MK, Lokesha R (2018) Estimation of genetic variability, heritability, genetic advance, correlations and path analysis in advanced mutant breeding lines of sesame (Sesamum indicum L.). J Pharmacogn Nat Prod 4:1–5

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure using Multilocus Genotype Data. 155:945–959

Qureshi M, Langham DR, Lucas SJ, et al (2022) Breeding history for shattering trait in sesame: classic to genomic approach. Mol Biol Rep 49:7185–7194

Ramalho MAP, Abreu A de FB, Santos JB dos, Nunes JAR (2012) Aplicações da genética quantitativa no melhoramento de plantas autógamas. Lavras: Ufla 365–456

Ramesh K, Suresh G, Qureshi AA, et al (2021) Plant density and nitrogen interaction on productivity of summer sesame (Sesamum indicum L.). Indian Society of Oilseeds Research 38:100–105

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution (N Y) 55:1095–1103

Resende MD V, Resende Jr MFR, Sansaloni CP, et al (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist 194:116–128

Rufo R, Alvaro F, Royo C, Soriano JM (2019) From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. PLoS One 14:e0219867

Sankar D, Rao MR, Sambandam G, Pugalendi K V (2006) Effect of sesame oil on diuretics or ß-blockers in the modulation of blood pressure, anthropometry, lipid profile, and redox status. Yale J Biol Med 79:19

Schulman AH (2007) Molecular markers to assess genetic diversity. Euphytica 158:313–321

Sensoy S, Büyükalaca S, Abak K (2007) Evaluation of genetic diversity in Turkish melons (Cucumis melo L.) based on phenotypic characters and RAPD markers. Genet Resour Crop Evol 54:1351–1365

Stamatav S, Ishpekov S, Deshev M, et al (2020) SEED RETAINING MODEL OF NON-DEHISCENCE SESAME (Sesamum indicum L.) GENOTYPES AT RIPENING. Scientific Papers Series A Agronomy, 63 (1) 541–546

Stamatov S, Deshev M (2018) Review of the methods for breeding of sesame varieties (sesamum indicum L.) in Bulgaria. Bulgarian Journal of Agricultural Science 24:411–416

Team RC (2025) R: A language and environment for statistical computing

Teklu DH, Shimelis H, Abady S (2022a) Genetic Improvement in Sesame (Sesamum indicum L.): Progress and Outlook: A Review. Agronomy 2022, 12, 2144. 2144

Teklu DH, Shimelis H, Tesfaye A, et al (2021) Genetic variability and population structure of Ethiopian sesame (Sesamum indicum L.) germplasm assessed through phenotypic traits and simple sequence repeats markers. Plants 10:1129

Teklu DH, Shimelis H, Tesfaye A, Shayanowako AIT (2022b) Analyses of genetic diversity and population structure of sesame (Sesamum indicum L.) germplasm collections through seed oil and fatty acid compositions and SSR markers. Journal of Food Composition and Analysis 110:104545

Uzun B, Arslan Ç, Furat Ş (2008a) Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (Sesamum indicum L.). J Am Oil Chem Soc 85:1135–1142

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Wang M, Huang J, Liu S, et al (2022) Improved assembly and annotation of the sesame genome. DNA Research 29:dsac041

Wei P, Zhao F, Wang Z, et al (2022a) Sesame (Sesamum indicum L.): A comprehensive review of nutritional value, phytochemical composition, health benefits, development of food, and industrial applications. Nutrients 14:4079

Wei P, Zhao F, Wang Z, et al (2022b) Sesame (Sesamum indicum L.): A comprehensive review of nutritional value, phytochemical composition, health benefits, development of food, and industrial applications. Nutrients 14:4079

Wellmann R (2023) Selection index theory for populations under directional and stabilizing selection. Genetics Selection Evolution 55:10

Zhang H, Langham D, Miao H (2021) Economic and academic importance of sesame. In: The Sesame Genome. Springer, Cham, pp 1–18

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Genomic designing of climate-smart oilseed crops. Springer, pp 371–440

2.7 Supplementary material

GENETIC PARAMETERS IN SESAME ACCESSIONS ESTIMATED BY DNA MOLECULAR MARKERS AND PREDICTED GENETIC VALUES

Table S1 – List of microsatellite markers used in the molecular characterization of the 27 sesame genotypes.

Name	Sequences (5' -> 3')	Amplicon (pb)	Motif	Allele number	Amplicon (pb)	Reference
HS03 F	ACT TGG CCT ACG AAC AGC TT			4	180 - 210	
HS03 R	GGA AAA ACA CCT CGG AAG AA					
HS105 F	CCA TCC CAC AGC TAA CAC AC			3	170 - 200	
HS105 R	GGG GTT TCT TTC TTT TTC CC					
HS152 F	GGA CCC CTT TAC GTG AAA AA			4	180 - 380	
HS152 R	CTA TCC AGG TTA AGG CCC AG					
HS211 F	CTC TTC GAG TCA ACA CCT GC			3	150 - 170	
HS211 R	ACC CAA TCT TCA CGG CTA TC					
HS218 F	TGC TAT GTG AGT GCG TGA AA			2	295 - 300	
HS218 R	TCT CCT AAT TCC CAA CGT CC					
HS22 F	AGC CCA CGA AAC TCT TCC TA			4	295 - 230	
HS22 R	GTT CGA GCC CTT CAA GAT TC					
HS226 F	AGG GGA GAA TTC AGG TGT TG			5	260 - 350	
HS226 R	GGG CAC TGG AGA TTT CTG TT					
HS255 F	TTG CAT CAG AGG ACC AAT AG			5	270 - 390	
HS255 R	AGC ATA GGA CCA GCT TCC TC					
HS260 F	GAA GCG CAA TTG TAG TTC CA			5	300 - 400	
HS260 R	GGA GGA GGA ATA GGG AAA GG					
HS265 F	CTC ATG CAC CAA CTC TCA CA			2	150 - 160	
HS265 R	CAA GGC GTT GTG TCT CAA AT					
HS38 F	TGA AAT GTG AAG AGC AAC GC			4	290 - 390	
HS38 R	TGA AGA CGG ACC CAT TAC AA					
HS53 F	GAA GCT TGA AGA GAG GAG GG			5	100 - 110	
HS53 R	ATG GAA CTT CTC CGA TCA CC					
HS71 F	GAG GGA AAG AGT CGT CCT TG			5	290 - 350	
HS71 R	TTG AAG AGG ATC ATC CCA CC					
HS83 F	ACT CCA GCC TCA AGC AAT CT			4	195 - 205	
HS83 R	ATC CCA GTC TTG AGC AGA GG					
HS84 F	GTT ACC AAT GCC CAC TTC CT			3	210 - 290	
HS84 R	TTC CAT CCG CTT CTA CCT TT					
ZM0667 F	AGA GAT CCA ATC AAA TGT GC			3	200 - 300	Wang et al., 2012
ZM0667 R	CAC CAA TAG GAA CAA ATA CTC G					
ZM0737 F	CAC CAC TCT CTC TCA CTT CAG			5	200 - 300	Wang et al., 2012
ZM0737 R	AGA AGA TGC TTC TGT CAG CTA					
ZM0878 F	ACG AGG TAC AAA ATC AAC CT			4	400 - 530	Wang et al., 2012
ZM0878 R	CAC TGC AGA ACT TGT CTA GGA					
ZM1043 F	TAA CCA AAC ACT ACC AAC CTC			3	340 - 400	Wang et al., 2012
ZM1043 R	CAA ACC AAG AAG AGC AAA TC					
ZM1062 F	GGA CTA GAT ATC GGT GAT CCT			2	395 - 480	Wang et al., 2012
ZM1062 R	GGT TAA ACA AGT CCG TCT TTC					
ZM1308 F	TGA GAT TTC CTC ATC TGT GAC			3	460 - 510	Wang et al., 2012
ZM1308 R	GCA CAT ACT TCA ACA GTT TCC					
ZM1384 F	AAT TGG AAA CGT AGA GGA GAG			3	400 - 500	

ZM1384 R	GAG TAG AAC ACG CGA ATA AAC					Wang et al., 2012
ZM1459 F	CCT GAC TTT CAC AAG AAC TGA			4	490 - 590	Wang et al., 2012
ZM1459 R	TGC CAC ATT TTG TTA CAC AC					
ZM1531 F	CAC CAC AAA CAT TCT TCT CTC			4	85 - 105	Wang et al., 2012
ZM1531 R	GGC TTA ACC AAT GAA GTT GT					
ZM1546 F	TTA AAC TCT TCG CGT GTA AGC			5	495 - 505	Wang et al., 2012
ZM1546 R	AAC CTT CCA CGA CAC TAC ATA					
M129539 F	CTC AGT TTG CCC AAG ATC AC			6	97 - 115	
M129539 R	CTA ACA TAT ATT GAC TGC AA					
SIN_1024652 F	AGC TGG AGT ATT ACT GTT GG	288	(TCAG)4	3	97 - 105	
SIN_1024652 R	ACC CTT AAC TCT GGA TTC TC					
SiCL1 F	CTA CTC CGA AGT CAG TGT TGG A	190	(TC)6			
SiCL1 R	TTG ACT ACA ATT TTC CTA TGT TTT C			4	290 - 305	
SIN_1011197 F	GTC CCA AGA ATT CTA CTT CC	319	(CTT)3			
SIN_1011197 R	AGA GCA GAT AAC ACA CTT CC			3	145 - 155	
SIN_1015742 F	GGA TAG TTA GTT CGT TGA GC	239	(AAG)$_3$			
SIN_1015742 R	CTT ATA CCT CTT CTC GAT GC			3	310 - 400	
SIN_1003263 F	TCT ATG TGG TTC TCT CTT CC	305	(GCT)$_3$			
SIN_1003263 R	AAT ATC CCT GCC TAG TAT GG			4	85 - 100	
SIN_1015387 F	CCC TTA TAC ATG GAC TAT GC	256	(AT)$_6$			
SIN_1015387 R	CTA AAA GCT ACA AGC TCA CC			4	230 - 300	
SIN_1005078 F	GGA GAA ATA CAT GAA TGA GC	249	(GATTG)$_3$			
SIN_1005078 R	GTC CAT GCA TAG AAA GAC TC			6	220 - 300	
SSI429-1 F	ATG CAT TGC ACG AAT ATC CCT C					
SSI429-1 R	CAT CAT CGG GGA AGT TTT TAT CAC			4	190 - 205	
SIN_1015383 F	AGC TGC TGG TAT TAG TAT GG	151	(TGCT)3			
SIN_1015383 R	TTC CTC TCT TCC ATA GAC TG			4	240 - 320	
SIN_1012084 F	GTA TAG GGA AAC CAG TAT GC	182	(CCT)$_4$			
SIN_1012084 R	CAG CAG CAG TAG ATG TAA CC			2	190 - 195	
M58311 F	ATT GTC TTC GTG GTG CTA CT					
M58311 R	CTA CAT GGT TAT CTA AAT GG			5	50 - 70	
M4E17-7 F	TGA GTC CAA ACC GGA CC					
M4E17-7 R	GAC TGC GTA CGA ATT CCA			5	85 - 100	
HS205 F	GAT GTG ATG GTG GTG AGA GC			3	180 - 205	
HS205 R	GCT ATG CGT TGA ATG AAG AC					
HS1956 F	CAC AGT TAC CAT GGG CAA AG					
HS1956 R	ACA CCC ATA TTT CCA GGC AT			3	205 - 270	
HS1638 F	TAG GAA GAG GCA TGT TCA CG			4	220 - 295	
HS1638 R	CCA TCT CCA CAT CTT GCA TC					
Y2129 F	GGG GCA CAG AGT GGA TGT AG					
Y2129 R	GGA CCA TGT AAT CCC AGC AC			2	105 - 120	
SIN_1003248 F	GTT CTT AAC ATT GGA CAT GG	195	(GAAGCT)3	2	220 - 250	
SIN_1003248 R	ATC GCC TCA TAG AAC TCC					
SIN_1009923 F	CCG ACT TAC CTT ATA GCA AC	269	(CT)$_8$			
SIN_1009923 R	TTC ATC TCT GCT TCT TCT TC			4	210 - 300	

