ALISON UBERTI

MOLECULAR BREEDING APPROACHES FOR THE IMPROVEMENT OF OIL CONTENT AND FATTY ACIDS COMPOSITION IN EXOTIC-DERIVED MAIZE GERMPLASM

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: Rodrigo Oliveira de Lima

Co-adviser: Thomas Lübberstedt

VIÇOSA - MINAS GERAIS
2024

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

T
U14m
2024

Uberti, Alison, 1993-
Molecular breeding approaches for the improvement of oil content and fatty acids composition in exotic-derived maize germplasm / Alison Uberti. – Viçosa, MG, 2024.
1 tese eletrônica (57 f.): il. (algumas color.).

Texto em inglês.
Orientador: Rodrigo Oliveira de Lima.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2024.
Referências bibliográficas: f. 33-48.
DOI: https://doi.org/10.47328/ufvbbt.2024.714
Modo de acesso: World Wide Web.

1. Milho - Melhoramento genético. 2. Mapeamento cromossômico. 3. Óleo de milho. 4. Ácidos graxos. I. Lima, Rodrigo Oliveira de, 1982-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.152

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

ALISON UBERTI

MOLECULAR BREEDING APPROACHES FOR THE IMPROVEMENT OF OIL CONTENT AND FATTY ACIDS COMPOSITION IN EXOTIC-DERIVED MAIZE GERMPLASM

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: August 16, 2024.

Assent:

Documento assinado digitalmente
gov.br ALISON UBERTI
Data: 26/10/2024 07:49:35-0300
Verifique em https://validar.iti.gov.br

Alison Uberti
Author

Documento assinado digitalmente
gov.br RODRIGO OLIVEIRA DE LIMA
Data: 16/10/2024 17:07:25-0300
Verifique em https://validar.iti.gov.br

Rodrigo Oliveira de Lima
Adviser

ACKNOWLEDGEMENTS

To God, for life, for everything He has given me and for being present in every moment of my life.

To my parents, Pedro Paulo Uberti and Lorita Fátima Uberti, my fiancée Alice Silva Santana, my brother Alex Uberti and sister-in-law Janaina Uberti, for their support, encouragement and love.

To the Universidade Federal de Viçosa (UFV), the Departamento de Agronomia, especially the Graduate Program in Genetics and Breeding, for the opportunity to join the program and learn from such brilliant professors.

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), to granting the scholarship.

To my advisor, Professor Rodrigo Oliveira de Lima, for the help and support in the research.

To Professor Thomas Lübberstedt for the opportunity to join the Lübberstedt Lab and Doubled Haploid Facility at Iowa State University during my sandwich doctorate.

To Programa Milho-UFV, undergraduate and graduate students, for their support, friendship and help in carrying out the project, as without them it would not have been able to conduct the experiments.

To the staff of the Campo experimental de Coimbra, Horta Nova and Vale da Agronomia, for all their dedication and support in conducting the experiment.

To my friends, for their valuable friendship and support in completing this part of my life.

To all of you, thank you very much!

ABSTRACT

UBERTI, Alison, D.Sc., Universidade Federal de Viçosa, August, 2024. **Molecular breeding approaches for the improvement of oil content and fatty acids composition in exotic-derived maize germplasm**. Adviser: Rodrigo Oliveira de Lima. Co-adviser: Thomas Lübberstedt.

Molecular breeding strategies such as genome-wide association studies (GWAS) and genomic prediction have revolutionized crop improvement by enhancing selection accuracy and genetic gains. Through a comprehensive evaluation of a large set of lines from Germplasm Enhancement of Maize (BGEM) and their testcross hybrids, we aimed to elucidate the genetic basis of oil content and fatty acid compositions and predict superior hybrids as well as breeding populations. Our study revealed wide phenotypic variation among BGEM lines and testcrosses for such traits, with promising genotype mean values. Leveraging GWAS, we identified significant genomic regions associated with oil content and fatty acids composition, unveiling candidate genes putatively for trait improvement. Incorporating both additive and non-additive genomic prediction model did not enhance the prediction accuracy, suggesting that the additive effects play a major role in the genetic architecture of these traits. Furthermore, our predictive modeling facilitated the identification of breeding populations with heightened oil content and superior fatty acid profiles compared to original BGEM lines. These findings highlight how effective we can be using BGEM germplasm and molecular breeding approaches to improve oil-related traits in maize germplasm, offering insights for future crop improvement endeavors.

Keywords: *Zea mays* L.; Germplasm Enhancement of Maize; GWAS; Genomic prediction; Breeding populations.

RESUMO

UBERTI, Alison, D.Sc., Universidade Federal de Viçosa, agosto, 2024. **Abordagens moleculares para a melhoria do teor de óleo e da composição de ácidos graxos em germoplasma de milho de origem exótica.** Orientador: Rodrigo Oliveira de Lima. Coorientador: Thomas Lübberstedt.

Estratégias de melhoramento molecular tais como estudo de associação genômica ampla (GWAS) e predição genômica revolucionaram o melhoramento de plantas ao aumentar a acurácia de seleção e o ganho genético. Neste contexto, o objetivo deste trabalho consistiu em elucidar a base genética do conteúdo de óleo e a composição de ácidos graxos e predizer híbridos superiores bem como populações de melhoramento ao avaliar um grande número de linhagens pertencentes ao *Germplasm Enhancement of Maize* (BGEM) e seus híbridos testcross. De acordo com os resultados, observou-se ampla variação fenotípica entre as linhagens BGEM e os testcrosses, sendo possível identificar genótipos com médias promissoras. Além disso, identificaram-se regiões genômicas significativas associadas com o conteúdo de óleo e composição de ácidos graxos, revelando genes candidatos para o melhoramento destes caracteres. A incorporação de efeitos aditivos e não-aditivos no modelo genômico não aumentou a acurácia de predição, sugerindo que os efeitos aditivos são os que mais contribuem para a arquitetura genética do conteúdo de óleo e ácidos graxos. Além disso, os modelos preditivos facilitaram a identificação de populações de melhoramento com maior conteúdo de óleo e melhor composição de ácidos graxos comparado com as linhagens BGEM parentais. Estes resultados reforçam o quão efetivo pode ser o uso do germoplasma BGEM bem como a adoção de abordagens de melhoramento molecular para melhorar caracteres relacionados ao óleo em milho, oferecendo uma melhor compreensão para futuros esforços de melhoramento de plantas.

Palavras-chave: *Zea mays* L.; *Germplasm Enhancement of Maize* (GEM); GWAS; Seleção genômica; Populações de melhoramento.

LIST OF TABLES

LIST OF FIGURES

SUMMARY

1. Introduction

Maize (*Zea mays* L.) is one of the most important cereal crop and holds significant value as a staple commodity worldwide (Tanumihardjo et al., 2020). Historically, maize breeding efforts have concentrated on bolstering stability and enhancing grain yield potential (Ruiz et al., 2023). Nonetheless, recent years have witnessed a notable shift towards enhancing maize for both animal feed and human nutrition, with consumers increasingly prioritizing kernel quality (Prasanna et al., 2020). The elevated nutritional profile of maize kernels primarily stems from their starch content (constituting approximately 72% of the kernel's dry matter), protein content (about 6 to 12% of the kernel's dry matter), and oil content (ranging from 3 to 5.5% of the kernel's dry matter; Watson 2003; Nuss et al., 2010; Kahrıman et al., 2015). Among these key kernel components, maize oil represents a significant reservoir of polyunsaturated fatty acids and serves as a valuable resource for human consumption, animal feed, and bioenergy purposes. It comprises five predominant fatty acids that collectively constitute over 98% of its concentration: palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), and linolenic (C18:3) acids (Yu et al., 2022). The elevated energy content and polyunsaturated fatty acid profile make maize oil a premium-quality edible oil with several benefits for human health. Consequently, an increase in oil content within maize kernels is poised to augment the intrinsic nutritional value of maize varieties. Past efforts in increasing oil content in maize includes the development of the Illinois High-Oil (IHO) population (Dudley and Lambert, 2004), the Alexho single-kernel synthetic population (Lambert et al., 2004) and the Beijing High-Oil (BHO) population (Song and Chen, 2004). These maize populations have paved the way for elucidating the genetic architecture of oil biosynthesis in maize kernels (Laurie et al., 2004; Yang et al., 2010). A comprehensive understanding of the relationship between the genetic architecture kernel oil traits and lipid metabolism in maize will be important to develop selection strategies for high-oil maize breeding and biotechnology-driven improvements.

The genetic improvement of temperate maize for increased oil content and fatty acids can be accelerated with the utilization of exotic germplasm since exotic germplasm can be a source of novel alleles that can be introduced into locally adapted germplasm (Hallauer and Carena, 2014; Vanous et al., 2019). Notably, the Germplasm Enhancement of Maize (GEM) project stands out as a significant source of exotic germplasm. Established in 1993, the GEM project aims to increase allelic diversity in temperate maize by releasing tropical and subtropical germplasm derived from exotic sources. Its overarching goal is to broaden the genetic base of commercial maize cultivars through the evaluation, identification, and incorporation of beneficial genes from maize landraces (Pollak and Salhuana, 2001; Pollak, 2003; Salhuana and Pollak, 2006; Rogers et al., 2022). In this context, both tropical and subtropical germplasm has potential to contribute useful alleles to grain quality and resistance to major maize pests to U.S. Corn Belt breeding programs (Carena et al., 2009; Sharma and Carena, 2012; Laude and Carena, 2013; Hallauer and Carena, 2014). Within the allelic diversity component of the GEM project, doubled haploid (DH) lines were derived from BC_1F_1 or F_1 crosses between tropical and subtropical accessions and the elite inbred lines PHB47 (stiff stalk) and PHZ51 (non-stiff stalk), both of which are expired plant variety protection (ex-PVP) lines (Brenner et al., 2012; Sanchez et al., 2023). Currently, the DH lines released through this initiative are referred to as BGEM lines, with the B indicating Iowa State University, the institution where these DH lines were developed. Recently, the BGEM panel was used to dissect the genetic basis of grain, plant height and anthesis to silking interval under contrasting nitrogen inputs (Sanchez et a., 2023), flowering time and plant stature (Vanous et al., 2018), and root-related traits (Sanchez et al., 2018; Ma et al., 2020; Zuffo et al., 2022). Even though Vanous et al. (2019) provided initial insights into the oil content of BGEM lines, further studies focusing on fatty acids composition have not been undertaken.

Genome-wide association studies (GWAS) have been successfully employed for dissecting the genetic architecture of complex traits in maize (Galli et al., 2020; Zhang et al., 2022; Dwiningsih and Al-Kahtani, 2022). A number of chromosomal regions and quantitative trait *loci* (QTL) controlling oil content and fatty acids composition in maize have been identified through molecular markers (Mangolin et al., 2004; Song and Chen, 2004; Clark et al., 2006; Zhang et al., 2008; Li et al., 2013; Fang et al., 2021; Zheng et al., 2021; Zhang et al., 2023). In these studies, the authors suggested that oil content and fatty acids composition are controlled by a large number of genes with a range of small to large effects and mainly additive gene action. For oil content, Zhang et al. (2023) detected 16 QTL associated with oil content in kernels from four DH populations. Fang et al. (2021) found 62 QTL for oil content and fatty acid composition in a recombinant inbred line (RIL) population. Additionally, genes for several key enzymes in fatty acid biosynthesis, including *ZmfatB* and *fad2*, have been identified by fine mapping and GWAS in maize. These genes are associated with C16:0 and C18:1 fatty acids concentration, respectively (Li et al., 2011; Li et al., 2013).

In recent years, genomic prediction has emerged as a widely employed genomic-based approach in both plant and animal breeding. Unlike GWAS, which involves identifying trait-associated markers, genomic prediction focuses on marker-based selection where genome-wide markers are utilized to capture the total genetic variance without the need for prior marker-trait association analysis (Meuwissen et al., 2001). The primary objective of genomic prediction is to predict the genetic potential, called genome-estimated breeding values (GEBVs), of genotypes without explicitly identifying specific genes or QTL associated with the traits of interest. One of the key statistical challenges in genomic prediction is improving prediction accuracy, which has been a focal point of research efforts (Jarquin et al., 2021; Atanda et al., 2021; Barreto et al., 2024). Several strategies have been explored to enhance prediction accuracy. Among them, adding non-additive effects such as dominance can provide a more

comprehensive understanding of trait inheritance and facilitate more effective breeding strategies. This endeavor aims to refine the predictive performance of genomic prediction models and enable more accurate selection of breeding candidates, ultimately facilitating accelerated genetic gain in breeding programs.

The goal of this study was to investigate the genetic basis of oil content and ten fatty acids in maize using a large set of BGEM lines *per se* and their testcross hybrids. Our objectives were to: (i) evaluate the field performance of BGEM lines *per se* and testcross hybrids for oil content and fatty acids composition; (ii) identify genomic regions associated with variation for oil content and fatty acids composition in BGEM lines and testcross hybrids; (iii) predict the best hybrid combinations and new breeding populations using additive and non-additive genomic prediction models to improve the oil content and fatty acids composition in the exotic-derived maize germplasm.

2 Material and methods

2.1 Plant materials

In total, 66 exotic maize landraces from the Germplasm Enhancement of Maize (GEM) project were crossed with both expired plant variety protection (PVP) lines PHB47 (stiff stalk heterotic group - SS) and PHZ51 (non-stiff stalk heterotic group - NSS). The F_1 plants were backcrossed once with either PHB47 or PHZ51 to produce the BC_1F_1 generation (Sanchez et al., 2018). The BC_1F_1 generation was used as source population to produce DH lines. DH lines were developed using the protocol described by Vanous et al. (2017). In brief, the hybrid inducer RWS 9 × RWK-76 (Röber et al., 2005) was crossed with plants from each BC_1F_1 population to produce haploid seeds that were identified based on the *R-nj* color marker (Liu et al., 2016a). Putative haploids were grown in the greenhouse and treated with colchicine at V3-V4 stage to promote genome doubling. After that, plants were transplanted to the field and self-pollinated to produce DH lines. Seeds of these DH lines were increased at the North-Central Region Plant

Introduction Station (NCRPI Station) of USDA-ARS in Ames, Iowa, and Agricultural Engineering and Agronomy Farm (AEAF) of the Iowa State University in Boone, Iowa. In total, 145 and 96 DH lines were produced from the BC_1F_1 crosses with the recurrent parents PHB47 (SS) and PHZ51 (NSS), respectively (total of 241 DH lines). From now onward, we termed this set of DH lines as "BGEM lines". In addition to BGEM lines *per se*, we developed a set of testcross hybrids by crossing our BGEM lines with two testers (PHB47 and PHZ51). Thus, BGEM lines from the SS group were crossed with PHZ51, whereas BGEM lines from the NSS group were crossed with PHB47, resulting in PHB47 × PHZ51 hybrids with exotic introgressions contributed by one of the GEM parents. A total of 187 testcross hybrids were obtained, comprising 74 hybrids from the SS group and 113 hybrids from the NSS group.

2.2 Trial management and experimental design

The 241 maize BGEM lines and the two recurrent parents (PHB47 and PHZ51) were evaluated across three environments: two environment (2013 and 2014 summer seasons) at the NCRPI Station, and one environment at the AEAF, during the 2014 summer season. The 187 maize testcross hybrids were evaluated during the 2015 summer season across three environments: one at the NCRPI Station, and two trials on the same farm at AEAF. All experiments were carried out under rain-fed conditions, and no irrigation was applied. Trial management was the same for all environments employing standard agricultural practices.

In all environments, the BGEM lines and testcross hybrids were planted in a randomized complete block design with three replications. Each plot was a double 5.64 m rows, with rows spaced 0.76 m apart. Planting density was 65,323 plants ha^{-1}.

2.3 Phenotypic data

We evaluated oil content (oil, %) and ten fatty acids traits: palmitic acid (C16:0, %), palmitoleic acid (C16:1, %), stearic acid (C18:0, %), oleic acid (C18:1, %), linoleic acid (C18:2, %), linolenic acid (C18:3, %), arachidic acid (C20:0, %), gadoleic acid (C20:1, %), docosanoic acid (C22:0, %) and tetracosanoic acid (C24:0, %) in maize BGEM lines and testcross hybrids in all environments. Briefly, random kernels from the bulked grain of each plot were ground and a HP6890 gas chromatogram (Agilent Technologies, USA) was employed for fatty acid analysis. All samples were measured as duplicate at the same time and averaged out. Lipids were extracted as described by Sukhija and Palmquist (1988) modified by Yang et al. (2010). The oil content was calculated as the sum of all identified fatty acid concentrations as percentage of kernel weight. Individual fatty acids were expressed as percentage of oil content.

2.4 Genotypic data

Young leaves from all BGEM lines and the testers PHB47 and PHZ51 were sampled for DNA extraction and subsequently genotyped using genotyping-by-sequencing (GBS) markers, as described by Elshire et al. (2011). The BGEM lines were genotyped using 955,690 GBS markers at the Cornell Institute for Genomic Diversity (IGD) laboratory, aligned with B73 AGPv2 reference genome. After filtering out missing data (>25%), minor allele frequency (<2.5%) and monomorphic markers, 62,077 markers were left and distributed across all 10 chromosomes. These markers were used for further analyzes.

We were not expecting a high number of recombination events per line, as the population source was a BC_1F_1 generation in the DH line induction. Therefore, the genotypic data were corrected for monomorphic markers that were located between flanking markers displaying donor parent genotypes. The correction was based on Bayes theorem as described by Sanchez et al. (2018), with an underlying assumption that very short distances of a marker with recurrent

parent genotype to flanking markers with donor genotype are more likely due to identity of marker alleles for that SNP between recurrent parent and donor, instead of a rare double recombination event. The testcross hybrids genotype data were generated using the 'create hybrid genotypes' function in TASSEL 5.2.90 (Bradbury et al., 2007) with genotype information from the BGEM lines and the PHB47 and PHZ51 testers.

2.5 Phenotypic data analysis

A mixed model implemented in R package "lme4" (Bates et al., 2015) was used to estimate the variance components and to predict the genotypic values for oil content and all fatty acid traits separately for the BGEM lines *per se* and testcross hybrids trials. Phenotypic values were modeled according to the following model: $y_{ilj} = \mu + g_i + e_l + ge_{il} + b_{j(l)} + \varepsilon_{ilj}$ (1), where y_{ilj} is the phenotypic value of i^{th} genotype at l^{th} environment in the j^{th} replication; μ is the mean; g_i is the random effect of i^{th} genotype with $g \sim N(0, \sigma_g^2)$, where σ_g^2 is the variance component of genotypes; e_l is the fixed effect of l^{th} environment; ge_{il} is the random effect of the genotype-by-environment interaction with $ge \sim N(0, \sigma_{ge}^2)$, where σ_{ge}^2 is the variance component of genotype-by-environment; $b_{j(l)}$ is the random effect of j^{th} block within the l^{th} environment with $b \sim N(0, \sigma_b^2)$, where σ_b^2 is the variance component of block within environment, and ε_{ilj} is the random effect of error with $\varepsilon \sim N(0, \sigma^2)$, where σ^2 is the residual variance. Variance components were estimated using a restricted maximum likelihood approach, and a likelihood ratio test deviance analysis was used to test random effects via the chi-square statistic (Resende, 2007). The genotypic values of genotypes were predicted using the best linear unbiased predictors (BLUP; Piepho et al., 2008). In the initial stage of the analyses, the genotype effect was treated as random to estimate the variance component and the BLUP values were used to Pearson's and Spearman's correlations. Subsequently, this effect was

treated as fixed to calculate the best unbiased linear estimator (BLUE). The BLUE values of genotype effect were used to perform the GWAS and genomic prediction analyses.

Broad-sense heritability ($\hat{h}^2_x$) on a BGEM lines-mean and testcross hybrids-mean basis of each trait across environments were estimated as follows (Hallauer et al., 2010): $\hat{h}^2_x = \frac{\hat{\sigma}_g}{\hat{\sigma}_g + \frac{\hat{\sigma}_{ge}}{e} + \frac{\hat{\sigma}^2}{er}}$, where $\hat{\sigma}_g$ and $\hat{\sigma}_{ge}$ are the genotypic and genotype-by-environment interaction variance components, respectively, $\hat{\sigma}^2$ is the residual variance estimate, and e and r are the number of environments and replications, respectively. Pearson's correlation coefficients ($\hat{r}$) between traits were estimated using the BLUP values of each BGEM lines *per se* and testcross hybrids. Spearman's correlation coefficients between BGEM DH lines and testcross hybrids performance for the same trait were also estimated. Correlation coefficients were estimated using the R package "Hmisc" (Harrell Jr and Dupont, 2023).

2.6 Genotypic data analyses

Genome-wide association studies

Adjusted mean of oil content and all fatty acid traits in BGEM lines *per se* and testcross hybrids were used for GWAS. GWAS were performed using the Fixed and Random Model Circulating Probability Unification (FarmCPU) model implemented in the R package "GAPIT" (Lipka et al., 2012; Liu et al., 2016b). The FarmCPU model incorporates principal component analysis (PCA) results as a covariate, along with kinship as an additional covariate to address the relatedness among individuals (VanRaden, 2008). Additionally, it applies algorithms designed to address confounding issues between test markers and covariates, thus controlling false positives and preventing overfitting, as outlined by Liu et al. (2016b). For the population structure in the GWAS analyses, we allocated the entire panel into two subpopulations, as described by Sanchez et al. (2018). For the BGEM lines, we employed an additive genetic model, while for the testcross hybrids, both additive and dominant genetic models were fitted.

The Bonferroni test was used to determine the multiple testing threshold level, which was equal to $P = 0.05/62{,}077 = 8.05\times10^{-7}$, where 62,077 is the total number of SNP markers after quality control (Holm, 1979). SNPs markers whose *p*-values exceeding the threshold were identified as candidate *loci*. The linkage disequilibrium (LD) decay distance for each chromosome reported by Sanchez et al. (2018) was used to determine the upstream and downstream range of the candidate *loci*. The MaizeGDB database was used to find linked candidate genes for each candidate *loci* based on the maize B73 RefGen_V5. The National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/gene) was used to search for functional annotations.

Genomic prediction for the best hybrid combinations

The genomic prediction was performed for the testcross hybrids with the extended best line unbiased prediction (GBLUP) model. Our strategy was to use the two heterotic groups (SS and NSS) to predict the best hybrid combinations by North Carolina II (NCII; Comstock and Robinson, 1948) design with high oil content and the five fatty acids traits with the highest concentration (C18:2, C18:1, C16:0, C20:0 and C24:0). We fitted additive (G_a) and dominant (G_d) relationship matrices for extended GBLUP models. Thus, the two extended GBLUP models were used: $\mathrm{Model}(G_a\#G_a): y = 1_n\mu + Z_1u_{aSS} + Z_2u_{aNSS} + \varepsilon$ (2) and $\mathrm{Model}(G_a\#G_a\#G_d): y = 1_n\mu + Z_1u_{aSS} + Z_2u_{aNSS} + Z_3u_d + \varepsilon$ (3), where y is the vector of hybrid BLUE values; 1_n is the n-dimensional vector of ones; μ is the overall mean; u_{aSS}, u_{aNSS} and u_d are the vectors of random effects for additive SS, additive NSS and dominance effects assumed to follow the normal distributions $N(0, G_{aSS}\sigma^2_{aSS})$, $N(0, G_{aNSS}\sigma^2_{aNSS})$ and $N(0, G_d\sigma^2_d)$, respectively; G_{aSS}, G_{aNSS} and G_d are the genomic relationship matrices corresponding to additive SS, additive NSS and dominance genotypic values, respectively; and ε is a normal residuals distribution vector, with $N(0, \sigma^2_e)$, where $\sigma^2_e$is the residual variance. Z_1, Z_2 and Z_3

represents the incidence matrices to aSS, aNSS and d effects, respectively. The genomic relationship matrices and genomic prediction models were performed using the R package "sommer" (Covarrubias-Pazaran, 2016).