SIN_1007213 F	AAC CAC CAC CTA CAG AAA C	236	(CCA)$_2$(TCA CCA)$_2$	5	185 - 280
SIN_1007213 R	GAG AAG TTT AGC TTT GAT GG				
M9E8-2 F	TGA GTC CAA ACC GGA GG				
M9E8-2 R	GAC TGC GTA CGA ATT CTG			5	170 - 270
M15E7-2 F	CAA ATG TGA ACC GGA TA			4	360 - 490
M15E7-2 R	GAC TGC GTA GCA ATT ATG				
SIN_1004921 F	ACT GGA GAA CTC ATT TCA AG	300	(ATC)$_3$		
SIN_1004921 R	AGT AAT TAT CAG CTG GGA TG		(GTT)$_3$	5	105 - 180
SIN_1025617 F	AGT TAT TAG GCA ACA GAT GG	274	(AAAC)$_3$		
SIN_1025617 R	ATC ATC AGC TTC TTG TTC TC				
SIN_1018135 F	GTT TAC ATG GAA AGA GTT GC	175	(TTA) 5		
SIN_1018135 R	GAT CAT GGA AAG TGA GTA GG			4	350 - 390
SIN_1004125 F	GTA TAT TCG ACT GGT GAA GG	290	(AGA)$_2$(GCG)$_2$	2	180 - 200
SIN_1004125 R	GAA GAT TAT AGC GGT AGA GC				
M19E17-3 F	TGA GTC CAA ACC GGA GG				
M19E17-3 R	GAC TGC GTA CGA ATT CCA			4	98 - 140
M13E11-2 F	TGA GTC CAA ACC GGC AT				
M13E11-2 R	GAC TGC GTA GCA ATT TCG			3	300 - 390
SIN_1017074 F	CAT ATT CTG AGG ACG TTA CC	225	(TTGCA)2		
SIN_1017074 R	TAC TTT TGA GGC TAT CAA CC			1	280
SIN_1003215 F	GTT TGA GGA AGC AGA TTA AG	160	(AAGG)3	4	270 - 300
SIN_1003215 R	GCA GTG TTC AAA GTG TAA GC				
SIN_1008249 F	GTC CAG TTC CAG AAA AGT G	182	(CGC)2(GGA)2		
SIN_1008249 R	TCG AAT ATT GAG TCG TAC AG			2	180 - 185
SIN_1013971 F	AGC TAA TCT ATG TGG GAA TG	175	(GGC)$_3$	2	80 - 85
SIN_1013971 R	TTC TCT TGT TCC TTG AGA AG				
SIN_1012023 F	TAA AGA TCG ACA GGT ACA CG	253	(AG)$_3$		
SIN_1012023 R	CTG CAG TTG TAG AAA GAT GC				
SIN_1005985 F	GAG TCA ATG TAC AGG TTT CC	232	(CTCA)$_3$		
SIN_1005985 R	TCA TAG GTG AGA ACT TGA GG			3	70 - 85
SIN_1018247 F	GCT ACT GGA AAA GAT GCT AC	153	(CGC) 3	3	280 - 305
SIN_1018247 R	CTT AAG ATC TTG CTT CTT GG				
SIN_1001964 F	AAG ATA GCA GCA ACT ACC C	343	(GAAG)$_3$	3	330 - 450
SIN_1001964 R	GCT ACT TCT GTT GCT ACT CC				
233				2	175 - 185
234					
235				2	220 - 230
236					
237				2	295 - 310
238					
239				2	230 - 240
240					
241				3	395 - 405
242					

3. CHAPTER II – QTL MAPPING FOR SESAME AGRONOMIC TRAITS VIA DARTSEQ

Abstract: Brazil's rapid expansion of sesame (Sesamum indicum L.) faces challenges such as low productivity and a lack of cultivars adapted to Brazilian cultivation conditions, particularly for mechanized farming. Adopting modern technologies, such as molecular markers for marker-assisted selection (MAS), has become essential in contemporary breeding programs, allowing for more efficient use of genetic and financial resources and significantly reducing the time needed to generate new cultivars. This study aimed to map quantitative trait loci (QTLs) using two sesame F_2 populations, SD and SJ, each with a common female parent, comprising 98 and 160 genotypes, respectively. The DArT-seq genotyping methodology was used to identify the single nucleotide polymorphisms (SNPs). There was considerable phenotypic variability in the 17 traits evaluated. From single marker analysis, 24 significant QTLs were identified when analyzing both populations. Separately, 29 and 26 QTLs were identified in SD and SJ populations, respectively. All traits have shown marker associations except for the traits stem color and capsule insertion position and highlighted the effectiveness of using these two F2 populations for QTL mapping in sesame. It was possible to find QTLs for traits directly related to field mechanization, such as those on chromosomes CM047434.1 and CM047433.1, related to growth habit and placental adherence, respectively, and productivity-related traits, such as those on chromosome CM047438.1 for number of capsules per plant, and CM047437.1 for number of seed per capsule, the greatest issues of sesame in Brazil. Our results represent a great achievement for sesame breeding in Brazil, being the first step to establish marker-assisted selection (MAS) in sesame breeding.

Keywords: *Sesamum indicum*. Quantitative trait loci. Single nucleotide polymorphism.

Resumo: A rápida expansão do gergelim (Sesamum indicum L.) no Brasil enfrenta desafios como baixa produtividade e falta de cultivares adaptadas às condições de cultivo brasileiras, particularmente para agricultura mecanizada. A adoção de tecnologias modernas, como marcadores moleculares para seleção assistida por marcadores (MAS), tornou-se essencial nos programas de melhoramento contemporâneos, permitindo o uso mais eficiente de recursos genéticos e financeiros, reduzindo significativamente o tempo necessário para gerar novas cultivares. Neste estudo, objetivamos mapear loci de características quantitativas (QTLs) usando duas populações F2 de gergelim, denominadas SD e SJ, cada uma com um genitor feminino comum, compreendendo 98 e 160 genótipos, respectivamente. A metodologia de genotipagem DArT-seq foi usada para identificação dos polimorfismos de nucleotídeo único (SNPs). Houve considerável variabilidade fenotípica nas 17 características avaliadas. A partir da análise de marcas simples, 24 QTLs significativos foram identificados ao analisar ambas as populações juntas. Separadamente, 29 e 26 QTLs foram identificados nas populações SD e SJ, respectivamente. Todas as características mostraram associações de marcadores, exceto as características cor do caule e posição de inserção da cápsula, e destacaram a eficácia do uso dessas duas populações F2 para mapeamento de QTL em gergelim. Foi possível encontrar QTLs para características diretamente relacionadas à mecanização dos cultivos de gergelim, como os localizados nos cromossomos CM047434.1 e CM047433.1 relacionadas ao hábito de crescimento e aderência placentária, respectivamente, e características relacionadas à produtividade, localizados no cromossomo CM047438.1 para número de cápsulas por planta e CM047437.1 para número de sementes por cápsula, os maiores problemas do gergelim no Brasil. Nossos resultados representam uma grande conquista para o melhoramento de gergelim no Brasil, sendo o primeiro passo para estabelecer a seleção assistida por marcadores (SAM) no melhoramento de gergelim.

Palavras-chave: *Sesamum indicum*. Locos de características quantitativas. Polimorfismo de nucleotídeo único.

3.1. Introduction

Sesame (*Sesamum indicum* L.) is an oilseed plant of great importance, being widely cultivated in various regions worldwide. The relevance of sesame has grown substantially in recent years due to a growing demand for grain driven by increased demand from industries. Moreover, the nutritional properties of sesame, which align with a change in the population's eating pattern, have led to an increase in grain consumption by the population, consequently increasing its demand (Islam et al., 2016; Rahman et al., 2019). Global sesame production jumped from 4.3 million tons in 2010 to almost 7 million tons in 2020 (FAO, 2022). This growing demand for sesame grain has allowed the expansion of sesame cultivation to regions and countries where cultivation is not traditionally practiced on a large scale.

Brazil, an agricultural giant known for being one of the most important players in food production in the world, has experienced a significant increase in exports of this grain, especially to important markets such as India, after the opening of exports (IBRAFE, 2023; MAPA, 2023). However, sesame production in Brazil faces significant challenges since large-scale cultivation requires adapted cultivars, especially for mechanization. The mechanization of sesame cultivation faces problems mainly related to harvesting since sesame is typically a dehiscent plant, with the dehiscence of the capsules being responsible for significant losses in production, starting with the early maturation of the capsules at the base that open releasing the seeds, up to the loss during mechanized harvesting that can reach 50% (Ahmed et al., 2023; Qureshi et al., 2022). Furthermore, the architecture of the plant, which normally has an indeterminate growth habit, leads to very tall plants with uneven flowering and growth, and sometimes it is possible to observe mature capsules at the base of the plant, concomitantly with green capsules at the top (Langham & Wiemers, 2002; Uzun et al., 2003; Uzun & Çağirgan, 2006). In addition to management, crop productivity and factors related to grain quality are important. In Brazil, a limited number of cultivars are available for cultivation, which are generally poorly adapted to new conditions (mechanized). Some cultivars, such as BRS Seda (Arriel et al., 2009) and BRS Anahí (Arriel et al., 2015), developed by the Brazilian Agricultural Research Corporation (Embrapa), have high production potential, but due to dehiscence and low adhesion of the grains to the capsules, crop productivity is compromised, leading the sesame breeding programs to seek for productivity and adaptation to mechanization primarily.

In the international scenario, sesame breeding programs to overcome these obstacles have advanced significantly. The incorporation of modern technologies arising from phenomics and genomics, such as marker-assisted selection (MAS) and quantitative trait loci (QTL) mapping, has allowed the identification of genes related to important agronomic characteristics of sesame, such as indehiscence and determinate growth habit, essential for the mechanization of the crop (Sabag et al., 2024; Uzun et al., 2003; Zhang et al., 2018). Two main genes related to sesame growth habit have already been identified. A gene of simple and recessive inheritance, *SiDt*, was mapped in a QTL region in the linkage group 8 (LG8) and regulates the transition from indeterminate to determinate growth (Zhang et al., 2016). Another gene was discovered by Zhang et al. (2018) in the linkage group 9 (LG9) and presents simple inheritance, with the determinate character being recessive in relation to the indeterminate, as in the previous gene. The difference between the two QTLs identified is that the first originated from an induction of mutation, while the second was mapped from a natural mutation.

For indehiscence, the first gene identified was the *SiCL1* gene from associative mapping, which, in addition to indehiscence, also controls the shape of the leaf in its recessive form (Zhang et al., 2018). Developing indehiscent cultivars was the target of sesame breeding programs for a long time. However, due to technical implications for mechanized harvesting imposed by the morphological characteristics of indehiscent capsules, given the greater difficulty in releasing the grains in fully closed capsules and grain quality compromise (Langham, 2001), the focus of breeding programs turned to the development of cultivars with placental adherence of grains, aiming to minimize losses during maturation and at the same time make mechanized harvesting more efficient, where the capsules would have a small opening and keep the seeds retained (Qureshi et al., 2022; Tripathy et al., 2019). Low information is available about the genetic mechanisms controlling this trait. However, studies developed to date suggest that this is a trait of complex and polygenic inheritance that involves modifications in the structure of the cell wall to retain the seeds in the embryonic axis of the capsules during maturation (Qureshi et al., 2022).

For yield, Mei et al. (2021) have found 46 QTLs for seven related traits in sesame (seed yield per plant, number of capsules per plant, number of seeds per capsule, thousand-grain weight, plant height, height to first capsule, and harvest index). Similar results were reported by Wu et al. (2014), who found 17 QTLs related to seven yield-related traits. QTL mapping has also been effectively utilized to identify QTLs linked to the seed coat color of sesame (Zhang et al., 2013) and Phytophthora resistance (Asekova et al., 2021). These findings show the importance of adopting modern technologies in breeding programs to accelerate the development of new sesame varieties.

In Brazil, sesame breeding research is still in its early stages, with few institutions involved and limited resources. Furthermore, the initial focus of sesame breeding in Brazil was on developing varieties adapted to manual harvesting on small properties, especially for the northeast region, which is poorly adapted to mechanization (Arriel et al., 2006). In Brazil, the adoption of technologies such as MAS in sesame breeding is still limited, but it represents an important path to accelerate and optimize the generation of new cultivars.