We used a five-fold cross validation approach to assess the ability of the tested genomic prediction models. Prediction accuracy was estimated by Pearson's correlation between the BLUE values from the validation set and the genomic estimated breeding values (GEBVs) predicted from a given genomic prediction model evaluated in the test set. The process was repeated 1,000 times. The genomic heritability was estimated by the following equations (de los Campos et al., 2015) for additive models: $h^2_{aSS} = \frac{\sigma^2_{aSS}}{\sigma^2_{aSS}+\sigma^2_{aNSS}+\sigma^2_{\varepsilon}}$ and $h^2_{aNSS} = \frac{\sigma^2_{aNSS}}{\sigma^2_{aSS}+\sigma^2_{aNSS}+\sigma^2_{\varepsilon}}$, and for additive-dominance models: $h^2_{aSS} = \frac{\sigma^2_{aSS}}{\sigma^2_{aSS}+\sigma^2_{aNSS}+\sigma^2_{d}+\sigma^2_{\varepsilon}}$, $h^2_{aNSS} = \frac{\sigma^2_{aNSS}}{\sigma^2_{aSS}+\sigma^2_{aNSS}+\sigma^2_{d}+\sigma^2_{\varepsilon}}$ and $h^2_{d} = \frac{\sigma^2_{d}}{\sigma^2_{aSS}+\sigma^2_{aNSS}+\sigma^2_{d}+\sigma^2_{\varepsilon}}$, where h^2_{aSS}, h^2_{aNSS} and h^2_{d} are the estimates of genomic heritability in the additive for SS, additive for NSS, and dominance models, respectively, σ^2_a is the estimate of additive genetic variance for SS or NSS, σ^2_d is the estimate of dominance genetic variance, and σ^2_{ε} is the estimates of residual variance.

Genomic prediction for breeding populations

Progeny of all possible pairwise cross combinations for each heterotic group (SS and NSS) were simulated, and their performance were predicted for high oil content and the five fatty acids traits with the highest concentration (C18:2, C18:1, C16:0, C20:0 and C24:0). The progeny generation process assumed that all progenies were developed using single seed descent method with each progeny generated from a separate meiotic event within F_1 hybrids. Each resulting S_0 population was then advanced through multiple generations of self-pollination (inbreeding) to generate a S_5 RIL. In total, 94 and 131 BGEM lines from NSS and SS heterotic groups, respectively, were used as parents for the generation of the progenies and prediction

within each heterotic group. Reciprocal crosses were not predicted since we assumed that maternal effects are not significant. The generic RIL population from each breeding population was generated using R package "qtl" (Broman et al., 2003). This requires ordered markers with genetic distances to account for recombination. The genetic map was generated considering the physical map positions of 62k SNP marker set and the maximum distance between markers at 0.85 cM. In general, at any maize breeding program, 200 progenies rows from each breeding population are grown for self-pollination derived from narrow-based parents (Hallauer et al., 2010). Therefore, the 200 RILs genotypes were generated from breeding population, and each population was evaluated based on the GEBVs predicted for their 200 progenies. The GEBVs for all progenies from a breeding population were generated using ridge regression BLUP (RR-BLUP) model with progeny as the validation set, and the parental lines as the training set. The GEBVs were generated using the R package "PopVar" (Mohammadi et al., 2015). In this, the value of each cross is assessed based on the marker effects. This value consists of two components, the genetic variance among progeny and the mean value of the cross' progeny (Mohammadi et al., 2015). Based on the variance and the mean values, the usefulness criterion (UC_m; Schnell and Utz, 1975) was estimated for each population and each group according to the following equation: $UC_m = \mu + ih\sigma_g$ (4), where UC_m is the cross value of a given cross m; μ is the adjusted mean value of cross m's progeny; i is selection intensity (10%); h is the selection accuracy and σ_g is the standard deviation of GEBVs predicted via genomic prediction (Zhong and Jannink, 2007; Neyhart and Smith, 2019). The selection accuracy was assumed to be one, as would be the case when selecting directly on genetic effects (Zhong and Jannink 2007).

3. Results

3.1 Performance of BGEM lines *per se* and testcross hybrids

We observed wide ranges of phenotypic values of the BGEM lines and testcross hybrids for all traits (Tables 1 and 2). Linoleic acid (C18:2) was the fatty acid with the highest, and palmitoleic acid (C16:1) with the lowest concentration in both BGEM line and testcross panels. We observed that the mean of BGEM lines for oil content and stearic acid (C18:0) were higher than the means of both recurrent parents PHB47 and PHZ51. For the C16:0, C18:1, C20:0 and C20:1 fatty acids, BGEM lines showed a higher mean than the recurrent parent PHZ51, but a lower than PHB47. For the other fatty acids (C16:1, C18:2, C18:3, C22:0 and C24:0), BGEM line means were higher than for the PHB47, but lower than for PHZ51. Concerning the analysis of variance, genetic variation among BGEM lines and testcross hybrids were highly significant ($P < 0.00001$) across all tested traits. The estimates of broad-sense heritability $\left(\hat{h}^2_{\bar{X}}\right)$ were high and ranged from 0.92 (oil) to 0.99 (C16:0) for BGEM lines, and from 0.62 (C22:0) to 0.82 (C18:2 and C20:1) for testcross hybrids.

Although most Pearson correlation coefficients among pairs of traits were not significant ($P > 0.01$) or showed low magnitude, we found some moderate-to-strong negative ($\hat{r} < -0.45$) and positive ($\hat{r} > 0.45$) correlations (Figure 1). We observed that the same traits showing strong Pearson correlation coefficients in the BGEM lines also exhibited such pattern in the testcross hybrids. This was evident in correlations such as between C18:1 and C18:2 ($\hat{r}$ = -0.97 and -0.96, respectively) and between C22:0 and C24:0 ($\hat{r}$ = 0.80 and 0.73, respectively). Moderate correlations were found for BGEM lines between C20:0 and C22:0 ($\hat{r}$ = 0.59), C18:1 and C20:1 ($\hat{r}$ = 0.59), C18:0 and C20:0 ($\hat{r}$ = 0.52), oil content and C18:3 ($\hat{r}$ = -0.60), C18:2 and C20:1 ($\hat{r}$ = -0.60), and C16:0 and C18:2 ($\hat{r}$ = -0.54). For testcross hybrids, we found moderate correlations between C20:1 and C18:2 ($\hat{r}$ = -0.57), C20:1 and C18:1 ($\hat{r}$ = 0.56), C20:0 and C22:0 ($\hat{r}$ = -0.56), and oil content and C22:0 ($\hat{r}$ = -0.52). In relation to Spearman correlation

coefficients, although most correlations were significant ($P < 0.01$) and ranged from weak to strong positive correlations. Notably, strong Spearman correlations were found between BGEM lines and their testcross hybrids for C18:0 ($\hat{r} = 0.81$) and C20:0 ($\hat{r} = 0.73$) fatty acids.

3.2 GWAS for oil content and fatty acids in BGEM lines and testcross hybrids

GWAS was conducted using the BLUEs obtained from the combined analysis for the BGEM lines and their testcrosses using the FarmCPU method. Two genetic models, additive and dominant, were used in the testcross hybrid panel, and only the additive model was used in the BGEM line panel. The quantile–quantile (QQ) plots showed that the population structure and family relatedness were well controlled (Supplementary Figure S1, S2 and S3). In the BGEM line panel, GWAS analysis identified 65 SNPs significantly associated with oil content and fatty acid traits (Table 3). As LD decay of the BGEM panel is very extensive (Sanchez et al., 2018), we found many candidate genes within the chromosomal region of known candidate *loci*. To make it simpler, we chose to focus solely on candidate genes associated with the oil and fatty acid pathways. Based on the candidate *loci*, we found seven candidate genes on six chromosomes associated with six traits (Table 4). The candidate genes can be classified into six biological processes, four cellular components, and six molecular functions (Supplementary Table S1). The candidate genes in biological processes are associated with the lipids and fatty acids biosynthetic and metabolic process. *Fatty acid desaturation2* (*fad2*) pathway was identified in our study for the S4_167081234 SNP. This pathway associated with *Zm00001eb188990* gene, on chromosome 4, contains the major enzyme responsible for the synthesis of linoleic (C18:2) fatty acids in the endoplasmic reticulum in *Arabidopsis*. The S9_20337850 SNP on chromosome 9 was significant and is associated with the gene *Zm00001eb376490*. This gene is involved in nine metabolic pathways including the long-chain fatty acid activation (PWY-5143), oleate biosynthesis I (C18:1; PWY-5147), palmitate

biosynthesis II (C16:0; PWY-5971), sporopollenin precursors biosynthesis (PWY-6733), phosphatidylcholine acyl editing (C18:2 and C18:3; PWY-6803), cutin biosynthesis (PWY-321), suberin monomers biosynthesis (C16-C18; PWY-1121), stearate biosynthesis II (PWY-5989) and IV (C18:0; PWY-8280). The *Zm00001eb249240* gene, identified on chromosome 5 with the S5_195306550 SNP, is responsible for fatty acid β-oxidation II (PWY-5136) and IV (PWY-5138), and also very long chain fatty acid biosynthesis II (C20:0; PWY-7036) pathways. A gene, *Zm00001eb010410*, which codes for Acyl-CoA-binding protein 5 (ACBP5), was identified on chromosome 1 for the S1_32868703 SNP. This gene is responsible for fatty-acyl-CoA binding and fatty acid metabolic process in Arabidopsis. We found the *Zm00001eb162340* gene (S3_231891794 SNP), on chromosome 3, which is responsible for fatty acid desaturase14 (*fad14* - PWY-5129) pathway. Also, a hit was found on chromosome 7 (S7_145632437 SNP) located within the gene *Zm00001eb318810*, which is responsible for very long chain fatty acid biosynthesis I (PWY-5080) pathway. However, among all the significant SNPs identified in our study, only one SNP (S3_189459204) was associated with a gene responsible for lipid transport. The *Zm00001eb149490* gene, on chromosome 3, is responsible for *phospholipid transfer protein homolog3* (*plt3*).

In the testcross hybrid panel using additive model, we found 32 SNPs significantly associated with fatty acids traits C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1 and C24:0 (Table 5). Based on candidate *loci*, we found two candidate genes for three fatty acid traits (C18:0, C20:0 and C24:0; Table 6). The candidate genes mainly focus on fatty acid biosynthesis and membrane transport (Supplementary Table S2). Among them, *Zm00001eb046050* gene placed on chromosome 1 is responsible for very long chain fatty acid biosynthesis I (PWY-5080) pathway. The *Zm00001eb432380* gene, on chromosome 10, which had two hits in the GWAS analysis (Supplementary Figure S2) for C20:0 and C24:0 traits, is responsible for *fatty acid export2* (*fax2*). Conversely, using dominant model in hybrid panel,

we found 35 SNPs significantly associated with oil content, and fatty acids traits C16:1, C18:0, C18:1, C18:2, C18:3 and C20:1 (Table 7). However, only two candidate genes were linked to two candidate *loci* on chromosome 7 (Table 6). The candidate genes are responsible for long-chain fatty-acyl-CoA metabolic and fatty acid biosynthetic processes (Supplementary Table S2). The sporopollenin precursor's biosynthesis pathway that was associated with *Zm00001eb308150* gene is responsible for hydrolysis, hydroxylase and catalysis of the C16:0, C18:0 and C18:1 fatty acids, whereas *Zm00001eb317340* gene is responsible for acyl-[acyl-carrier-protein] hydrolase in fatty acid biosynthetic process.

3.3 Genomic prediction

Hybrid combinations between SS and NSS lines

In the additive and additive-dominance genomic models, we estimated the variance components of general combining ability (GCA) for both NSS ($\sigma^2_{GCA\_NSS}$) and SS ($\sigma^2_{GCA\_SS}$) groups of BGEM parent lines, as well as the specific combining ability (σ^2_{SCA}) resulting from combination between two parent lines (Table 8). In the additive model, the $\sigma^2_{GCA\_NSS}$ was highly significant ($P < 0.001$) by the likelihood ratio test for all traits in the genomic prediction, whereas $\sigma^2_{GCA\_SS}$ was significant ($P < 0.001$) just for C16:0, C18:1 and C18:2. The estimates of genomic heritability associated to GCA (h^2_{GCA}) ranged from 0.36 (C18:1 and C20:0) to 0.54 (C24:0), and from 0.05 (oil) to 0.29 (C16:0) in the NSS and SS group, respectively. Notably, the h^2_{GCA} values were much higher in the NSS than in the SS group for all tested traits. Similarly, in the additive-dominance genomic model, we observed that $\sigma^2_{GCA\_NSS}$ was significant ($P < 0.02$) for all traits, where $\sigma^2_{GCA\_SS}$ was not significant ($P > 0.28$) for any trait. In agreement with additive genomic model results, $h^2_{GCA\_NSS}$ values were much higher than $h^2_{GCA\_SS}$ for all traits. In relation to σ^2_{SCA}, there was no effect ($P > 0.30$) for any trait. The h^2_{SCA} values were close to zero for all traits, except for palmitic acid (C16:0; $h^2_{SCA} = 0.48$). The predictive accuracy of ability to

predict hybrids based on the additive genomic model showed the same values as using the additive-dominant genomic model for all traits (Figure 2). The mean values of predictive accuracy for both models were 0.26 for oil content, 0.28 for C24:0, 0.35 for C20:0, 0.37 for C16:0, 0.43 for C18:1 and 0.44 for C18:2.

As both additive and additive-dominant genomic models had similar values of predictive accuracy and there was no dominance (σ^2_{SCA}; $P > 0.30$) effect in the additive-dominance genomic model, we used results from additive model for further analysis since it is less parameterized. Regarding the combining ability among BGEM lines, we found different sets of DH lines with high and positive GCA effects among tested traits in both NSS and SS heterotic groups (Supplementary Table S3). The BGEM lines that showed high and positive GCA effect for linoleic acid (C18:2), showed low and negative GCA for the other five traits for both heterotic groups. This pattern was also evident for predicted genotypic means, where the best predicted hybrids differed among the traits (Supplementary Table S4). Additionally, an increase in the mean values of the top 20 predicted hybrids (untested, but genotyped hybrids) was observed compared to the top 20 testcross hybrids tested for oil content, C16:0, C18:1, C18:2 and C24:0 fatty acid traits. In general, untested hybrids between NSS and SS heterotic groups showed higher oil content and fatty acid compositions than the tested testcross hybrids. We observed that 242, 159, 139, 79, 27 and 11 untested hybrid combinations between NSS and SS from BGEM lines presented higher predicted mean than the best testcross hybrid tested for C20:0, C24:0, C16:0, C18:2, oil content and C18:1, respectively.

Breeding populations

We observed reasonable oil content and fatty acid compositions in 4.371 and 8.515 predicted breeding populations within NSS and SS heterotic groups, respectively (Supplementary Table S5). Among the predicted breeding populations, the mean of predicted

progeny for C18:2, the highest fatty acid concentration among the evaluated traits, ranged from 68.8% (BGEM_0191_N/BGEM_0207_N) to 39.5% (BGEM_0248_N/BGEM_0106_N) for NSS heterotic group, and from 63.7% (BGEM_0094_S/BGEM_0076_S) to 51.9% (BGEM_0095_S/BGEM_0101_S) for SS heterotic group. For C18:1, the predicted breeding populations ranged from 43.8% (BGEM_0248_N/BGEM_0106_N) to 16.5% (BGEM_0191_N/BGEM_0207_N) for NSS heterotic group, and from 30.4% (BGEM_0011_S/BGEM_0101_S) to 22.2% (BGEM_0094_S/BGEM_0076_S) for SS heterotic group. The prediction of superior genotypes was notably influenced by mid-parent GEBVs across all tested traits in both NSS and SS heterotic groups. The predicted genetic variances (V_g) were higher in the NSS heterotic group than in the SS for all traits. The estimates of predicted performance variances of 200 progenies were smaller among breeding populations with BGEM lines parents derived from the same GEM landraces in both heterotic groups. In addition, the highest predicted performance variances were generally found when the BGEM lines of extreme means were crossed (i.e., low × high mean) in both heterotic groups. However, progenies derived from these crosses showed low to intermediate means for all tested traits. Aligned with the results of the hybrid predictions, breeding populations that showed high predicted performance for linoleic acid (C18:2), showed low predicted performance for the other five traits in both heterotic groups. We observed that the top 10% simulated progenies (RIL's) of the best 20 breeding populations showed higher predicted values in relation to untested hybrids for all traits in both heterotic groups (Supplementary Table S4). The highest positive increases were found in the NSS heterotic group, which ranged from 0.2% (C16:0) to 55.8% (C18:1), in contrast to the SS heterotic group, where the range extended from 5.9% (C24:0) to 18.0% (C20:0). In our efforts to enhance the concentration of total fatty acids with high oleic (C18:1) and low linoleic (C18:2) fatty acid composition in the new breeding populations, our focus was on evaluating the mean performance and predictive variance. This

will help us in the selection and recommendation of the new breeding populations that will derive new inbred lines. Considering these, we selected 1% of the predicted breeding populations in both heterotic groups based on the populations with higher mean and genetic gain. The selected breeding populations increased the oil content, C16:0, C18:1 and C20:0 fatty acids and decreased C18:2 and C24:0, as both fatty acids showed high BLUP correlation (Figure 1).

4. Discussion

Prioritizing improvements in oil content and fatty acid compositions in maize contributes to enhancing its economic, nutritional, and industrial value, thereby benefiting both producers and consumers alike. Unravelling the genetic basis of these traits using GWAS approaches is a critical step for setting up breeding strategies targeting the development of superior maize varieties with greater oil content in a shorter timeframe and, consequently, speed up the breeding process. In our study, we evaluated a large set of maize DH lines from the GEM project and their testcross hybrids for oil content and ten fatty acid traits across three environments. The DH lines were derived from backcrossing among 66 tropical maize landraces and two expired-PVP temperate lines. We observed that linoleic acid (C18:2) was the predominant fatty acid component, followed by oleic (C18:1) and palmitic (C16:0) fatty acids, which agrees with previous studies (Saoussem et al., 2009; Egesel et al., 2016; Baldin et al., 2018; Ortíz-Islas et al., 2018). There was a wide genotypic variation observed among both BGEM lines and testcross hybrids associated with high $\hat{h}^2_{\bar{x}}$ values, indicating that selection for increasing oil content and fatty acid compositions will allow good genetic progress in our set of maize lines (Zheng et al., 2021; Ndlovu et al., 2022). We observed that the overall mean of the BGEM lines across the different traits was higher than the mean of the testers only for oil content and C18:0 fatty acid concentration. This result suggests that the genetic background of the BGEM lines

plays a significant role in determining these particular traits. On the other hand, the observation that hybrids from either tester had higher means for the remaining fatty acids implies that these traits are more influenced by the interaction between the inbred lines and the specific testers. This suggests that there may be additive and dominant genetic effects contributed by the testers, influencing the expression of these fatty acid traits in the hybrids. Overall, these findings highlight the complex genetic architecture underlying oil content and fatty acid composition in maize. Understanding these genetic mechanisms is crucial for designing effective breeding strategies aimed at improving oil traits in maize varieties.

The strong negative correlation observed between oleic (C18:1) and linoleic (C18:2) fatty acids in both BGEM lines and testcross hybrids panels was previously observed in other maize studies (Pamin et al., 1986; Baldin et al., 2018; Ortíz-Islas et al., 2018; Ray et al., 2019). According to Alrefai et al. (1995), this connection among both fatty acids is associated with the biosynthetic pathway of fatty acids. In one of the biosynthetic processes of fatty acids, the n-6 desaturase enzyme converts oleic to linoleic fatty acid by inserting a double bond at the n-6 carbon position in the oleic fatty acid chain (Alrefai et al., 1995). This insertion leads to an increase in linoleic fatty acid concentration and a decrease in oleic fatty acid in maize kernels. One application possible application of our discoveries refers to silage maize hybrid breeding programs that have been focusing on the reduction of linoleic fatty acid concentration (Karnatam et al., 2023). It has been reported that diets abundant in C18:2 increase the risk for biohydrogenation-induced milk fat depression in dairy cows (He and Armentano, 2011; He et al., 2012). However, a positive correlation has been reported between C18:2 and total fatty acid in maize silage (Khan et al., 2012). Therefore, an optimal approach to decrease biohydrogenation-induced milk fat depression risk caused by C18:2 entails selecting against n-6 desaturase enzyme activity and identifying hybrids with high C18:1 and moderate to high total fatty acid levels. In our GWAS, we found the *Zm00001eb188990* gene on chromosome 4,

responsible for the synthesis of linoleic (C18:2) fatty acid. The *fatty acid desaturation2* (*fad2*) is the main enzyme responsible for polyunsaturated lipid synthesis (Byrum et al., 1997; Yang et al., 2012; Zhang et al., 2012; Launhardt et al., 2023). Therefore, variation within this gene can be targeted for marker-assisted selection among maize genotypes in silage maize breeding programs. Doubled haploids lines lacking the *Zm00001eb188990* gene exhibit reduced synthesis of linoleic fatty acid and increased accumulation of oleic fatty acid in kernels making them viable. Therefore, these lines could be used for hybrid combinations. Furthermore, considering the performance of the best hybrids combinations in the genomic prediction, we found DH lines with high and positive GCA for C18:1, and high and negative GCA for C18:2 fatty acid in both heterotic groups (NSS and SS). This suggests that these DH lines have lower *fad2* enzymatic activity and provide a greater accumulation of C18:1 in the kernels. The DH lines BGEM_0090_N, BGEM_0248_N, BGEM_0259_N and BGEM_0089_N from NSS heterotic group and BGEM_0270_S, BGEM_0074_S, BGEM_0200_S and BGEM_0097_S from SS heterotic group are potential parents for new hybrid combinations due to their high GCA values and could contribution to develop hybrids with superior performance for C18:1 fatty acid.

In addition to GWAS for DH lines, we carried out GWAS for testcross hybrids panel to identify candidate genes for oil content and fatty acids using additive and additive-dominance models. The background of the GWAS panel from homozygous lines developed from multiple synthetic populations is more controllable, which has been confirmed in the nested association mapping population from previous studies (Tian et al., 2011; Wu et al., 2016). In total, eleven candidate genes associated with tested traits were detected in our study with BGEM lines and testcross hybrid panels. For the BGEM lines, we found seven candidate genes strongly associated with fatty acid traits. Notably, a highly significant locus on chromosome 4 was tightly linked with the known lipid metabolic process gene *fad2* (Launhardt et al., 2023) as well

as other reported to be involved in metabolic and biosynthetic fatty acid related processes, including *ZmLACS5* (Wang et al., 2023), *IDP290* (Latimer et al., 2018), *acb5* (Narayanan et al., 2019), *fad14* (Li et al., 2019), *kcs38* (Campbell et al., 2019) and *plt3* (Qiao et al., 2020). Four candidate genes were detected in the hybrid panel by GWAS using additive-dominant model, emphasizing the importance of both additive and dominance effects in the inheritance of maize oil content and fatty acids. These candidate genes were mostly involved in biosynthetic and export of fatty acids, aligning with previously reported genes such as *ad1* (Batsale et al., 2023), *FAR1* (Domergue et al., 2010), *acyl-ACP-hydrolase* (Imai et al., 1992) and *fax2* (Tian et al., 2019) genes. Therefore, in maize breeding for oil content and fatty acids, marker-assisted selection can be employed to aid in the selection of these highly significant regions in new breeding populations. In addition to these 11 candidate genes found in the BGEM lines and testcross hybrid panels using the additive and additive-dominant models, several other candidate *loci* were found in the GWAS analysis, however none of them showed a direct participation in the oil content and fatty acid pathways and thus we did not present them. This reinforces the complexity of the traits under study. Oil content and fatty acid metabolism are governed by a network of genes interacting with each other and with environmental factors. It's possible that some candidate genes are part of this intricate network, but their effects might be indirect or modulatory rather than directly causal. Our GWAS analyses in the BGEM lines and their testcross hybrids allowed the identification of genomic regions significantly associated with oil content and fatty acid compositions in maize. By pinpointing specific regions of the genome associated with these traits, our study contributes to a deeper understanding of the molecular mechanisms controlling oil metabolism in maize. Furthermore, our results demonstrate that the genetic diversity present in the BGEM lines can serve as a valuable source of genes for enhancing oil content and fatty acid compositions in maize varieties.