The integration of molecular tools into conventional breeding is facilitated by QTL mapping (Collard et al., 2005; Zhou et al., 2003). Nowadays, QTL mapping is a precondition for applying marker-assisted selection to important agronomic traits, particularly quantitative ones (Mei et al., 2021). A common problem in QTL mapping is the need to genotype a relatively large number of genotypes, with many genetic markers, to construct the maps (Mauricio, 2001). With the advent of next-generation sequencing (NGS) technologies, it has become possible to genotype populations with thousands of markers and create high-resolution genetic maps relatively inexpensively. A genetic map describes markers' position and relative distances on chromosomes (Collard et al., 2005; Nadeem et al., 2018). However, as demonstrated by Rebai et al. (1995) and Coffman et al. (2003), in studies where the reference genome is available, and the number of markers is large enough to provide good genome coverage, methodologies that do not require genetic maps, such as single marker analysis, can deliver reliable results.

Generally, populations for QTL mapping are derived from biparental crosses, which can limit the number of recombinant events, reduce the number of polymorphic markers, and sometimes result in low genetic diversity. Therefore, one strategy employed by breeding programs is to use multiple populations to incorporate more founder genotypes and increase the number of recombinant events in the mapping populations (Colasuonno et al., 2021). This strategy, in addition to increasing the resolution power of the mapping, also helps mitigate another big issue in QTL mapping: the limitation of using these markers in populations other than those in which they were identified. The effects estimated for biparental mapping populations may be restricted to that specific population and are often not transferable to other populations, which limits the application of the markers in a marker-assisted selection program (Holland, 2007; Würschum, 2012).

As discussed, developing markers from QTL mapping and their subsequent validation is highly beneficial for breeding programs. We hypothesize that QTL mapping populations can be developed by crossing parents from germplasm already present in Brazil, allowing us to identify QTLs related to agronomic traits of great importance for the crop that can be utilized in crop breeding in Brazil. Although we do not have multiple populations to map QTLs, the use of three parents to generate the mapping population can increase the power to detect QTLs in sesame. Thus, this work aimed to develop two segregating F2 populations and identify QTLs for traits of agronomic interest in sesame.

3.2. Material and methods

3.2.1. Plant material

The mapping populations were developed from crosses between contrasting sesame lines, resulting in two populations with a common female parent. The first population (SD) was derived from the cross between the commercial cultivar BRS Seda and line DMDU1(BRS Seda x DMDU1). The second population (SJ) was obtained from the cross between the cultivar BRS Seda and line JMDU1 (BRS Seda x JMDU1). The crosses were performed in a greenhouse at the Sebra Agrícola SA research department, located in Canarana-MT, and the F1 plants were subjected to self-fertilization to generate F2 populations.

The commercial variety BRS Seda, developed by the Empresa Brasileira de Pesquisa Agropecuária (Embrapa), features grains with a sweet flavor, an indeterminate growth habit, tall stature, high production potential, branching, an alternate arrangement of capsules, low trichome presence on capsules and stem, and no placental adherence. In contrast, the genotypes DMDU1 and JMDU1, from the Sebra Agrícola SA breeding, program, have grains with a bitter flavor, determinate growth, short stature, short internodes, low branching, high trichome presence on capsules and stem, and placental adherence. They differ in capsule arrangement, with capsules arranged oppositely in the JMDU1 line and alternately in the DMDU1 line, along with some subtle differences related to leaves.

3.2.2. Phenotypic evaluation and statistical analysis

In the F2 generation, 98 plants from the SD cross and 160 plants from the SJ cross were grown, along with two plants from each parent in the greenhouse. Each plant was treated as an experimental plot without replication and was subjected to morphological analyses for the subsequent QTL mapping analysis. The traits evaluated were (i) quantitative: number of capsules per plant (CPP), capsule length (CLA, in cm), plant height (PH, in cm), internode length (ILA, in cm), height to the first capsule (HFC, in cm), number of seeds per capsule (SCA), and period up to 50% flowering (FPF, in days), and (ii) categorical: growth habit (GH), internode number to the first capsule (INFC), primary (PBN) and secondary branching (SBN), stem color (SC), capsule hairiness (CH), stem hairiness (SH), capsule insertion position (SI), placental adherence of grains (PA), capsule opening (CO) (Table 2.1).

Table 2.1. List of categorical variables evaluated to characterize the F2 SD and SJ populations, and the respective classes evaluated for each.

Traits	Type	Classes measured	Score
Growth habit	Binary	Indeterminated	0
		Determinated	1
Internode number to first capsule	Counting		-
Primary branch Number	Counting		-
Secondary branch Number	Counting		-
Stem color	Multicategorical	Green	1
		Yellow green	2
		Yellow	3
Capsule hairiness	Multicategorical	Glabra	0
		Weak or sparse	3
		Median	5
		Strong or abundant	7
Stem hairiness	Multicategorical	Hairless	0
		Weak or sparse	3
		Median	5
		Strong or abundant	7
Capsule insertion position	Binary	Alternate	0
		Opposite	1
Placental adherence of grains	Multicategorical	No adherence	0
		Weak	3

		Median	5
		Strong	7
Capsular openning	Multicategorical	Open from tip to base	0
		Open to mid-length	3
		Open from the tip to the upper third	5
		Slightly openned	7

The data obtained from the morphological analyses were tabulated, and an exploratory analysis was conducted. Initially, histograms were plotted to show the variation among the genotypes in each population and among the parental genotypes. Also, basic statistics, such as mean, standard deviation error, and variation coefficient (%), were calculated for each population. Subsequently, principal component analysis (PCA) was performed to study the genetic structure among the progenies. The correlation between traits was examined using nonparametric analysis with Spearman's correlation at 5% confidence.

3.2.3. Genotyping essays and bioinformatics analysis

For genotyping via SNP markers, individual leaves from each F2 plant and those from the parents were collected and sent for DArT-seq genotyping (DArT Pty Ltd, Canberra, Australia). This method uses methylation filtration and next-generation sequencing technologies (Kilian et al., 2012; Sansaloni et al., 2011). A total of 264 genotypes were genotyped, including duplicates of the parental genotypes.

The Bowtie2 program (Langmead & Salzberg, 2012) was used to align the reads to the reference genome (Wang et al., 2022). Then, variant calling was performed using FreeBayes (Garrison & Marth, 2012). Biallelic variants with high quality were selected, and genotypes were filtered to remove duplicates and keep genotypes with a minimal call rate of 0.8. In addition, SNPs were filtered to a minimum call rate of 0.8 and a minimum allele frequency of 5%.

3.2.4. QTL mapping

Single marker analysis methodology was used to verify the association of the set of SNPs with the morphological traits. To perform a joint analysis, we used a model as described below.:

$$y_{ij} = \mu + p_j + \beta x_i + e_{ij}$$

where y_{ij}: is the phenotypic value of the i-th individual in the j-th population; μ: is the general mean; p_j: is the population effect; β is the marker effect, x_i: is the indicator variable of individual marker genotype (0=SS, 1=SB, 2=BB); and e_{ij}: is the residual error.

The significance of the markers, i.e. β effect, was assessed via likelihood ratio tests using the R package R/qtl2 (Broman et al., 2019) and expressed as logarithm-of-the-odds (LOD) score. QTLs were declared when LOD scores were greater than three (LOD>3) (Moskvina & Schmidt, 2008). The QTLs were defined in terms of the most significant SNP within a 20 Mbp window. For the identified QTLs, the dominance effect of these markers was also estimated using a square matrix of contrast, as described in the user guide of the R/qtl2 package.

3.3. Results

3.3.1. Phenotypic characterization of the F2 population

The F2 progenies displayed high variability across all evaluated traits, as evidenced by the distribution of the bars along the x-axis at the histograms (Figure 2.1), and reflects the genetic segregation derived from the cross carried out. The population SD (pink bars) exhibit higher variation than SJ (blue bars) for most traits, especially for CLA, PH, and ILA.

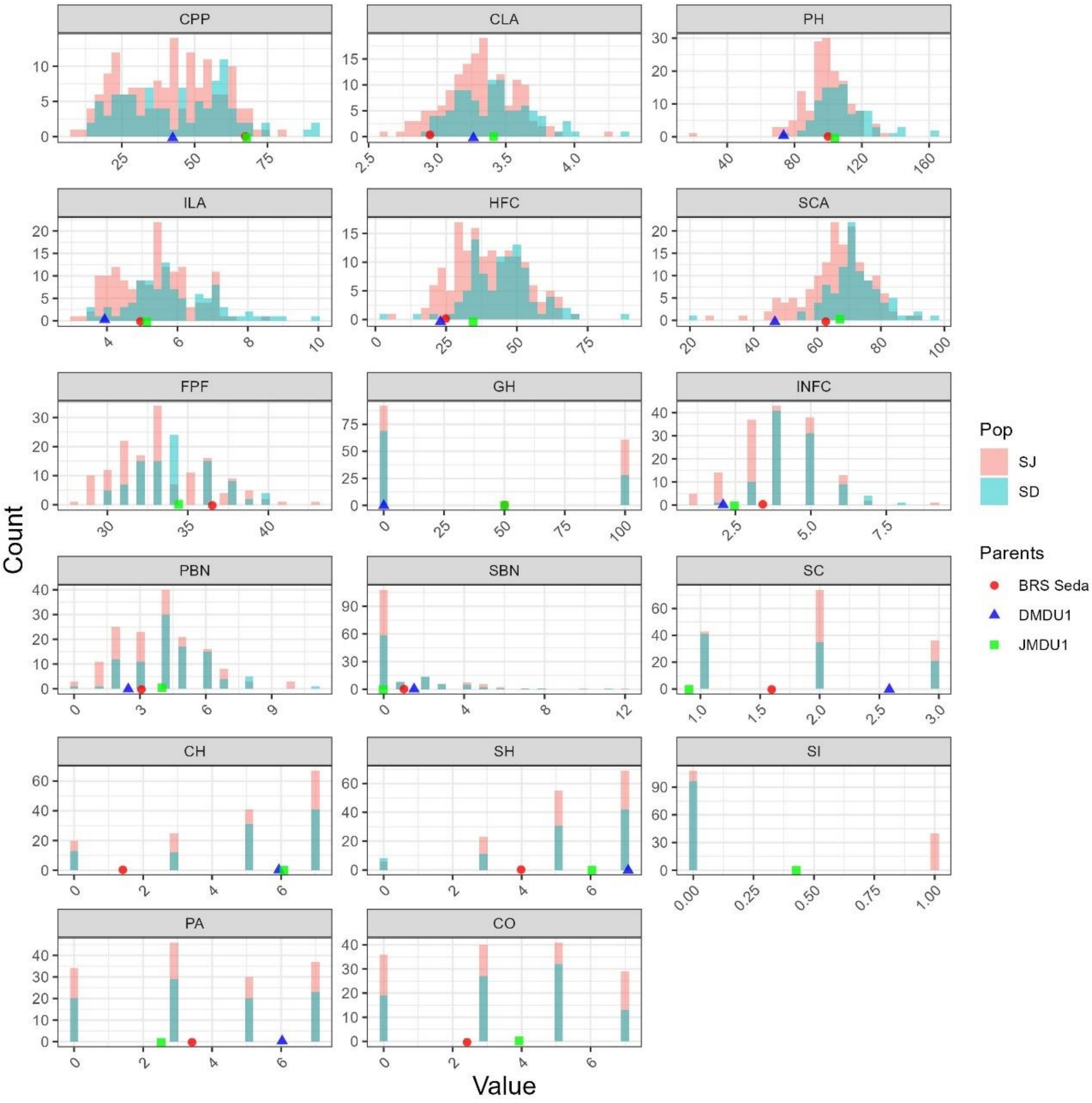

Figure 2.1. Phenotypic variability of the two F2 progenies. The bar plots represent the frequency distribution of the different phenotypic classes for the traits evaluated for the population SD e SJ. The variation for each trait in the population's SD, SJ e parents (BRS Seda, DMDU1, and JMDU1) is shown along the x-axis, where the different colored dots highlight the values for the parents.