Relative merits of genomic prediction over phenotypic selection are that it allows the reduction in the duration of breeding cycle and cost compared with phenotypic selection accelerating genetic gains in a breeding program (Beyene et al., 2021; Budhlakoti et al., 2022), especially when combined with DH technology (Fu et al., 2022). We performed genomic prediction analysis on the hybrid panel to explore unrealized hybrid combinations and on the BGEM panel to explore new breeding populations due to the lack of knowledge about the genetic architecture of oil content and fatty acid compositions. In hybrid panel, the inclusion of dominance matrix in the additive-dominance genomic prediction model provided no significant improvement in prediction accuracy in relation to the additive genomic model. This suggests that the additive effects are more important than the non-additive effects (dominance) in oil content and fatty acid compositions in maize. Also, implementing the additive-dominance genomic prediction model introduces additional parameters into the model, particularly with the inclusion of the SCA effect. As a result, the model became more complex, and the variance explained by SCA effects masked the contributions of GCA effects to trait variation in hybrids.

Inheritance studies have shown that oil content and fatty acid compositions in maize are quantitatively inherited traits, with no contribution of non-additive effects, and they present high heritability (Lambert et al., 2004; Laurie et al., 2004; Dudley et al., 2007; Wassom et al., 2008; Zhang et al., 2008; Yang et al., 2010; Môro et al., 2012; Singh et al., 2014). Therefore, breeding strategies, such as intrapopulation recurrent selection, can be used to increase additive effects within each heterotic group and, subsequently, the best performing lines will be used to obtain hybrids with high oil content and fatty acid compositions. Reported results of intrapopulation recurrent selection have shown steady increases in oil content in two maize populations (Dudley and Lambert, 2004; Song and Chen, 2004). Thus, we can develop hybrids with high levels of oil content and fatty acid compositions.

Regarding the genomic prediction of hybrids using additive model, NSS heterotic group had a greater $\hat{\sigma}^2_{\text{GCA\_NSS}}$ and also its lines showed higher GCA effects than SS group for all tested traits. Therefore, BGEM germplasm from NSS heterotic group had high genetic variability and genetic potential for maximizing oil content and fatty acid compositions in maize hybrids. Conversely, even in the additive model, there were no significant $\hat{\sigma}^2_{GCA\_SS}$ effects for oil content, C20:0 and C24:0 fatty acid traits. For traits exhibiting significant effects (C16:0, C18:1 and C18:2), we observed that the estimated variance $\hat{\sigma}^2_{\text{GCA\_SS}}$ was five times smaller than that for the NSS group ($\hat{\sigma}^2_{\text{GCA\_NSS}}$). This suggests that, under selection, there will be minimal genetic gain for these traits due to the reduced genetic variability within the SS group compared to the NSS group. Thus, an alternative to this is the introduction of genetic variability in the SS heterotic group from a genetically close population with high favorable alleles frequency for oil content and fatty acid compositions. The performance of 8,363 potential hybrids was predicted using genotypic and phenotypic data from 187 hybrids. Compared to the predicted hybrids and testcross hybrids performance, we found a proportion ranging from 0.2% to 2.9% of the predicted hybrids with better performance, which provides greater genetic gain for these traits. The increase in maize genetic gain based on genomic prediction was reported by Technow et al. (2014), Marulanda et al. (2016), Dias et al. (2020), Li et al. (2020) and Luo et al. (2023). Thus, we suggested that the best predicted hybrids be developed and then evaluating in the field for oil content and fatty acid compositions. Furthermore, the phenotypic data of these hybrids will be used to "re-feed" the model, providing an increase in predictive accuracy over seasons

The BGEM lines have potential to be used as parental lines in the development of breeding populations targeting grain quality. Also, they have potential as source germplasm for diversifying the genetic pool of U.S. Corn Belt germplasm. A useful application of genomic prediction to maize quality breeding would be in the selection of parents from BGEM panel to develop new breeding populations recycling lines and, consequently, improving their

performance. For that, the mid-parent values have been considered an important criterion to determine which combinations should be performed (Bernardo, 2014; Mohammadi et al., 2015). In the current study, the prediction of superior RILs was strongly influenced by mid-parent GEBVs in all tested traits in both NSS and SS heterotic groups, which agrees with the results from Lehermeier et al. (2017), Beckett et al. (2019) and Almeida et al. (2020). Therefore, when the midparent along with the predicted variance are considered when choosing initial biparental crosses, it is more likely that genetic gain will be maximized. Considering that the DH lines from GEM germplasm were very diverse, we found the most interesting crosses to generate new superior lines in both NSS and SS heterotic groups. Especially for the SS heterotic group, we found a satisfactory increase which was not observed in the initial BGEM lines.

In summary, our study revealed wide variations in phenotypic values among the BGEM lines and testcross hybrids for all traits. Notably, mean values of BGEM lines often surpassed those of the hybrids, indicating their genetic potential for oil and acid fatty breeding. Furthermore, our investigation identified several candidate genes contributing oil and fatty acid composition variation. Also, predictive accuracy unchanged between the additive and additive-dominance genomic models for all traits in the hybrids. This indicated that additive genetic effects have the major contribution for genetic control of oil and acid fatty composition. Overall, our findings suggest that leveraging predictive models based on genetic markers holds immense promise in identifying hybrids with superior performance compared to traditional testing methods. Additionally, the method employed for selecting parents in breeding populations has the potential to be a valuable tool for breeders in identifying optimal parental combinations with the greatest potential for genetic progress in hybrid crop breeding programs.

5. References

Almeida, V. C., Trentin, H. U., Frei, U. K., Lübberstedt, T. (2020). Genomic prediction of maternal haploid induction rate in maize. The Plant Genome, 13(1), e20014. https://doi.org/10.1002/tpg2.20014

Alrefai, R., Berke, T. G., Rocheford, T. R. (1995). Quantitative trait locus analysis of fatty acid concentrations in maize. Genome, 38(5), 894-901. https://doi.org/10.1139/g95-118

Atanda, S. A., Olsen, M., Burgueño, J., Crossa, J., Dzidzienyo, D., Beyene, Y., Gowda, M., Dreher, K., Zhang, X., Prasanna, B.M., Tongoona, P., Danquah, E.Y., Olaoye, G., Robbins, K. R. (2021). Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. Theoretical and Applied Genetics, 134, 279-294. https://doi.org/10.1007/s00122-020-03696-9

Baldin, M., Ying, Y., Fan, Y., Roth, G., Casper, D. P., Harvatine, K. J. (2018). Characterization of linoleic acid (C18: 2) concentration in commercial corn silage and grain hybrids. Journal of dairy science, 101(1), 222-232. https://doi.org/10.3168/jds.2017-12972

Barreto, C. A. V., das Graças Dias, K. O., de Sousa, I. C. Azevedo, C. F., Nascimento, A. C. C., Guimarães, L. J. M., Guimarães, C. T., Pastina, M. M., Nascimento, M. (2024). Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. Sci Rep 14, 1062. https://doi.org/10.1038/s41598-024-51792-3

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 1-48. https://doi.org/10.18637/jss.v067.i01

Batsale, M., Alonso, M., Pascal, S., Thoraval, D., Haslam, R. P., Beaudoin, F., Domergue, F., Joubès, J. (2023). Tackling functional redundancy of Arabidopsis fatty acid elongase complexes. Frontiers in Plant Science, 14, 1107333. https://doi.org/10.3389/fpls.2023.1107333

Beckett, T. J., Rocheford, T. R., Mohammadi, M. (2019). Reimagining maize inbred potential: identifying breeding crosses using genetic variance of simulated progeny. Crop Science, 59(4), 1457-1468. https://doi.org/10.2135/cropsci2018.08.0508

Bernardo, R. (2014). Genomewide selection of parental inbreds: classes of loci and virtual biparental populations. Crop Science, 54(6), 2586-2595. https://doi.org/10.2135/cropsci2014.01.0088

Beyene, Y., Gowda, M., Pérez-Rodríguez, P., Olsen, M., Robbins, K. R., Burgueño, J., Prasanna, B. M., Crossa, J. (2021). Application of genomic selection at the early stage of breeding pipeline in tropical maize. Frontiers in Plant Science, 12, 685488. https://doi.org/10.3389/fpls.2021.685488

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinform. 23(19), 2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Brenner, E. A., Blanco, M., Gardner, C., Lübberstedt, T. (2012). Genotypic and phenotypic characterization of isogenic doubled haploid exotic introgression lines in maize. Mol. Breed. 30, 1001–1016. https://doi.org/10.1007/s11032-011-9684-5

Broman, K. W., Wu, H., Sen, Ś., Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19(7), 889-890.

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., Kumar, S. (2022). Genomic selection: A tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. Frontiers in Genetics, 13, 832153. https://doi.org/10.3389/fgene.2022.832153

Byrum, J. R., Kinney, A. J., Stecca, K. L., Grace, D. J., Diers, B. W. (1997). Alteration of the omega-3 fatty acid desaturase gene is associated with reduced linolenic acid in the A5

soybean genotype. Theoretical and Applied Genetics, 94, 356-359. https://doi.org/10.1007/s001220050423

Campbell, A. A., Stenback, K. E., Flyckt, K., Hoang, T., Perera, M. A. D., Nikolau, B. J. (2019). A single-cell platform for reconstituting and characterizing fatty acid elongase component enzymes. PloS one, 14(3), e0213620. https://doi.org/10.1371/journal.pone.0213620

Comstock, R. E., Robinson, H. F. (1948). The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics, 4(4), 254-266.

Carena, M. J., Pollak, L., Salhuana, W., Denuc, M. (2009). Development of unique and novel lines for early-maturing hybrids: moving GEM germplasm northward and westward. Euphytica, 170, 87-97. https://doi.org/10.1007/s10681-009-9985-5

Clark, D., Dudley, J. W., Rocheford, T. R., LeDeaux, J. R. (2006). Genetic analysis of corn kernel chemical composition in the random mated 10 generation of the cross of generations 70 of IHO× ILO. Crop science, 46(2), 807-819. https://doi.org/10.2135/cropsci2005.06-0153

Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the r package sommer. PloS One 11. https://doi.org/10.1371/journal.pone.0156744

de Los Campos, G., Sorensen, D., Gianola, D. (2015). Genomic heritability: what is it?. PLoS Genetics, 11(5), e1005048.

Dias, K. O. D. G., Piepho, H. P., Guimarães, L. J. M., Guimarães, P. D. O., Parentoni, S. N., Pinto, M. D. O., Noda, R. W., Magalhães, J. V., Guimarães, C. T., Garcia, A. A. F., Pastina, M. M. (2020). Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theoretical and Applied Genetics, 133, 443-455. https://doi.org/10.1007/s00122-019-03475-1

Domergue, F., Vishwanath, S. J., Joubès, J., Ono, J., Lee, J. A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., Lessire, R., Rowland, O. (2010). Three Arabidopsis fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. Plant physiology, 153(4), 1539-1554. https://doi.org/10.1104/pp.110.158238

Dudley, J. W., Lambert, R. J. (2004). 100 generations of selection for oil and protein in corn. Plant breeding reviews, 24(1), 79-110. https://doi.org/10.1002/9780470650240.ch5

Dudley, J. W., Clark, D., Rocheford, T. R., LeDeaux, J. R. (2007). Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of IHP× ILP. Crop Science, 47(1), 45-57. https://doi.org/10.2135/cropsci2006.03.0207

Dwiningsih, Y., Al-Kahtani, J. (2022). Genome-wide association study of complex traits in maize detects genomic regions and genes for increasing grain yield and grain quality. Advance Sustainable Science Engineering and Technology, 4(2), 0220209. https://doi.org/10.26877/asset.v4i2.12678

Egesel, C. Ö., Kahrıman, F., Ekinci, N., Kavdır, İ., Büyükcan, M. B. (2016). Analysis of fatty acids in kernel, flour, and oil samples of maize by NIR spectroscopy using conventional regression methods. Cereal Chemistry, 93(5), 487-492. https://doi.org/10.1094/CCHEM-12-15-0247-R

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloSOne 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Fang, H., Fu, X., Ge, H., Zhang, A., Shan, T., Wang, Y., Li, P., Wang, B. (2021). Genetic basis of maize kernel oil-related traits revealed by high-density SNP markers in a recombinant inbred line population. BMC Plant Biology, 21, 1-12. https://doi.org/10.1186/s12870-021-03089-0

Fu, J., Hao, Y., Li, H., Reif, J. C., Chen, S., Huang, C., Wang, G., Li, X., Xu, Y., Li, L. (2022). Integration of genomic selection with doubled-haploid evaluation in hybrid breeding: From GS 1.0 to GS 4.0 and beyond. Molecular Plant, 15(4), 577-580. https://doi.org/10.1016/j.molp.2022.02.005

Galli, G., Alves, F. C., Morosini, J. S., Fritsche-Neto, R. (2020). On the usefulness of parental lines GWAS for predicting low heritability traits in tropical maize hybrids. PloS one, 15(2), e0228724. https://doi.org/10.1371/journal.pone.0228724

Hallauer, A. R., Carena, M. J., Filho, J. D. M. (2010). Quantitative genetics in maize breeding. Springer Science & Business Media.