There seems to be a tendency toward higher means for all quantitative traits on the population SD, although it would not be possible to assure its superiority with the higher standard errors associated with the estimated means. A higher variation coefficient (CV) is observed for the traits, up to 40% for the CPP trait, reinforcing the presence of variability between the F2 genotypes (Table 2.2).

Table 2.2. Descriptive analysis (mean, standard deviation, and variation coefficient (CV%)) for the quantitative traits analyzed in each population.

	Population SD			**Population SJ**		
	Mean	**Standard deviation**	**CV (%)**	**Mean**	**Standard deviation**	**CV (%)**
CPP	44.11	18.03	40.87	42.68	16.47	38.58
CLA	3.41	0.27	7.87	3.31	0.26	7.8
PH	108.31	15.88	14.67	98.74	13.9	14.08
ILA	5.9	1.22	20.7	5.3	1.05	19.78
HFC	45.09	11.61	25.74	39.2	12.27	31.3
SCA	72.03	9.61	13.35	66.15	9.85	14.89
FPF	34.02	2.54	7.48	33.49	3.05	9.1

Although variation was observed among the genotypes in all qualitative traits evaluated, no clear differences were observed among the populations. In most cases, the frequency distribution of the classes was similar. However, for some traits, such as INFC, SBN, and SI, a higher frequency of a specific class was observed, suggesting possible dominance effects or measurement errors. Phenotypic variation was not observed in the population SD for the trait stem insertion, and only two phenotypic classes were observed in SJ. Higher class frequencies are observed in the SJ population for all the traits, which is already expected since the number of genotypes in the SJ population is considerably higher than in the SD.

Clear phenotypic variability was observed among the parents. The highest differences are noted between the female genotype BRS Seda (red marker) and the male genotypes JMDU1 (green marker) and DMDU1 (blue marker) (Figure 1). Low variability is observed between DMDU1 and JMDU1 for most traits, highlighting the close relationship between these two genotypes. All parents have shown phenotypic values close to each other for the traits CLA, PH, ILA, SCA, PBN, and SBN. However, the progenies still vary, indicating the presence of genetic variability (such as dominance) or environmental effects acting to generate variability.

The principal component analysis of phenotypic values (Figure 2.2) plot shows high dispersion of the F2 genotypes, indicating considerable divergence between them. Besides SD and SJ being distinct populations, it is possible to observe that the two populations were close to each other, as shown by the overlapping of the ellipses of each population.

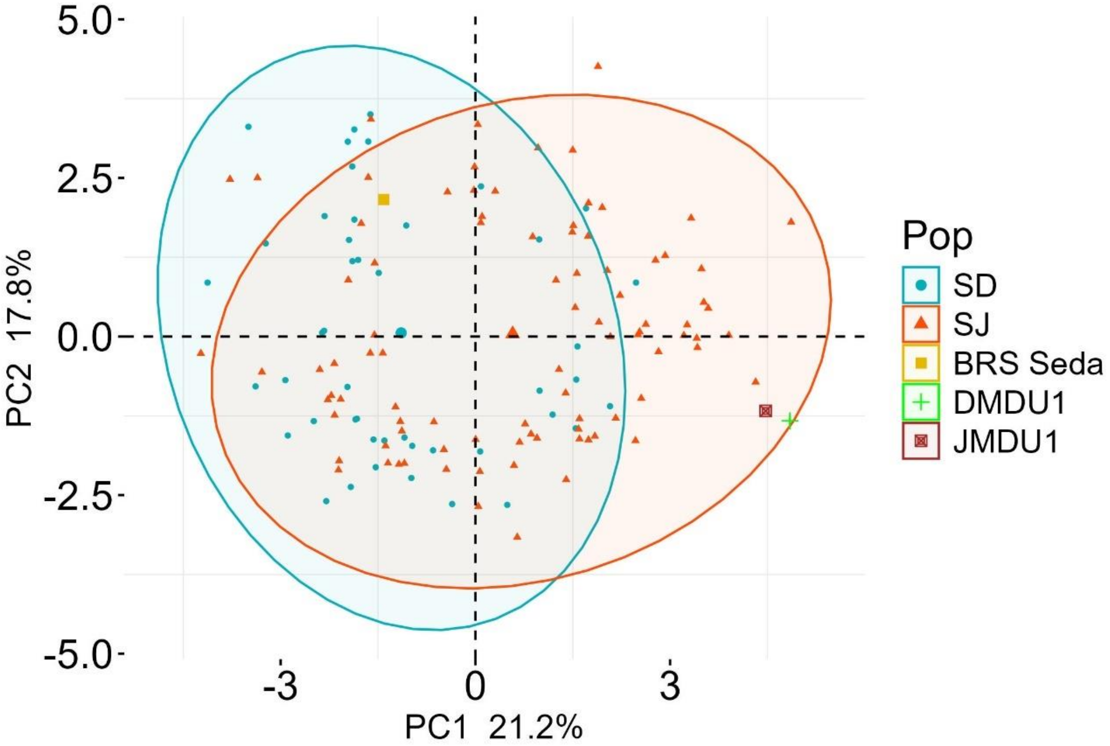

Figure 2.2. Principal component analysis carried out on the 17 phenotypic traits evaluated on the progenies SD and SJ, along with their parents BRS Seda, DMDU1, and JMDU1**.** The ellipses show at least 95% of the genotypes within each respective population.

A relatively low correlation was observed between the phenotypic traits evaluated in the F2 progenies (Figure 2.3), with the significant correlations ($P > 0.05$) listed below. The highest correlation was found between stem hairiness (SH) and capsule hairiness (CH) (0.87), followed by placental adherence (PA) and capsule opening (CO) (0.75). Height to the first capsule (HFC) showed a moderate correlation with plant height (PH) (0.56), days to 50% flowering (FPF) (0.46), and internode number to the first capsule (INFC) (0.54). Although weak, days to 50% flowering (FPF) also showed a correlation with the internode number to the first capsule (INFC) (0.32).

A negative correlation was observed between height to the first capsule (HFC) and the number of capsules per plant (CPP) (-0.36), indicating that plants that begin flowering at lower heights tend to produce more capsules per plant. Additionally, the number of capsules per plant (CPP) was significantly correlated with the number of primary and secondary branches (0.36 and 0.30, respectively).

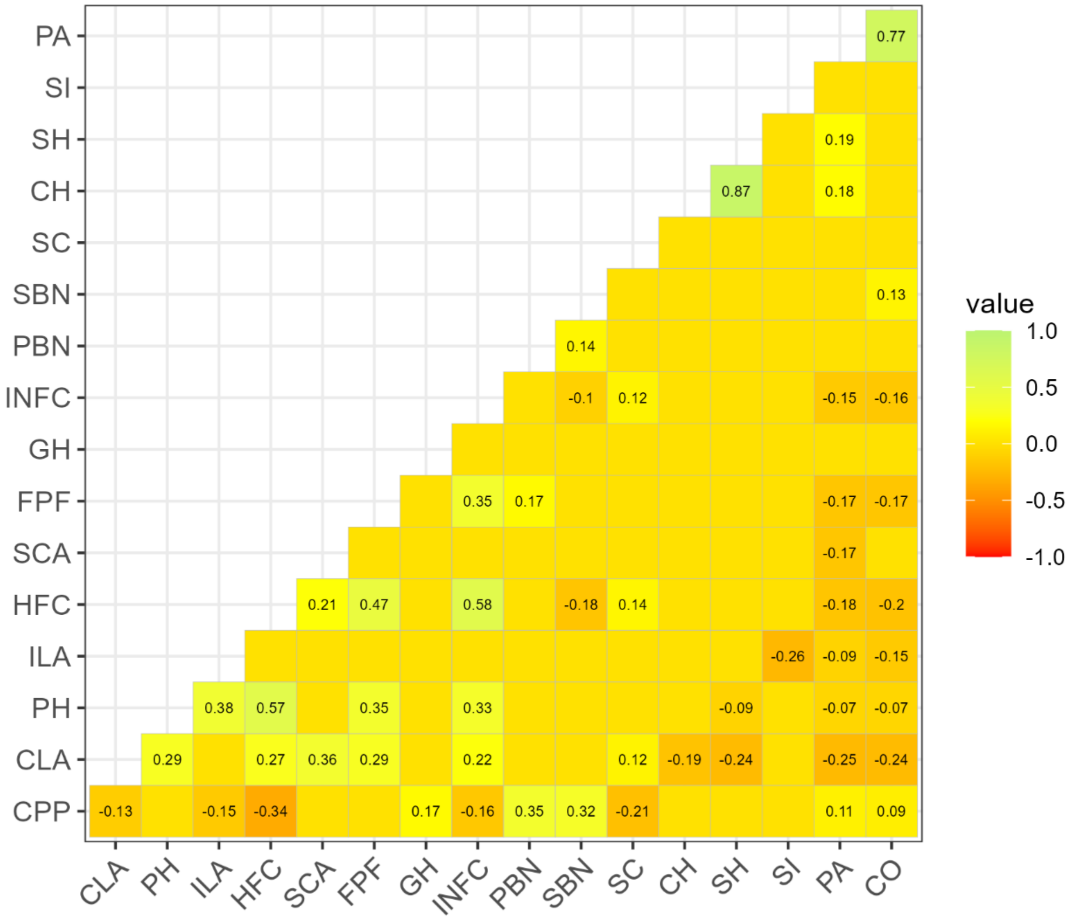

Figure 2.3. Correlation analysis using Spearman's method among phenotypic traits evaluated in F2 populations. The empty windows mean the non-significative correlations (p-value > 0,05). The quantitative traits: number of capsules per plant (CPP), capsule length (CLA, in cm), plant height (PH, in cm), internode length (ILA, in cm), height to the first capsule (HFC, in cm), number of seeds per capsule (SCA), and period up to 50% flowering (FPF, in days); and categorical: growth habit (GH), internode number to the first capsule (INFC), primary (PBN) and secondary branching (SBN), stem color (SC), capsule hairiness (CH), stem hairiness (SH), capsule insertion position (SI), placental adherence of grains (PA), capsule opening (CO).

3.3.2. Genotyping, SNP mining, and QTL mapping

A total of 33,057 SNPs were obtained from variant calling data. Seven genotypes showed a call rate lower than 0.8 and were removed. Similarly, 15,121 SNPs showed a call rate lower than 0.8 and were removed. Regarding the minimum allele frequency (MAF), 10,439 variants showed MAF lower than 5% and were removed. Ultimately, 7,497 SNP markers were obtained for QTL mapping in the two sesame populations. It was observed that the 7,497 SNPs were relatively well distributed along the 13 sesame chromosomes, ranging from 270 (chromosome 4) to 1,185 (chromosome 3) SNPs (Table 2.3).

Table 2.3. Distribution of the SNPs along the sesame chromosomes.

Chromosome Code	Chromosome	SNP Number/ Chromosome	Percentage (%)
1	CM047432.1	684	9.12
2	CM047433.1	382	5.1
3	CM047434.1	1,185	15.81
4	CM047435.1	270	3.6
5	CM047436.1	334	4.46
6	CM047437.1	647	8.63
7	CM047438.1	396	5.28
8	CM047439.1	737	9.83
9	CM047440.1	901	12.02
10	CM047441.1	432	5.76
11	CM047442.1	576	7.68
12	CM047443.1	628	8.38
13	CM047444.1	323	4.31

Fourteen significant markers were identified for 14 of the 16 traits initially analyzed, considering both progenies together (Figure 2.4) since there were no significant markers for the traits' primary branch number and stem color. The capsule insertion position trait was just analyzed in the SJ population. For most traits, only one region showed significant markers. However, the traits capsule per plant, growth habit, capsule length average, plant height, internode length average, and height to the first capsule each had more than one significant region, ranging from two to five significant regions.

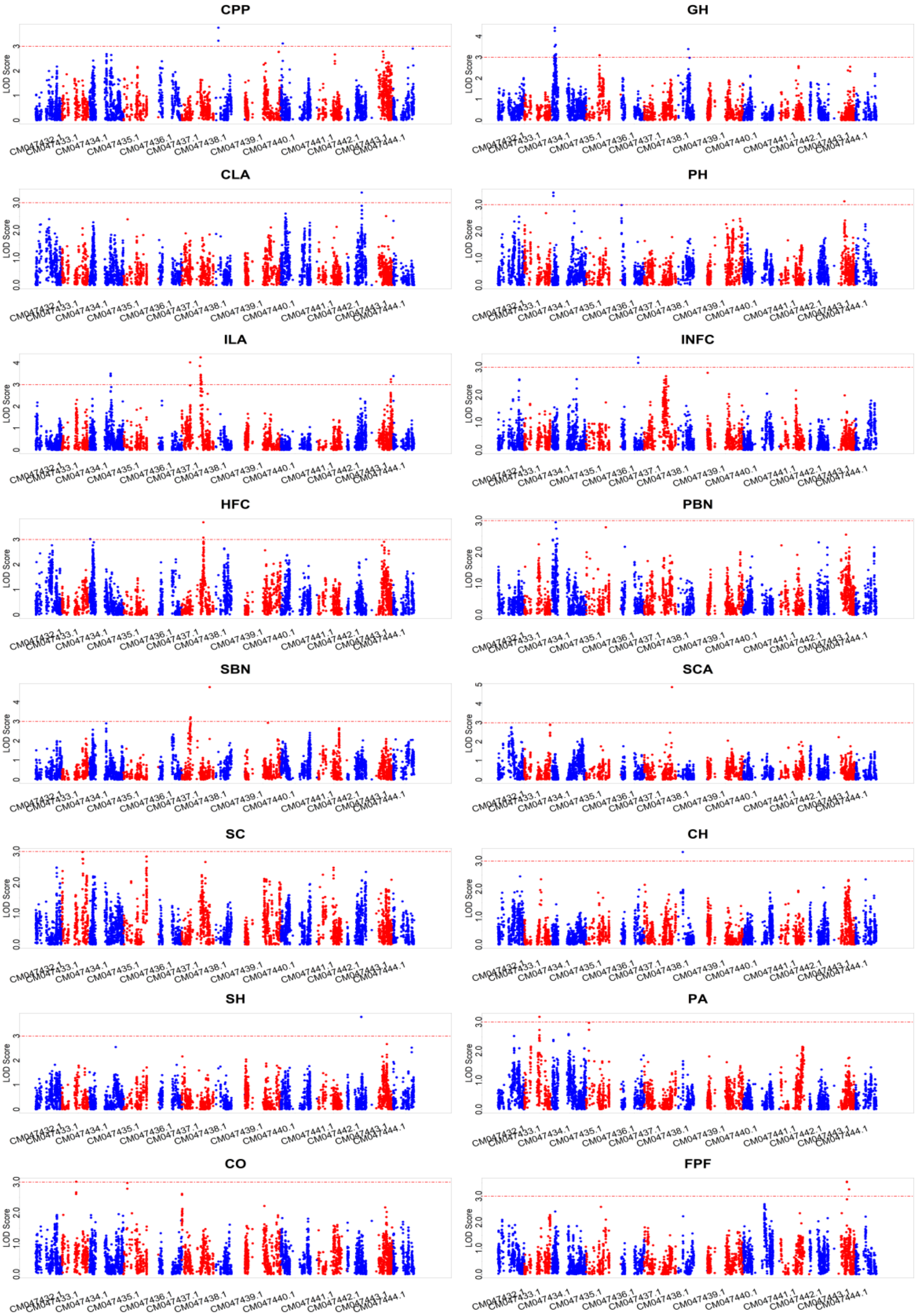

Figure 2.4. Manhattan plots display the LOD scores for each SNP along the chromosomes for the two populations. The dashed red line represents the significance threshold LOD > 3. Points above the dashed red line indicate significant SNPs. The quantitative traits: number of capsules per plant (CPP), capsule length (CLA, in cm),

plant height (PH, in cm), internode length (ILA, in cm), height to the first capsule (HFC, in cm), number of seeds per capsule (SCA), and period up to 50% flowering (FPF, in days); and categorical: growth habit (GH), internode number to the first capsule (INFC), primary (PBN) and secondary branching(SBN), stem color (SC), capsule hairiness (CH), stem hairiness (SH), placental adherence of grains (PA), capsule opening (CO).

Regions with more than one significant SNP were identified for most traits, except for capsule hairiness, stem hairiness, placental adherence, capsule opening, capsule length average, seeds per capsule, and 50% flowering, and the window size of the QTL was considered (Table 2.4). The percentage of variance explained by the markers ranged from 5.69% to 15.37%. The markers explaining the lowest percentage of phenotypic variance were associated with capsule opening (5.69%), capsule per plant (5.89%), and placental adherence and capsule hairiness (5.94%). The highest values were observed in the QTLs for number of seeds per capsule (15.37%) and plant height, ranging from 14.85% to 15.36% (Table 2.4).

Table 2.4. Significant markers in the joint analyses, the genotypic effects, and the percentage of variance explained for each marker.

Traits	Chromossome	Position (Mpb)	Lod Score	Interval (Mpb)	Mean	Additive effect (a)	Dominance effect (d)	Explained QTL variation (%)
Number of capsules per plant	CM047438.1	44.799	3.758	34.799 - 54.799	42.617	-1.893	5.978	6.992
Number of capsules per plant	CM047440.1	16.884	3.116	6.884 - 26.884	43.492	1.223	5.75	5.886
Capsule length	CM047442.1	132.295	3.372	122.295 - 142.295	3.333	0.014	-0.077	9.443
Growth habit	CM047434.1	34.762	4.409	24.762 - 44.762	0.388	0.026	0.24	6.309
Growth habit	CM047435.1	122.63	3.099	112.630 - 132.630	0.334	-0.174	0.098	4.435
Growth habit	CM047438.1	111.187	3.39	101.187 - 121.187	0.372	-0.094	0.128	4.851
Plant height	CM047434.1	22.118	3.45	12.118 - 32.118	11058284.43	16587269.7	-11058184	15.36
Plant height	CM047443.1	90.437	3.123	80.437 - 100.437	103.048	2.681	-3.778	14.849
Internode length	CM047434.1	196.618	3.509	186.618 - 206.618	5.477	0.354	-0.19	12.27
Internode length	CM047437.1	73.6	4.026	63.600 - 83.600	5.74	-0.028	0.747	13.112
Internode length	CM047437.1	166.41	4.254	156.410 - 176.410	5.453	-0.145	0.356	13.48
Internode length	CM047443.1	169.026	3.25	159.026 - 179.026	5.538	0.019	-0.324	11.845
Internode length	CM047444.1	5.891	3.399	-19.999	2.37	-11.343	3.131	12.09
Internode number to first capsule	CM047436.1	155.33	3.353	145.330 - 165.330	4.089	-0.502	-0.057	10.963
Height to first capsule	CM047434.1	15.28	3.019	5.280 - 25.280	40.756	3.137	-1.806	9.621
Height to first capsule	CM047437.1	191.264	3.689	181.264 - 201.264	40.998	1.898	-3.598	10.731
Secondary branch number	CM047437.1	77.103	3.222	67.103 - 87.103	1.308	0.26	0.705	6.069
Secondary branch number	CM047437.1	246.749	4.775	236.749 - 256.749	6594.32	-0.262	13186.991	8.718
Number of seeds per capsule	CM047437.1	245.835	4.872	235.835 - 255.835	-21750230.3	-3.302	-43500592	15.371
Capsule Hairiness	CM047438.1	62.667	3.322	52.667 - 72.667	5.033	0.299	-0.743	5.939
Stem Hairiness	CM047442.1	128.838	3.781	118.838 - 138.838	-5.622	-16.518	10.793	7.109
Placental Adherence of grain	CM047433.1	135.48	3.181	125.480 - 145.480	4.101	0.712	0.106	5.944
Capsular Opening	CM047433.1	128.366	3.01	118.366 - 138.366	3.32	0.155	-1.317	5.685
Period up to 50% flowering	CM047443.1	113.021	3.556	103.021 - 123.021	33.744	0.837	0.286	7.184

Most identified QTLs exhibited additive and dominance effects (Table 2.4). A predominantly additive effect was found only in the QTL on chromosome CM047436.1, position 155.33 for the trait internode number to the first capsule (a = -0.502; d = -0.057). On the other hand, predominantly dominant effects were observed in the QTLs for the traits secondary branch number, plant height, and internode length. Indeed, the trait internode length is associated with five QTLs, two of which (chromosome CM047437.1, position 73.60 Mbp, and chromosome CM047443.1, position 169.03 Mbp) presented a stronger dominance effect over the additive effect. Three others (chromosome CM047434.1, position 196.618 Mbp, chromosome CM047437.1, position 166.410 Mbp, and chromosome CM047444.1, position 5.891 Mbp) presented both additive and dominance effects. In general, the QTLs identified for the remaining traits displayed both additive and dominance effects, showing that controlling these characters is not simple.

The population SD individually presented 29 significant QTLs that were identified across 12 of the 16 traits analyzed (Figure 2.5). In this population, no QTL was found for the number of capsules per plant (CPP), growth habit (GH), primary branch number (PBN), and stem color (SC). Most of the significant markers in the SD population were in genomic regions different from those in the joint analysis (Table 2.5). An exception was observed for traits capsule length (CLA), internode length (ILA), and number of seeds per capsule (SCA), which shared regions of significant QTLs in both analyses. For CLA, the shared QTL was found in the chromosome CM047442.1, at the window 122.295 – 142.295 Mpb; for ILA, it was found in the chromosome CM047437.1, at the window 63.600 – 83.600 Mpb; and for SCA it was found in chromosome CM047437.1, at the window 235.835 – 255.835 Mpb.

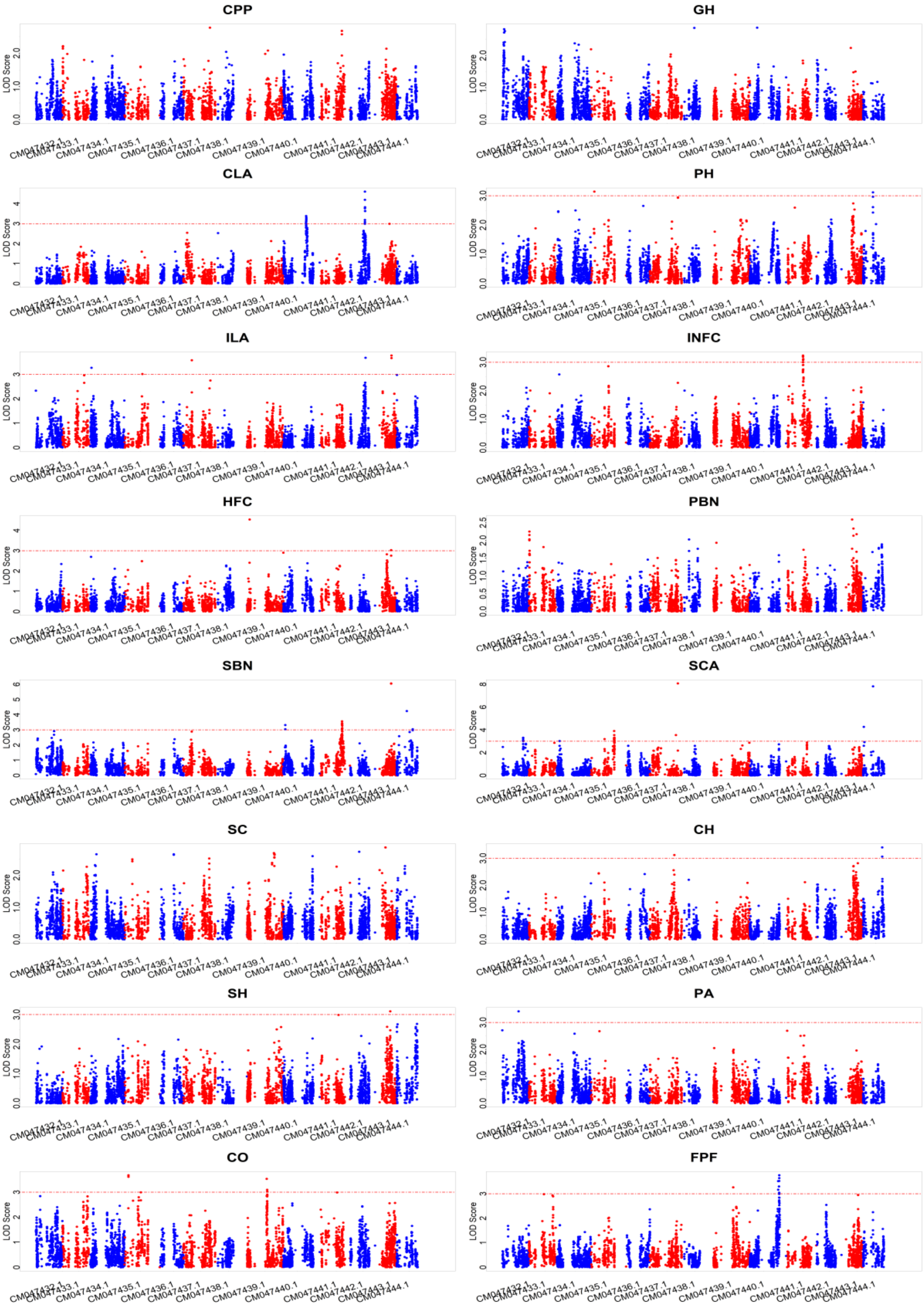

Figure 2.5. Manhattan plots showing the LOD scores for each SNP along the chromosomes for the SD population. The dashed red line indicates the threshold to consider significative SNPs. The points above the dashed red line indicate significative SNPs for the trait under study. The quantitative traits: number of capsules per plant

(CPP), capsule length (CLA, in cm), plant height (PH, in cm), internode length (ILA, in cm), height to the first capsule (HFC, in cm), number of seeds per capsule (SCA), and period up to 50% flowering (FPF, in days); and categorical: growth habit (GH), internode number to the first capsule (INFC), primary (PBN) and secondary branching(SBN), stem color (SC), capsule hairiness (CH), stem hairiness (SH) placental adherence of grains (PA), capsule opening (CO).