Hallauer, A. R., Carena, M. J. (2014). Adaptation of tropical maize germplasm to temperate environments. Euphytica, 196, 1-11. https://doi.org/10.1007/s10681-013-1017-9

Harrell Jr, F. R., Dupont, C. (2023). Package 'Hmisc', Harrell Miscellaneous. R package version 5.0-1. https://cran.r-project.org/web/packages/Hmisc/index.html

He, M., Perfield, K. L., Green, H. B., Armentano, L. E. (2012). Effect of dietary fat blend enriched in oleic or linoleic acid and monensin supplementation on dairy cattle performance, milk fatty acid profiles, and milk fat depression. Journal of dairy science, 95(3), 1447-1461. https://doi.org/10.3168/jds.2011-4635

He, M., Armentano, L. E. (2011). Effect of fatty acid profile in vegetable oils and antioxidant supplementation on dairy cattle performance and milk fat depression. Journal of dairy science, 94(5), 2481-2491. https://doi.org/10.3168/jds.2010-3755

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian J. Stat 6, 65–70.

Imai, H., Nishida, I., Murata, N. (1992). Acyl-(acyl-carrier protein) hydrolase from squash cotyledons specific to long-chain fatty acids: purification and characterization. Plant molecular biology, 20, 199-206. https://doi.org/10.1007/BF00014488

Jarquin, D., De Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., Edwards, J., Ertl, D., Flint-Garcia, S., Gore, M. A., Graham, C., Hirsch, C. N., Holland, J. B., Hooker, D., Kaeppler, S. M., Knoll, J., Lee, E. C., Lawrence-Dill, C. J., Lynch, J. P., Moose, S. P., Murray, S. C., Nelson, R., Rocheford, T., Schnable, J. C., Schnable, P. S., Smith, M., Springer, N., Thomison, P., Tuinstra, M., Wisser, R. J., Xu, W., Yu, J., Lorenz, A. (2021). Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. Frontiers in genetics, 11, 592769. https://doi.org/10.3389/fgene.2020.592769

Kahrıman, F., Egesel, C. Ö., Egesel, B. (2015). A comparative study on changes and relationships of kernel biochemical components in different types of maize. Journal of the American Oil Chemists' Society, 92, 1451-1459. https://doi.org/10.1007/s11746-015-2716-2

Karnatam, K. S., Mythri, B., Un Nisa, W., Sharma, H., Meena, T. K., Rana, P., Vikal, Y., Gowda, M., Dhillon, B. S., Sandhu, S. (2023). Silage maize as a potent candidate for sustainable animal husbandry development—perspectives and strategies for genetic enhancement. Frontiers in Genetics, 14, 1150132. https://doi.org/10.3389/fgene.2023.1150132

Khan, N. A., Cone, J. W., Fievez, V., Hendriks, W. H. (2012). Causes of variation in fatty acid content and composition in grass and maize silages. Animal Feed Science and Technology, 174(1-2), 36-45. https://doi.org/10.1016/j.anifeedsci.2012.02.006

Lambert, R. J., Alexander, D. E., Mejaya, I. J. (2004). Single kernel selection for increased grain oil in maize synthetics and high-oil hybrid development. Plant breeding reviews, 24(1), 153-176. https://doi.org/10.1002/9780470650240.ch8

Latimer, S., Li, Y., Nguyen, T. T., Soubeyrand, E., Fatihi, A., Elowsky, C. G., Block, A., Pichersky, E., Basset, G. J. (2018). Metabolic reconstructions identify plant 3-

methylglutaconyl-CoA hydratase that is crucial for branched-chain amino acid catabolism in mitochondria. The Plant Journal, 95(2), 358-370. https://doi.org/10.1111/tpj.13955

Laude, T. P., Carena, M. J. (2014). Diallel analysis among 16 maize populations adapted to the northern US Corn Belt for grain yield and grain quality traits. Euphytica, 200, 29-44. https://doi.org/10.1007/s10681-014-1125-1

Launhardt, L., Uhlenberg, J., Stellmach, H., Schomburg, M., Hause, B., Heilmann, I., Heilmann, M. (2023). Association of the Arabidopsis oleoyl Δ12-desaturase FAD2 with pre-cis-Golgi stacks at endoplasmic reticulum-Golgi-exit sites. The Plant Journal. 117, 242–263. https://doi.org/10.1111/tpj.16492

Laurie, C. C., Chasalow, S. D., LeDeaux, J. R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T. R., Dudley, J. W. (2004). The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics, 168(4), 2141-2155. https://doi.org/10.1534/genetics.104.029686

Lehermeier, C., Teyssèdre, S., Schön, C. C. (2017). Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. Genetics, 207(4), 1651-1661. https://doi.org/10.1534/genetics.117.300403

Li, L., Li, H., Li, Q., Yang, X., Zheng, D., Warburton, M., Chai, Y. C., Zhang, P., Guo, Y. Q., Yan, J. B., Li, J. (2011). An 11-bp insertion in *Zea mays fatb* reduces the palmitic acid content of fatty acids in maize grain. Plos one, 6(9), e24699. https://doi.org/10.1371/journal.pone.0024699

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M. L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., Yan, J. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nature genetics, 45(1), 43-50. https://doi.org/10.1038/ng.2484

Li, H., Thrash, A., Tang, J. D., He, L., Yan, J., Warburton, M. L. (2019). Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. The Plant Journal, 98(5), 853-863. https://doi.org/10.1111/tpj.14282

Li, G., Dong, Y., Zhao, Y., Tian, X., Würschum, T., Xue, J., Chen, S., Reif, J. C., Xu, S., Liu, W. (2020). Genome-wide prediction in a hybrid maize population adapted to Northwest China. The Crop Journal, 8(5), 830-842. https://doi.org/10.1016/j.cj.2020.04.006

Li, J., Cheng, D., Guo, S., Chen, C., Wang, Y., Zhong, Y., Qi, X., Liu, Z., Wang, D., Wang, Y., Liu, W., Liu, C., Chen, S. (2023). Genome-wide association and genomic prediction for resistance to southern corn rust in DH and testcross populations. Frontiers in Plant Science, 14, 1109116. https://doi.org/10.3389/fpls.2023.1109116

Liu, Z., Wang, Y., Ren, J., Mei, M., Frei, U.K.B., Trampe, B., Lübberstedt, T. (2016a). Maize doubled haploids. Plant Breed. Rev. 40, 123. https://doi.org/10.1002/9781119279723.ch3

Liu, X., Huang, M., Fan, B., Buckler, E. S., Zhang, Z. (2016b). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics, 12(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Luo, P., Wang, H., Ni, Z., Yang, R., Wang, F., Yong, H., Zhang, L., Zhou, Z., Song, W., Li, M., Yang, J., Weng, J., Meng, Z., Zhang, D., Han, J., Chen, Y., Zhang, R., Wang, L., Zhao, M., Gao, W., Chen, X., Li, W., Hao, Z., Fu, J., Zhang, X., Li, X. (2023). Genomic prediction of yield performance among single-cross maize hybrids using a partial diallel cross design. The Crop Journal, 11(6), 1884-1892. https://doi.org/10.1016/j.cj.2023.09.009

Ma, L., Qing, C., Frei, U., Shen, Y., and Lübberstedt, T. (2020). Association mapping for root system architecture traits under two nitrogen conditions in germplasm enhancement of maize doubled haploid lines. Crop J. 8 (2), 213–226. https://doi.org/10.1016/ j.cj.2019.11.004

Mangolin, C. A., De Souza, C. L., Garcia, A. A. F., Garcia, A. F., Sibov, S. T., De Souza, A. P. (2004). Mapping QTLs for kernel oil content in a tropical maize population. Euphytica, 137, 251-259. https://doi.org/10.1023/B:EUPH.0000041588.95689.47

Marulanda, J. J., Mi, X., Melchinger, A. E., Xu, J. L., Würschum, T., & Longin, C. F. H. (2016). Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. Theoretical and applied genetics, 129, 1901-1913. https://doi.org/10.1007/s00122-016-2748-5

Mohammadi, M., Tiede, T., Smith, K. P. (2015). PopVar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. Crop Science, 55(5), 2068-2077. https://doi.org/10.2135/cropsci2015.01.0030

Môro, G. V., Santos, M. F., Bento, D. A. V., Aguiar, A. M., de Souza, C. L. (2012). Genetic analysis of kernel oil content in tropical maize with design III and QTL mapping. Euphytica, 185, 419-428. https://doi.org/10.1007/s10681-011-0604-x

Meuwissen, T. H., Hayes, B. J., Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. genetics, 157(4), 1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Narayanan, S. P., Liao, P., Taylor, P. W., Lo, C., Chye, M. L. (2019). Overexpression of a Monocot Acyl-CoA-Binding Protein Confers Broad-Spectrum Pathogen Protection in a Dicot. Proteomics, 19(12), 1800368. https://doi.org/10.1002/pmic.201800368

Neyhart, J. L., Smith, K. P. (2019). Validating genomewide predictions of genetic variance in a contemporary breeding program. Crop Science, 59(3), 1062-1072. https://doi.org/10.2135/cropsci2018.11.0716

Ndlovu, N., Spillane, C., McKeown, P. C., Cairns, J. E., Das, B., Gowda, M. (2022). Genome-wide association studies of grain yield and quality traits under optimum and low-nitrogen

stress in tropical maize (*Zea mays* L.). Theoretical and Applied Genetics, 135(12), 4351-4370. https://doi.org/10.1007/s00122-022-04224-7

Nuss, E. T., Tanumihardjo, S. A. (2010). Maize: a paramount staple crop in the context of global nutrition. Comprehensive reviews in food science and food safety, 9(4), 417-436. https://doi.org/10.1111/j.1541-4337.2010.00117.x

Ortíz-Islas, S., García-Lara, S., Preciado-Ortíz, R. E., Serna-Saldívar, S. O. (2019). Fatty acid composition and proximate analysis of improved high-oil corn double haploid hybrids adapted to subtropical areas. Cereal Chemistry, 96(2), 182-192. https://doi.org/10.1002/cche.10109

Qiao, P., Bourgault, R., Mohammadi, M., Matschi, S., Philippe, G., Smith, L. G., Gore, M. A., Molina, I., Scanlon, M. J. (2020). Transcriptomic network analyses shed light on the regulation of cuticle development in maize leaves. Proceedings of the National Academy of Sciences, 117(22), 12464-12471. https://doi.org/10.1073/pnas.2004945117

Pamin, K., Compton, W. A., Walker, C. E., Alexander, D. E. (1986). Genetic Variation and Selection Response for Oil Composition in Corn 1. Crop science, 26(2), 279-28. https://doi.org/10.2135/cropsci1986.0011183X002600020013x

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228. https://doi.org/10.1007/s10681-007-9449-8

Prasanna, B. M., Palacios-Rojas, N., Hossain, F., Muthusamy, V., Menkir, A., Dhliwayo, T., Ndhlela, T., Vicente, F. S., Nair, S. K., Vivek, B. S., Zhang, X., Olsen, M., Fan, X. (2020). Molecular breeding for nutritionally enriched maize: status and prospects. Frontiers in genetics, 10, 509029. https://doi.org/10.3389/fgene.2019.01392

Pollak, L. M. (2003). The history and success of the public–private project on germplasm enhancement of maize (GEM). Adv. Agron. 78, 45–87. https://doi.org/10.1016/S0065 2113(02)78002-4

Pollak, L. M., Salhuana, W. (2001). The Germplasm Enhancement of Maize (GEM) project: private and public sector collaboration. In Cooper, H. D., Spillane, C., Hodgkin, T. (eds.), Broadening the genetic base of crop production (CABI Publishing). p.319-329. https://doi.org/10.1079/9780851994116.0319

Ray, K., Banerjee, H., Dutta, S., Hazra, A. K., Majumdar, K. (2019). Macronutrients influence yield and oil quality of hybrid maize (*Zea mays* L.). PloS one, 14(5), e0216939. https://doi.org/10.1371/journal.pone.0216939

Resende, M. D. V. (2007). Selegen-REML/BLUP: statistical system and computerized genetic selection via linear mixed models (Colombo: Embrapa Florestas).