Table 2.5. Significant QTL found for the traits in the SD population, with its position, genetic effect, and explained trait variation.

Traits	Chromosome	Position (Mpb)	Lod Score	Interval (Mpb)	Mean	Additive effect (a)	Dominance effect (d)	Explained variation (%)
Capsule length	CM047440.1	200.689	3.385	190.689 - 210.689	3.427	-0.077	-0.111	15.588
Capsule length	CM047442.1	132.295	4.608	122.295 - 142.295	3.438	0.098	-0.158	20.598
Plant height	CM047435.1	30.373	3.14	20.373 - 40.373	115.121	-13.547	-8.762	14.401
Plant height	CM047444.1	90.216	3.116	80.216 - 100.216	106.493	-5.394	-5.11	14.296
Internode length	CM047434.1	19.502	3.27	9.502 - 29.502	5.903	-0.196	-0.619	15.098
Internode length	CM047435.1	157.294	3.015	147.294- 167.294	5.924	0.32	0.55	14.009
Internode length	CM047437.1	73.6	3.58	63.600- 83.600	5.974	-0.139	0.729	16.406
Internode length	CM047442.1	137.852	3.682	127.852 - 147.852	5.888	-0.297	-0.649	16.833
Internode length	CM047443.1	144.226	3.773	134.226 - 154.226	5.733	0.573	0.654	17.209
Internode number to first capsule	CM047441.1	148.324	3.236	138.324 - 158.324	4.551	0.365	-0.383	14.806
Height to first capsule	CM047439.1	23.323	4.543	13.323 - 33.323	90.558	-7.855	78.646	20.143
Height to first capsule	CM047443.1	141.316	3.027	131.316 - 151.316	1349909.347	-2024887.668	-1350046	13.921
Secondary branch number	CM047440.1	16.871	3.32	6.871 - 26.871	1.251	-0.079	1.204	15.159
Secondary branch number	CM047441.1	196.395	3.566	186.395 - 206.395	1.49	0.893	-0.902	16.185
Secondary branch number	CM047443.1	141.497	6.059	131.497 - 151.497	-14875.597	22350.061	14946.582	25.921
Secondary branch number	CM047444.1	92.862	4.242	82.862 - 102.862	4.507	0.309	6.302	18.946
Number of seeds per capsule	CM047432.1	187.77	3.317	177.770 - 197.770	63.62	-0.125	-17.634	15.298
Number of seeds per capsule	CM047434.1	34.116	3.02	24.116 - 44.116	69.648	-6.558	7.434	14.031
Number of seeds per capsule	CM047435.1	204.812	3.901	194.812 - 214.812	70.972	-4.475	3.709	17.739
Number of seeds per capsule	CM047437.1	245.835	8.077	235.835 - 255.835	-37979247.22	0	-75958640	31.585
Number of seeds per capsule	CM047444.1	90.724	7.82	80.724 - 100.724	55.979	0.766	-34.866	32.391
Capsule Hairiness	CM047437.1	214.98	3.121	204.980 - 224.980	4.7	0.947	0.841	14.319
Capsule Hairiness	CM047444.1	171.618	3.404	161.618 - 181.618	4.711	-0.014	-1.489	15.511
Stem Hairiness	CM047443.1	133.363	3.11	123.363 - 143.363	5.378	-0.328	-1.042	15.02

Significant QTLs were found for 13 of the 17 traits analyzed for the SJ population, with 26 QTLs identified (Figure 2.6). Contrasting with the previous analysis, one QTL was found for the primary branch number (PBN) in this population. Still, like in the previous analysis, no QTL was found for stem color. No QTL was found for the trait capsule insertion position (SI), which was analyzed only in the SJ population. No QTL was found for Capsule Hairiness (CH) and Placental Adherence (PA). In this population, from the 26 QTLs identified, six of them also presented an association in the joint analysis of the two populations. The shared region was found for the number of capsules per plant (chromosome CM047438.1, window 34.799 - 54.799 Mbp), growth habit (chromosome CM047434.1, window 24.762 - 44.762 Mbp), internode length (chromosome CM047437.1, window 156.410 - 169.410 Mbp), secondary branch number (chromosome CM047437.1, window 236.749 - 256.749 Mbp) and stem hairiness (chromosome CM047442.1, window 118.838 - 138.838 Mbp). The QTL on chromosome CM047437.1, window 183.617 - 203.617 Mbp for height to the first capsule was also identified in the population SD. The QTLs for internode length (chromosome CM047437.1, positions 166.410 and 169.445) were located in the same window size, but the first one show a dominance effect of 0.372, and additive effect near zero, but in the second one, the opposite is observed, where the additive effect is predominant (-0.158) (Table 2.6), and the dominance effect was near zero, indicating that this two markers might have importance in the control of the trait.

Another important aspect of QTL mapping is the identification of regions that control more than one trait. The QTL for the number of capsules per plant at chromosome CM047440.1, position 16.884 in the joint analysis, was also identified for the secondary branch number in the population SD. The same is observed for the QTL at chromosome CM047442.1, position 132.295, for capsule length in joint analysis, which was also identified for stem hairiness in joint analysis and the SJ population. The QTL for growth habit CM047435.1 at position 154.142 was also identified for internode length at position 157.294. The same QTL for growth habit in the joint analysis was found for plant height (CM047434.1, positions 34.762 and 22.118, in the joint analysis and SD population, respectively). At this same region, QTL for capsule opening (position 22.531 in SJ population, and 34.779 in SD population), height to the first capsule (position 15.280 in joint analysis), and number of seeds per capsule (position 34.116) was found. At chromosome CM047434.1, the QTL for Internode length (position 196.618) in the joint analysis shared the same region with the QTL found for the number of capsules per plant (position 201.737) in the SJ population. A shared region of QTL was also found for the internode number to the first capsule (INFC) (CM047437.1, position 193.911) in the SJ population and height to the first capsule (HFC) (CM047437.1, position 193.617 in SJ, and 191.263 in joint analysis). In the population SJ the same QTL (CM047443.1, position 150.424) was found for number of capsules per plant and height to first capsule.

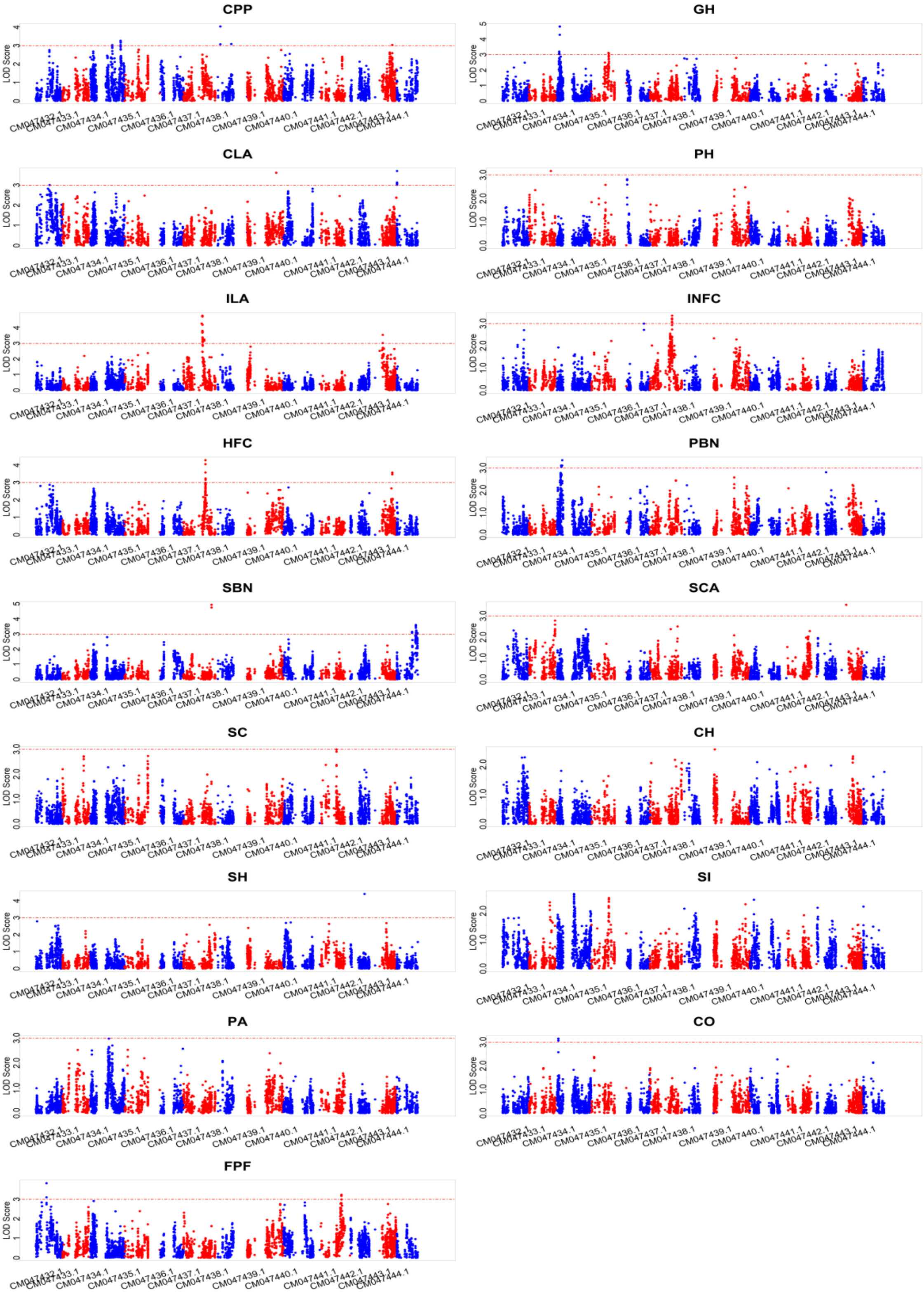

Figure 2.6. Manhattan plots showing the LOD scores for each SNP along the chromosomes for population 2. The dashed red line indicates the threshold to consider significative SNPs. The point above the dashed red line indicates significative SNPs for the trait under study.

Table 2.6. Significant QTL found for the traits in the SJ population, with its position, genetic effect, and explained variation on the trait.

Traits	Chromosome	Position (Mbp)	Lod Score	Range (Mpb)	Mean	Additive effect (a)	Dominance effect (d)	Explained variation (%)
Number of capsules per plant	CM047434.1	201.737	3.028	191.736 - 211.736	45.052	1.692	2.199	8.498
Number of capsules per plant	CM047434.1	275.603	3.267	265.602 - 285.602	44.559	0.279	-3.334	9.138
Number of capsules per plant	CM047438.1	44.799	4.041	34.798 - 54.798	43.729	0.577	4.542	11.177
Number of capsules per plant	CM047438.1	141.026	3.09	131.025 - 151.025	44.385	2.673	0.119	8.664
Number of capsules per plant	CM047443.1	150.424	3.035	140.424 - 160.424	44.591	-0.176	1.176	8.518
Growth habit	CM047434.1	34.762	4.832	24.762 - 44.762	0.315	0.029	0.098	6.31
Growth habit	CM047435.1	154.142	3.114	144.141 - 164.141	0.298	-0.073	-0.095	2.81
Capsule length	CM047432.1	125.028	3.005	115.027 - 135.027	3.402	0.029	0.047	8.488
Capsule length	CM047439.1	257.29	3.614	247.289 - 267.289	3.398	-0.01	-0.024	10.119
Capsule length	CM047444.1	5.9	3.7	0 - 20	3.217	0.351	0.193	10.348
Plant height	CM047433.1	193.816	3.165	183.816 - 203.816	206.727	164.815	-133.161	8.865
Internode length	CM047437.1	166.41	4.776	156.409 - 176.409	5.823	0.05	0.372	13.23
Internode length	CM047437.1	182.864	3.246	172.863 - 192.863	5.908	-0.158	-0.084	9.194
Internode length	CM047443.1	67.205	3.541	57.204 - 77.204	5.885	-0.26	0.039	9.986
Internode number to first capsule	CM047437.1	193.911	3.358	183.911 - 203.911	4.466	0.119	0.128	9.379
Height to first capsule	CM047437.1	193.617	4.281	183.617 - 203.617	44.631	0.119	-0.194	11.801
Height to first capsule	CM047443.1	150.425	3.57	140.425 - 160.425	44.347	0.487	0.194	9.942
Primary branch number	CM047434.1	56.266	3.344	46.265 - 66.265	4.352	-0.228	0.245	9.342
Secondary branch number	CM047437.1	246.749	4.935	236.748 - 256.748	1977853.273	0	3955704.12	13.475
Secondary branch number	CM047444.1	138.446	3.148	128.446 - 148.446	1.241	-0.394	-0.597	8.821
Secondary branch number	CM047444.1	171.429	3.603	161.429 - 181.429	1.22	-0.45	0.058	10.029
Number of seeds per capsule	CM047443.1	40.258	3.526	30.258 - 50.258	71.979	-0.034	-0.039	9.827
Stem Hairiness	CM047442.1	128.838	4.401	118.837 - 138.837	216.36	316.548	-211.342	12.11
Capsular Opening	CM047434.1	22.531	3.149	12.531 - 32.531	3.71	-0.103	-0.112	9.214
Period up to 50% flowering	CM047432.1	98.199	3.825	88.198 - 108.198	33.987	-0.624	-0.187	10.742
Period up to 50% flowering	CM047441.1	191.088	3.238	181.088 - 201.088	34.078	0.176	-0.633	9.171

In both populations, SD and SJ, genetic effects varied significantly, with the additive and dominant simultaneously being predominant, following the same pattern identified in the joint analysis (Tables 3, 4, and 5). In population SD, none of the QTLs identified exhibited a predominantly additive effect. In contrast, of the two QTLs found for CLA, one of the four found for SBN, two of the three found for SCA, and one of the two found for CH exhibited predominantly dominant effects. The mean percentage of variance explained by the markers was higher in the separate analyses than in the joint analyses. In the SD population, the mean-variance explained by the markers was 17.55% (Table 4), while in SJ, it was 9.58% (Table 5), whereas, in the joint analysis, it was 9.34% (Table 3). The highest values of explained variance in SD were found for SCA, up to 32.39%, while the lowest values were found for height to the first capsule (13.92%). In SJ, the highest value was 13.48% in QTL for the secondary branch number trait, while the lowest value was 2.81% for growth habit.

3.4. Discussion

Genetic variability is crucial for any genetic association method, including QTL mapping, regardless of the type of population used. The two segregating F2 progenies used here exhibited high phenotypic variability for almost all traits analyzed, demonstrating their suitability for QTL mapping. The observed high phenotypic variability in the F2 progenies underscores the effective choice of parents for developing these segregating populations. Differences among the populations were observed, with SD population showing higher means for most quantitative traits, while SJ population exhibited greater phenotypic variability for some traits (Figure 1). This variability may be related to the genetic differences among the founder parents. Although the genetic structure among the populations being very low, identifying these effects and incorporating models that account for them in QTL mapping is essential to avoid missing significant QTLs or detecting false positives (Raghavan et al., 2017; Yamamoto et al., 2014). The limited structure observed in the PCA can be attributed to the genetic similarity between DMDU1 and JMDU1, the male parents of Progenies 1 and 2, respectively, as they share a common ancestor.

Enhancing productivity is the primary goal of sesame breeding programs, along with developing plants suitable for mechanized cultivation. While productivity could not be directly measured in the F2 population, many analyzed traits are well-documented in the literature as highly correlated with productivity. Mei et al. (2021) described that the traits number of capsules per plant, seeds per capsule, and seed weight are the primary determinants of sesame plant production, and the two first ones were analyzed here. In addition to productivity, traits such as placental adherence and growth habit are crucial in contemporary sesame breeding programs because they directly impact the mechanized management of the crop. Developing sesame cultivars with a determinate growth habit is essential for modernizing sesame cultivation (Y. Zhang et al., 2018), as it provides more uniform crops in flowering, capsule maturation (Çağırgan, 2006), shorter plants, and reduces the risk of lodging (Uzun & Çağirgan, 2006). Additionally, placental adhesion is a critical trait for modern sesame cultivars, as it addresses one of the major challenges in mechanized harvesting, where losses due to capsule dehiscence can be significant (Stamatav et al., 2020).

Despite the low correlation observed among the traits used for phenotypic characterization of the F2 progenies, some correlations proved noteworthy from a breeding perspective. Notably, the number of capsules per plant was not positively affected by greater plant height or height to the first capsule. In fact, the opposite effect was observed, suggesting that it is possible to select shorter plants with lower height to the first capsule without compromising the productive potential. These findings align with the observations of Biabani & Pakniyat (2008), who found that the number of capsules per plant, harvest index, thousand-grain weight, and height to the first capsule were key traits affecting sesame productivity. Ibrahim & Khidir (2012) identified the number of capsules per plant as the second most correlated trait with productivity (0.74), while Mei et al. (2021) reported that height to the first capsule was the most negatively correlated (-0.71) with sesame productivity. Other phenotypic traits analyzed in this study, such as the number of seeds per capsule (0.33) (Emamgholizadeh et al., 2015; Ibrahim & Khidir, 2012) and primary branch number (0.52) (Ibrahim & Khidir, 2012), have also been shown to correlate with productivity in literature. Furthermore, the results indicate that plants with a lower height to the first capsule and shorter plant height tend to enter the reproductive period earlier. This trait is particularly important given that sesame is a second-season crop in Brazil, where water availability is often limited (Arvor et al., 2014; Nóia Júnior & Sentelhas, 2019). Therefore, plants that can begin the reproductive period earlier are advantageous. Since productivity is a complex trait influenced by multiple factors, breeders should focus their efforts on QTL regions that have the potential to impact production (Mei et al., 2021).

Pleiotropy and gene linkage are two primary reasons for correlations among different traits (Chebib & Guillaume, 2021; Chen & Lübberstedt, 2010). This study observed that some significant correlations corresponded with the proximity of the markers identified for these traits. For instance, markers significantly associated with placental adherence and capsular opening (correlation of 0.75) were found in a nearby region (135.48 and 128.37 Mbp, respectively) on chromosome CM047433.1 on the joint analysis. Langham (2014) discusses that seed retention within the capsule is influenced by capsule architecture, which also affects capsule opening. The relationship between capsule opening and seed retention is further supported by Stamatav et al. (2020) findings. Additionally, the traits height to the first capsule (HFC) and internode number to the first capsule (INFC), which have significant markers located relatively close to each other on chromosome CM047437.1 (at 193.911and 193.617, respectively, in SJ population), also showed significant correlations. Similarly, on chromosome CM047437.1, markers significantly associated with secondary branch number (at 77.10 Mbp) and internode length average (73.60 Mbp) in the joint analysis were correlated. The traits plant height (PH) and height to the first capsule (HFC) presented a significative correlation of 0.57 and shared a QTL on the chromosome CM047443.1 at 150.424 Mpb.

Conversely, no relationship was observed between markers associated with capsule hairiness and stem hairiness, despite this being the highest correlation coefficient noted, as these markers were located on different chromosomes. Pleiotropic effects have already been reported for various phenotypic traits in sesame, such as dehiscence, shape, and size of the seeds (Mei et al., 2023; Teboul et al., 2022; Weldemichael & Gebremedhn, 2023). It is important to note that several factors, such as population size, marker density, and heritability, can influence the independence between two QTLs (Xiaohong et al., 2001).

Considering the joint and separate analyses, the traits stem color and capsule insertion position did not present any significant QTL. A limited number of stable QTL markers were identified across the mapping analyses. Just three QTLs were found in the SD population, and the joint analysis simultaneously, and six QTLs were identified in the SJ population and in the joint analysis. Also, one QTL was shared between the SJ and SD populations. The variability in QTL detection could be attributed to several factors related to the populations and markers, such as allele frequency and gene effects, which may differ when populations are analyzed together or separately. Moreover, gene interactions intrinsic to the population, such as epistasis, might cause markers detected in individual populations not to be detected in joint analyses or vice versa, a phenomenon known as QTL x background interaction (Bernardo, 2008). Additionally, the statistical power in joint analyses might be greater, thus enhancing the detection of more complex QTLs that could be challenging to identify in smaller or individual populations.

As cited by Bernardo (2008), the studies by Beavis (1994), Melchinger et al. (1998), and Schön et al. (2004) clearly demonstrate the effect of population size on QTL detection. Beavis (1994) found that when conducting QTL mapping analyses with 100 F2:3 maize genotype subsets, only one marker was detected in each subset. In contrast, the complete analysis of 400 F2:3 maize genotypes identified four QTLs. Similarly, Melchinger et al. (1998) observed that of the 31 QTLs detected in 344 F3 maize families, only 6 were found in a subset of 100 F3 families. Schön et al. (2004), while studying QTLs for plant height in a testcross population, identified 30 QTLs when analyzing the full population of 976 maize F2:3 families. However, in subsets of 488, 244, and 122 F2:3 families, an average of 17.6, 12, and 9.1 QTLs were detected, respectively. Thus, a smaller population size can result in fewer QTL detections and biased estimates of QTL effects (Bernardo, 2008).

The number of significant QTLs detected in SD was higher than in SJ (29 and 26 markers, respectively). Notably, in SD, no QTL was detected for the trait capsules by plant, whereas QTLs for this trait were detected in the joint analysis and SJ. Another important observation is that the QTLs common to both SJ and joint analysis also had the highest LOD scores in situations where more than one marker was significant for a trait, underscoring the high effect of these markers on the associated traits.

One of the advantages of QTL mapping studies using F2 populations is the ability to study additive and dominance effects related to markers on the traits under investigation (Zhang et al., 2008). In the present study, dominance effects were highly prevalent across most QTLs identified. Generally, additive effects are more manageable in a breeding program because they represent the proportional change in phenotype caused by substituting one allele for another. In contrast, dominance effects can hide the influence of other alleles within a genotype, thereby reducing the predictability of trait inheritance and complicating the selection process (Wang et al., 2004).

Despite placental adherence and capsular opening traits being in closely related genomic regions and exhibiting a high correlation (r = 0.75), the genetic effects of the associated markers differ significantly in the QTLs detected on the joint analysis. For placental adherence, a greater additive effect was observed (a = 0.712), whereas capsular opening showed a stronger dominance effect (d = -1.317). This suggests that even though these traits are highly correlated, their selection strategies may need to be quite different.