Röber, F.K., Gordillo, G.A., Geiger, H.H. (2005). In vivo haploid induction in maize-performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica, 50, 275-283.

Rogers, A. R., Bian, Y., Krakowsky, M., Peters, D., Turnbull, C., Nelson, P., Holland, J. B. (2022). Genomic prediction for the germplasm enhancement of maize project. The Plant Genome, 15(4), e20267. https://doi.org/10.1002/tpg2.20267

Ruiz, A., Trifunovic, S., Eudy, D. M., Sciarresi, C. S., Baum, M., Danalatos, G. J., Elli, E. F., Kalogeropoulos, G., King, K., dos Santos, C., Thies, A., Pico, L. O., Castellano, M. J., Schnable, P. S., Topp, C., Graham, M., Lamkey, K. R., Vyn, T. J., Archontoulis, S. V. (2023). Harvest index has increased over the last 50 years of maize breeding. Field Crops Research, 300, 108991. https://doi.org/10.1016/j.fcr.2023.108991

Salhuana, W., Pollak, L. (2006). Latin American maize project (LAMP) and germplasm enhancement of maize (GEM) project: generating useful breeding germplasm. Maydica 51 (2), 339–355.

Sanchez, D.L., Liu, S., Ibrahim, R., Blanco, M., Lübberstedt, T. (2018). Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). Plant Sci. 268, 30-38. https://doi.org/10.1016/j.plantsci.2017.12.004

Sanchez, D. L., Santana, A. S., Morais, P. C., Peterlini, E., Castellano, M. J., Blanco, M., & Lubberstedt, T. Phenotypic and genome-wide association analyses for nitrogen use efficiency related traits in maize (*Zea mays* L.) exotic introgression lines. Frontiers in Plant Science, 14, 1270166. https://doi.org/10.3389/fpls.2023.1270166

Saoussem, H., Sadok, B., Habib, K., Mayer, P. M. (2009). Fatty acid accumulation in the different fractions of the developing corn kernel. Food chemistry, 117(3), 432-437. https://doi.org/10.1016/j.foodchem.2009.04.038

Schnell, F. W., Utz, H. F. (1975) F1-leistung und elternwahl euphy-der züchtung von selbstbefruchtern, 243–248 in Bericht über die Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter. BAL Gumpenstein, Gumpenstein, Austria.

Sharma, S., Carena, M. J. (2012). NDSU EarlyGEM: Increasing the genetic diversity of northern US hybrids through the development of unique exotic elite lines. Maydica, 57, 34-42.

Singh, N., Vasudev, S., Yadava, D. K., Chaudhary, D. P., Prabhu, K. V. (2014). Oil improvement in maize: potential and prospects. Maize: nutrition dynamics and novel uses, 77-82. https://doi.org/10.1007/978-81-322-1623-0_6

Song, T. M., Chen, S. J. (2004). Long term selection for oil concentration in five maize populations. Maydica, 49, 9–14.

Sukhija, P. S., Palmquist, D. L. (1988). Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. Journal of agricultural and food chemistry, 36(6), 1202-1206.

Tanumihardjo, S. A., McCulley, L., Roh, R., Lopez-Ridaura, S., Palacios-Rojas, N., Gunaratna, N. S. (2020). Maize agro-food systems to ensure food and nutrition security in reference to the Sustainable Development Goals. Global Food Security, 25, 100327. https://doi.org/10.1016/j.gfs.2019.100327

Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics, 197(4), 1343-1355. https://doi.org/10.1534/genetics.114.165860

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T. R., McMullen, M. D., Holland, J. B., Buckler, E. S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature genetics, 43(2), 159-162. https://doi.org/10.1038/ng.746

Tian, Y., Lv, X., Xie, G., Wang, L., Dai, T., Qin, X., Chen, F., Xu, Y. (2019). FAX2 mediates fatty acid export from plastids in developing Arabidopsis seeds. Plant and Cell Physiology, 60(10), 2231-2242. https://doi.org/10.1093/pcp/pcz117

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of dairy science, 91(11), 4414-4423. https://doi.org/10.3168/jds.2007-0980

Vanous, K., Vanous, A., Frei, U. K., Lübberstedt, T. (2017). Generation of maize (*Zea mays*) doubled haploids via traditional methods. Curr. Protoc. Plant Biol. 2(2), 147-157. https://doi.org/10.1002/cppb.20050

Vanous, A., Gardner, C., Blanco, M., Martin-Schwarze, A., Lipka, A. E., Flint-Garcia, S., Bohn, M., Edwards, J., Lübberstedt, T. (2018). Association mapping of flowering and height traits

in germplasm enhancement of maize doubled haploid (GEM-DH) lines. Plant Genome 11 (2), 170083. https://doi.org/10.3835/plantgenome2017.09.0083

Vanous, A., Gardner, C., Blanco, M., Martin-Schwarze, A., Wang, J., Li, X., Lipka, A. E., Flint-Garcia, S., Bohn, M., Edwards, J., Lübberstedt, T. (2019). Stability analysis of kernel quality traits in exotic-derived doubled haploid maize lines. The plant genome, 12(1), 170114. https://doi.org/10.3835/plantgenome2017.12.0114

Wang, X., Tian, X., Zhang, H., Li, H., Zhang, S., Li, H., Zhu, J. (2023). Genome-wide analysis of the maize LACS gene family and functional characterization of the ZmLACS9 responses to heat stress. Plant Stress, 10, 100271. https://doi.org/10.1016/j.stress.2023.100271

Wassom, J. J., Mikkelineni, V., Bohn, M. O., Rocheford, T. R. (2008). QTL for fatty acid composition of maize kernel oil in Illinois High Oil × B73 backcross-derived lines. Crop science, 48(1), 69-78. https://doi.org/10.2135/cropsci2007.04.0208

Watson, S. A. (2003) Description, development, structure and composition of the corn kernel. In: White, P. J., Johnson, L. A. (eds) Corn: chemistry and technology. Am Assn Cereal Chemists Inc St. Paul, MN, pp 69–106.

Wu, X., Li, Y., Shi, Y., Song, Y., Zhang, D., Li, C., Buckler, E. S, Li, Y., Zhang, Z., Wang, T. (2016). Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. Plant Biotechnology Journal, 14(7), 1551-1562. https://doi.org/10.1111/pbi.12519

Yang, Q., Fan, C., Guo, Z., Qin, J., Wu, J., Li, Q., Fu, T., Zhou, Y. (2012). Identification of FAD2 and FAD3 genes in Brassica napus genome and development of allele-specific markers for high oleic and low linolenic acid contents. Theoretical and Applied Genetics, 125, 715-729. https://doi.org/10.1007/s00122-012-1863-1

Yang, X., Guo, Y., Yan, J., Zhang, J., Song, T., Rocheford, T., Li, J. S. (2010). Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil

maize. Theoretical and Applied Genetics, 120, 665-678. https://doi.org/10.1007/s00122-009-1184-1

Yao, J., Zhao, D., Chen, X., Zhang, Y., Wang, J. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). Crop Journal, 6, 353–365. https://doi.org/10.1016/j.cj.2018.05.003

Yu, B. G., Chen, X. X., Zhou, C. X., Ding, T. B., Wang, Z. H., Zou, C. Q. (2022). Nutritional composition of maize grain associated with phosphorus and zinc fertilization. Journal of Food Composition and Analysis, 114, 104775. https://doi.org/10.1016/j.jfca.2022.104775

Zhang, J., Lu, X. Q., Song, X. F., Yan, J. B., Song, T. M., Dai, J. R., Rocheford, T., Li, J. S. (2008). Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. Euphytica, 162, 335-344. https://doi.org/10.1007/s10681-007-9500-9

Zhang, J., Liu, H., Sun, J., Li, B., Zhu, Q., Chen, S., Zhang, H. (2012). Arabidopsis fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. PloS one, 7(1), e30355. https://doi.org/10.1371/journal.pone.0030355

Zhang, X., Ren, Z., Luo, B., Zhong, H., Ma, P., Zhang, H., Hu, H., Wang, Y., Zhang, H., Liu, D., Wu, L., Nie, Z., Zhu, Y., He, W., Zhang, S., Su, S., Shen, Y., Gao, S. (2022). Genetic architecture of maize yield traits dissected by QTL mapping and GWAS in maize. The Crop Journal, 10(2), 436-446. https://doi.org/10.1016/j.cj.2021.07.008

Zhang, X., Wang, M., Guan, H., Wen, H., Zhang, C., Dai, C., Wang, J., Pan, B., Li, J., Liao, H. (2023). Genetic dissection of QTLs for oil content in four maize DH populations. Frontiers in Plant Science, 14, 1174985. https://doi.org/10.3389/fpls.2023.1174985

Zheng, Y., Yuan, F., Huang, Y., Zhao, Y., Jia, X., Zhu, L., Guo, J. (2021). Genome-wide association studies of grain quality traits in maize. Scientific reports, 11(1), 9797. https://doi.org/10.1038/s41598-021-89276-3

Zhong, S., Jannink J. L. (2007). Using QTL results to discriminate among crosses based on their progeny mean and variance. Genetics. 177:567–576. https://doi.org/10.1534/genetics.107.075358

Zuffo, L. T., DeLima, R. O., and Lübberstedt, T. (2022). Combining datasets for maize root seedling traits increases the power of GWAS and genomic prediction accuracies. J. Exp. Bot. 73 (16), 5460–5473. https://doi.org/10.1093/jxb/erac236

Table 1. Estimates of variance components due to genotypes ($\hat{\sigma}^2_g$) and genotypes by environments interaction ($\hat{\sigma}^2_{ge}$), quadratic components of environmental ($\hat{\Phi}^2_e$) and replication within environment ($\hat{\Phi}^2_{e:r}$), residual variance estimates ($\hat{\sigma}^2$), best linear unbiased prediction estimates of ranges and means, means of BGEM lines, and means of PHB47 and PHZ51 lines, coefficient of variation (CV%) and broad sense heritability estimates ($\hat{h}^2_{\bar{x}}$) for kernel oil content and fatty acid composition in BGEM lines across three environments.

Parameter	Traits										
	Oil	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1	C22:0	C24:0
$\hat{\sigma}^2_g$	0.18*	2.32*	0.0011*	0.08*	25.68*	33.59*	0.044*	0.0043*	0.0029*	0.0012*	0.0021*
$\hat{\sigma}^2_{ge}$	0.03*	0.06*	0.0001*	0.01*	3.03*	2.84*	0.004*	0.0004*	0.0002*	0.0002*	0.0003*
$\hat{\Phi}^2_e$	0.07	4.93*	0.0164*	2.19*	370.02*	567.03*	0.140*	0.1807*	0.0209*	0.0650*	0.1499*
$\hat{\Phi}^2_{e:r}$	0.22*	0.33*	0.0008*	0.02	3.01	3.09	0.036*	0.0009	0.0009*	0.0031*	0.0034*
$\hat{\sigma}^2$	0.05	0.13	0.0002	0.01	2.73	3.04	0.009	0.0007	0.0003	0.0003	0.0005
General mean (range)[2/]	2.42 (1.32-3.35)	11.81 (6.93-15.99)	0.15 (0.08-0.27)	1.51 (0.87-2.54)	25.58 (16.43-43.12)	58.48 (39.57-67.70)	1.25 (0.69-1.98)	0.46 (0.32-0.72)	0.35 (0.22-0.50)	0.17 (0.11-0.34)	0.23 (0.14-0.43)
BGEM mean	2.42	11.81	0.15	1.52	25.59	58.47	1.25	0.46	0.35	0.17	0.23
PHB47	2.18	11.99	0.13	1.43	29.49	54.45	1.18	0.47	0.41	0.18	0.23
PHZ51	1.72	10.63	0.23	1.34	19.55	65.62	1.49	0.42	0.31	0.21	0.25
CV%	9.08	3.11	8.11	6.92	6.45	2.98	7.43	5.89	5.09	9.45	9.30
$\hat{h}^2_{\bar{x}}$	0.92	0.99	0.96	0.96	0.95	0.96	0.95	0.95	0.96	0.93	0.93

[1/] Significant at the 0.01 probability level by the likelihood ratio test for random effect and F-test for fixed effect. 2/BLUP values of means and ranges.