3.5. Conclusion

The present study revealed considerable phenotypic variability and significant genetic diversity between the two F2 populations studied, highlighting the effectiveness of using these populations for QTL mapping in sesame breeding programs. The joint and individual analyses of the populations resulted in identifying significant QTLs for most of the phenotypic traits analyzed, with the predominant presence of additive and dominance effects for the same QTL, varying according to the trait and population. The presence of shared and specific QTLs in different populations suggests a complex interaction among genetic effects, including epistatic and background genetic effects. These results provide valuable information for marker-assisted selection, especially in selecting traits of interest, such as growth habit and placental adherence, which are essential for developing sesame cultivars adapted to mechanized harvesting. Furthermore, identifying genomic regions associated with traits related to yield and management reinforces the potential of QTL mapping to improve sesame yield and adaptability under Brazilian cultivation conditions. Finally, the variability in QTL detection between joint and individual analyses underscores the importance of considering population size and genetic structure when planning QTL mapping studies.

3.6. References

Ahmed, J., Qadir, G., Ansar, M., Wattoo, F. M., Javed, T., Ali, B., Marc, R. A., & Rahimi, M. (2023). Shattering and yield expression of sesame (Sesamum indicum L) genotypes influenced by paclobutrazol concentration under rainfed conditions of Pothwar. *BMC Plant Biology*, *23*(1), 137.

Arriel, N. H. C., ARAUJO, A. E., SOARES, J. J., BELTRÃO, N. E. M., & FIRMINO, P. T. (2006). Cultivo do gergelim. *Sistemas de Produção, Embrapa Algodão*, *06*.

Arriel, N. H. C., Gondim, T. M. de S., Firmino, P. de T., Beltrão, N. E. de M., Vasconcelos, R. A., Costa, I. L. da, Silveira, N. A. da, Sousa, S. L. de, Dantas, E. S. B., & Pereira, J. R. (2009). *Gergelim BRS Seda* (Issue Figura 2).

Arriel, N. H. C., Sousa, S. L. de, Heurt, J., Medeiros, A. A. de, Gondim, T. M. de S., Firmino, P. de T., Vasconcelos, R. A., Costa, I. L. da, & Dantas, E. S. B. (2015). *Gergelim BRS Anahí* (Issue 83).

Arvor, D., Dubreuil, V., Ronchail, J., Simões, M., & Funatsu, B. M. (2014). Spatial patterns of rainfall regimes related to levels of double cropping agriculture systems in Mato Grosso (Brazil). *International Journal of Climatology*, *34*(8).

Asekova, S., Oh, E., Kulkarni, K. P., Siddique, M. I., Lee, M. H., Kim, J. I., Lee, J.-D., Kim, M., Oh, K.-W., & Ha, T.-J. (2021). An integrated approach of QTL mapping and genome-wide association analysis identifies candidate genes for phytophthora blight resistance in sesame (Sesamum indicum L.). *Frontiers in Plant Science*, *12*, 604709.

Beavis, W. D. (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference*, *250*, 266.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science*, *48*(5), 1649–1664.

Biabani, A. R., & Pakniyat, H. (2008). Evaluation of seed yield-related characters in sesame (Sesamum indicum L.) using factor and path analysis. *Pakistan Journal of Biological Sciences: PJBS*, *11*(8), 1157–1160.

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś., Yandell, B. S., & Churchill, G. A. (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*, *211*(2), 495–502.

Çağırgan, M. İ. (2006). Selection and morphological characterization of induced determinate mutants in sesame. *Field Crops Research*, *96*(1), 19–24.

Chebib, J., & Guillaume, F. (2021). Pleiotropy or linkage? Their relative contributions to the genetic correlation of quantitative traits and detection by multitrait GWA studies. *Genetics*, *219*(4), iyab159.

Chen, Y., & Lübberstedt, T. (2010). Molecular basis of trait correlations. *Trends in Plant Science*, *15*(8), 454–461.

Coffman, C. J., Doerge, R. W., Wayne, M. L., & Mcintyre, L. M. (2003). *Intersection tests for single marker QTL analysis can be more powerful than two marker QTL analysis*. http://www.biomedcentral.com/1471-2156/4/10

Colasuonno, P., Marcotuli, I., Gadaleta, A., & Soriano, J. M. (2021). From genetic maps to QTL cloning: an overview for durum wheat. *Plants*, *10*(2), 315.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, *142*(1), 169–196.

Emamgholizadeh, S., Parsaeian, M., & Baradaran, M. (2015). Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*, *68*, 89–96.

FAO. (2022). *FAO - Food and Agriculture Organization of the United Nations*. https://www.fao.org/faostat/en/#data/QCL

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint ArXiv:1207.3907*.

Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, *10*(2), 156–161.

IBRAFE. (2023, December 1). *Space for Brazilian Sesame in the International Market*. Https://Ibrafe.Org/Artigo/Espaco-Para-o-Gergelim-Brasileiro-No-Mercado-Internacional.

Ibrahim, S. E., & Khidir, M. O. (2012). *Genotypic correlation and path coefficient analysis of yield and some yield components in sesame (Sesamum indicum L.).*

Islam, F., Gill, R. A., Ali, B., Farooq, M. A., Xu, L., Najeeb, U., & Zhou, W. (2016). Sesame. In *Breeding Oilseed Crops for Sustainable Production* (pp. 135–147). Elsevier.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., & Hopper, C. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Data Production and Analysis in Population Genomics: Methods and Protocols*, 67–89.

Langham, D. R. (2001). *Shatter resistance in sesame*.

Langham, D. R. (2014). *Method for mechanical harvesting of improved non-dehiscent sesame*. Google Patents.

Langham, D. R., & Wiemers, T. (2002). *Progress in mechanizing sesame in the US through breeding.*

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.

MAPA. (2023, November 30). *O mundo precisa do Brasil para ter alimentos*. Https://Www.Gov.Br/Agricultura/Pt-Br/Assuntos/Noticias-2022/o-Mundo-Precisa-Do-Brasil-Para-Ter-Alimentos-Diz-Ministro#:~:Text=Segundo%20ele%2C%20como%20importante%20player,Disse.

Mauricio, R. (2001). Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics*, *2*(5), 370–381.

Mei, H., Cui, C., Liu, Y., Du, Z., Wu, K., Jiang, X., Zheng, Y., & Zhang, H. (2023). QTL analysis of traits related to seed size and shape in sesame (Sesamum indicum L.). *Plos One*, *18*(11), e0293155.

Mei, H., Liu, Y., Cui, C., Hu, C., Xie, F., Zheng, L., Du, Z., Wu, K., Jiang, X., & Zheng, Y. (2021). QTL mapping of yield-related traits in sesame. *Molecular Breeding*, *41*(7), 1–14.

Melchinger, A. E., Utz, H. F., & Schön, C. C. (1998). Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, *149*(1), 383–403.

Moskvina, V., & Schmidt, K. M. (2008). On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *32*(6), 567–573.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., & Labhane, N. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, *32*(2), 261–285.

Nóia Júnior, R. de S., & Sentelhas, P. C. (2019). Soybean-maize off-season double crop system in Brazil as affected by El Niño Southern Oscillation phases. *Agricultural Systems*, *173*, 254–267.

Qureshi, M., Langham, D. R., Lucas, S. J., Uzun, B., & Yol, E. (2022). Breeding history for shattering trait in sesame: classic to genomic approach. *Molecular Biology Reports*, *49*(7), 7185–7194.

Raghavan, C., Mauleon, R., Lacorte, V., Jubay, M., Zaw, H., Bonifacio, J., Singh, R. K., Huang, B. E., & Leung, H. (2017). Approaches in characterizing genetic structure and mapping in a rice multiparental population. *G3: Genes, Genomes, Genetics*, *7*(6), 1721–1730.

Rahman, A., Bhattarai, S., Akbar, D., Thomson, M., Trotter, T., & Timilsina, S. (2019). *Market analysis of sesame seed*. CQUniversity Australia. https://crcna.com.au/sites/default/files/2020-02/200111 Sesame Market Analysis.pdf

Rebai, A., Goffinet, B., & Mangin, B. (1995). Comparing power of different methods for QTL detection. *Biometrics*, 87–99.

Sabag, I., Bi, Y., Sahoo, M. M., Herrmann, I., Morota, G., & Peleg, Z. (2024). Leveraging genomics and temporal high-throughput phenotyping to enhance association mapping and yield prediction in sesame. *The Plant Genome*, e20481.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, *5*(Suppl 7), P54.

Schön, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., & Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics*, *167*(1), 485–498.

Stamatav, S., Ishpekov, S., Deshev, M., Vankova, E., & Dallev, M. (2020). SEED RETAINING MODEL OF NON-DEHISCENCE SESAME (Sesamum indicum L.) GENOTYPES AT RIPENING. *Scientific Papers. Series A. Agronomy, 63 (1)*, 541–546.

Teboul, N., Magder, A., Zilberberg, M., & Peleg, Z. (2022). Elucidating the pleiotropic effects of sesame kanadi1 locus on leaf and capsule development. *The Plant Journal*, *110*(1), 88–102.

Tripathy, S. K., Kar, J., & Sahu, D. (2019). Advances in sesame (Sesamum indicum L.) breeding. *Advances in Plant Breeding Strategies: Industrial and Food Crops: Volume 6*, 577–635.

Uzun, B., & Çağirgan, M. I. (2006). Comparison of determinate and indeterminate lines of sesame for agronomic traits. *Field Crops Research*, *96*, 13–18. https://doi.org/10.1016/j.fcr.2005.04.017

Uzun, B., Lee, D., Donini, P., & ÇAğirgan, M. L. (2003). Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. *Plant Breeding*, *122*(1), 95–97.

Wang, J., van Ginkel, M., Trethowan, R., Ye, G., DeLacy, I., Podlich, D., & Cooper, M. (2004). Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. *Crop Science*, *44*(6), 2006–2018.

Weldemichael, M. Y., & Gebremedhn, H. M. (2023). QTL Mapping in Sesame (Sesamum indicum L.): A review. *Journal of Biotechnology*.

Wu, K., Liu, H., Yang, M., Tao, Y., Ma, H., Wu, W., Zuo, Y., & Zhao, Y. (2014). High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (Sesamum indicum L.) based on RAD-Seq technology. *BMC Plant Biology*, *14*, 1–14.

Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *Theoretical and Applied Genetics*, *125*, 201–210.

Xiaohong, H., Chenwu, X., Jianmin, K., Shiliang, G., & Tao, L. (2001). Principal factors affecting the power of detection and accuracy of QTL mapping. *Zuo Wu Xue Bao*, *27*(4), 469–475.

Yamamoto, E., Iwata, H., Tanabata, T., Mizobuchi, R., Yonemaru, J., Yamamoto, T., & Yano, M. (2014). Effect of advanced intercrossing on genome structure and on the power to detect linked quantitative trait loci in a multi-parent population: a simulation study in rice. *BMC Genetics*, *15*, 1–17.

Zhang, H., Miao, H., Li, C., Wei, L., Duan, Y., Ma, Q., Kong, J., Xu, F., & Chang, S. (2016). Ultra-dense SNP genetic map construction and identification of SiDt gene controlling the determinate growth habit in Sesamum indicum L. *Scientific Reports*, *6*(1), 31556.

Zhang, H., Miao, H., Wei, L., Li, C., Duan, Y., Xu, F., Qu, W., Zhao, R., Ju, M., & Chang, S. (2018). Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. *BMC Plant Biology*, *18*(1), 1–12.

Zhang, H., Miao, H., Wei, L., Li, C., Zhao, R., & Wang, C. (2013). Genetic analysis and QTL mapping of seed coat color in sesame (Sesamum indicum L.). *PloS One*, *8*(5), e63898.

Zhang, L., Li, H., Li, Z., & Wang, J. (2008). Interactions between markers can be caused by the dominance effect of quantitative trait loci. *Genetics*, *180*(2), 1177–1190.

Zhang, Y., Wang, L., Gao, Y., Li, D., Yu, J., Zhou, R., & Zhang, X. (2018). Genetic dissection and fine mapping of a novel dt gene associated with determinate growth habit in sesame. *BMC Genetics*, *19*(1), 1–10.

Zhou, W., Kolb, F. L., Bai, G., Domier, L. L., Boze, L. K., & Smith, N. J. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding*, *122*(1), 40–46.