Table 2. Estimates of variance components due to genotypes ($\hat{\sigma}^2_g$) and genotypes by environments interaction ($\hat{\sigma}^2_{ge}$), quadratic components of environmental ($\hat{\Phi}^2_e$) and replication within environment ($\hat{\Phi}^2_{e:r}$), best linear unbiased prediction estimates of ranges and means, means of testcross hybrids from crossing among BGEM lines and PHB47 (TC_PHB47), and BGEM lines and PHZ51 (TC_PHB47) testers, coefficient of variation (CV%) and broad sense heritability estimates ($\hat{h}^2_{\bar{x}}$) for kernel oil content and fatty acid composition in testcross hybrids across three environments.

Parameter	Traits										
	Oil	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1	C22:0	C24:0
$\hat{\sigma}^2_g$	0.07*	0.42*	0.0003*	0.02*	4.77*	5.93*	0.013*	0.0008*	0.0010*	0.0002*	0.0003*
$\hat{\sigma}^2_{ge}$	0.02*	0.05	0.0001*	0.01*	0.60	0.69*	0.005*	0.0004*	0.0002*	0.0002*	0.0003*
$\hat{\Phi}^2_e$	3.53*	6.43	0.0452*	0.66*	18.65	6.55	0.055	0.0655*	0.0015	0.0196*	0.0391*
$\hat{\Phi}^2_{e:r}$	2.06*	17.71*	0.0174*	0.50*	30.42	54.04	0.385*	0.0948*	0.0313*	0.0451*	0.1262*
$\hat{\sigma}^2$	0.08	1.13	0.0003	0.02	5.59	6.70	0.015	0.0014	0.0008	0.0004	0.0006
General mean (range)[2/]	2.28 (1.56-3.09)	11.58 (10.34-13.56)	0.17 (0.12-0.23)	1.39 (1.10-1.78)	22.63 (19.03-31.25)	61.80 (51.62-66.40)	1.28 (0.87-1.53)	0.42 (0.37-0.51)	0.34 (0.30-0.39)	0.18 (0.12-0.30)	0.21 (0.16-0.26)
TC_PHB47	2.33	11.65	0.16	1.39	23.43	60.96	1.24	0.43	0.34	0.19	0.20
TC_PHZ51	2.25	11.54	0.18	1.39	22.10	62.36	1.30	0.42	0.34	0.17	0.21
CV%	11.99	9.17	10.59	10.11	10.45	4.19	9.73	8.73	8.26	11.30	12.27
$\hat{h}^2_{\bar{x}}$	0.77	0.67	0.80	0.81	0.81	0.82	0.75	0.69	0.82	0.62	0.63

[1/] Significant at the 0.01 probability level by the likelihood ratio test for random effect and F-test for fixed effect. 2/BLUP values of means and ranges.

Table 3. SNPs significantly associated with kernel oil content and fatty acid composition traits in BGEM lines across three environments.

Trait	SNP	Chr.	Position	P-value	MAF[1/]	PVE[2/]
Oil	S1_54739371	1	54,739,371	1.20E-07	0.165	11.83
	S4_136639469	4	136,639,469	2.77E-07	0.452	0.21
	S5_71851438	5	71,851,438	3.82E-14	0.405	4.62
	S7_145632437	7	145,632,437	3.59E-07	0.410	0.71
C16:0	S1_8626591	1	8,626,591	1.22E-11	0.148	3.31
	S4_188829006	4	188,829,006	3.79E-07	0.383	5.06
	S6_3870870	6	3,870,870	6.24E-10	0.482	2.01
	S9_14537291	9	14,537,291	2.35E-23	0.491	11.50
	S9_20337850	9	20,337,850	7.35E-12	0.104	7.12
	S9_63146049	9	63,146,049	4.25E-08	0.390	2.81
C16:1	S1_41102630	1	41,102,630	8.43E-10	0.476	2.35
	S2_196221631	2	196,221,631	5.22E-16	0.425	7.78
	S3_183557621	3	183,557,621	6.62E-07	0.117	3.56
	S3_189459204	3	189,459,204	1.00E-08	0.458	1.90
	S5_206079600	5	206,079,600	4.92E-07	0.081	2.34
	S5_214938579	5	214,938,579	2.69E-08	0.441	5.02
C18:0	S3_231891794	3	231,891,794	1.84E-09	0.134	35.44
C18:1	S1_188717024	1	188,717,024	1.14E-09	0.335	2.46
	S4_94817189	4	94,817,189	1.81E-08	0.066	11.57
	S4_170652253	4	170,652,253	1.52E-08	0.339	5.17
	S6_22725025	6	22,725,025	2.97E-07	0.167	1.22
	S6_96906774	6	96,906,774	3.51E-09	0.361	5.09
	S6_104865219	6	104,865,219	2.51E-07	0.073	11.23
C18:2	S2_42676420	2	42,676,420	8.02E-09	0.441	1.22
	S4_3482817	4	3,482,817	1.36E-10	0.156	2.73
	S4_3515229	4	3,515,229	1.16E-09	0.449	3.34
	S4_163995572	4	163,995,572	4.60E-09	0.476	3.33
	S4_167081234	4	167,081,234	5.11E-19	0.132	5.61
	S6_91640544	6	91,640,544	3.32E-07	0.172	1.54
	S6_96906774	6	96,906,774	1.55E-07	0.361	2.64
	S6_104865219	6	104,865,219	1.75E-23	0.073	18.45
C18:3	S1_11944222	1	11,944,222	2.28E-07	0.141	2.84
	S3_6165571	3	6,165,571	1.32E-07	0.057	4.48
	S3_177676923	3	177,676,923	8.53E-10	0.471	2.56
	S4_53675067	4	53,675,067	2.48E-11	0.498	2.82
	S5_5619518	5	5,619,518	1.09E-08	0.399	1.70
	S6_26578652	6	26,578,652	3.11E-08	0.339	2.63
	S7_156209692	7	156,209,692	1.64E-15	0.421	6.47
	S8_138317908	8	138,317,908	7.05E-11	0.436	2.87
C20:0	S1_32868703	1	32,868,703	4.77E-09	0.090	4.35
	S1_64414894	1	64,414,894	2.93E-07	0.154	1.93
	S2_211497872	2	211,497,872	2.99E-11	0.143	4.00
	S3_170120146	3	170,120,146	2.79E-14	0.227	4.74
	S5_195306550	5	195,306,550	1.83E-10	0.392	1.72
	S6_96186343	6	96,186,343	1.33E-07	0.222	2.76
	S6_102939524	6	102,939,524	3.25E-14	0.372	3.82
C20:1	S3_205818128	3	205,818,128	2.43E-08	0.086	6.83
	S5_212682568	5	212,682,568	5.56E-09	0.194	1.48
	S6_93190536	6	93,190,536	1.60E-14	0.390	3.38
	S6_101228022	6	101,228,022	1.50E-07	0.233	2.14
	S7_165815642	7	165,815,642	1.53E-09	0.474	2.58
	S8_165773305	8	165,773,305	6.34E-07	0.465	1.49
	S10_145015431	10	145,015,431	2.97E-09	0.441	2.54
C22:0	S1_262811984	1	262,811,984	5.71E-08	0.165	2.10
	S2_201192162	2	201,192,162	8.37E-09	0.346	4.96
	S5_1726939	5	1,726,939	7.32E-09	0.388	3.36
	S5_200672739	5	200,672,739	1.61E-07	0.361	5.20
	S6_107477545	6	107,477,545	2.12E-13	0.405	3.72
	S8_138882450	8	138,882,450	1.80E-07	0.443	0.96
	S9_104885901	9	104,885,901	9.01E-10	0.328	2.14
C24:0	S3_229905382	3	229,905,382	1.22E-08	0.401	5.32
	S4_188300113	4	188,300,113	4.55E-07	0.427	3.98
	S6_104450979	6	104,450,979	9.43E-08	0.163	3.32
	S6_105833721	6	105,833,721	3.96E-21	0.456	11.42
	S6_121390584	6	121,390,584	4.03E-07	0.436	5.51

[1/]MAF: Minor allele frequency; [2/]PVE: Phenotypic variance explained

Table 4. Putative candidate genes associated with kernel oil content and fatty acid composition traits in BGEM lines across three environments.

Trait	SNP	B73 Gene ID[1/]	Zm Gene ID[1/]	Function[2/]
Oil	S7_145632437	GRMZM2G020740	Zm00001eb318810	kcs38 - 3-ketoacyl-CoA synthase38 (EC 2.3.1.199)
C16:0	S9_20337850	GRMZM5G812228	Zm00001eb376490	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3); 4-coumarate--CoA ligase (EC 6.2.1.12)
C16:1	S3_189459204	GRMZM2G126397	Zm00001eb149490	plt3 - phospholipid transfer protein homolog3
C18:0	S3_231891794	GRMZM2G430729	Zm00001eb162340	fad14 - fatty acid desaturase14
C18:2	S4_167081234	GRMZM2G064701	Zm00001eb188990	fad2 - fatty acid desaturase2
C20:0	S1_32868703	GRMZM2G108138	Zm00001eb010410	acb5 - Acyl-CoA-binding protein5
	S5_195306550	GRMZM2G398500	Zm00001eb249240	IDP290 - Methylglutaconyl-CoA hydratase

[1/]Based on B73 RefGen_v5. [2/]Obtained from MaizeGDB and NCBI.

Table 5. SNPs significantly associated with kernel oil content and fatty acid composition traits using additive GWAS model in testcross hybrids across three environments.

Trait	SNP	Chr.	Position	P-value	MAF[1/]	PVE[2/]
C16:0	S8_174215468	8	174,215,468	6.92E-08	0.115	1.77
	S9_14537291	9	14,537,291	7.09E-10	0.465	45.65
C16:1	S1_17725254	1	17,725,254	8.16E-09	0.088	6.33
	S5_172565790	5	172,565,790	1.63E-07	0.492	11.69
	S8_165878940	8	165,878,940	4.85E-07	0.249	0.87
	S10_43965032	10	43,965,032	1.62E-07	0.096	2.62
C18:0	S1_72805198	1	72,805,198	1.68E-08	0.058	11.47
	S1_238342155	1	238,342,155	3.60E-07	0.053	13.73
	S2_42676420	2	42,676,420	2.20E-07	0.497	6.43
	S5_194647704	5	194,647,704	1.32E-08	0.500	5.71
	S8_165995519	8	165,995,519	2.68E-07	0.080	4.75
C18:1	S1_103826231	1	103,826,231	6.31E-08	0.471	0.62
	S5_191770094	5	191,770,094	8.86E-10	0.064	5.66
	S6_104865219	6	104,865,219	4.83E-14	0.050	44.39
C18:2	S6_104865219	6	104,865,219	2.15E-18	0.050	45.99
	S10_70535885	10	70,535,885	3.41E-08	0.457	22.90
C18:3	S2_45526342	2	45,526,342	1.16E-08	0.102	32.24
	S2_144628433	2	144,628,433	2.71E-08	0.086	2.78
	S6_62076466	6	62,076,466	1.25E-07	0.179	4.29
C20:0	S1_296023753	1	296,023,753	4.42E-07	0.484	5.02
	S2_219156069	2	219,156,069	2.92E-09	0.484	3.96
	S4_237390228	4	237,390,228	1.11E-08	0.452	14.50
	S5_197958120	5	197,958,120	1.25E-09	0.484	6.40
	S6_102939524	6	102,939,524	4.16E-07	0.479	10.73
	S10_148116133	10	148,116,133	4.87E-07	0.487	10.60
C20:1	S1_19790681	1	19,790,681	6.97E-08	0.422	4.06
	S6_88743568	6	88,743,568	5.33E-07	0.492	17.37
	S6_110285241	6	110,285,241	7.00E-09	0.495	28.38
	S8_172692977	8	172,692,977	1.02E-08	0.473	2.19
C24:0	S6_109292729	6	109,292,729	6.07E-08	0.479	3.11
	S9_133095146	9	133,095,146	6.43E-07	0.460	1.96
	S10_148107187	10	148,107,187	3.30E-07	0.390	12.69

[1/]MAF: Minor allele frequency; [2/]PVE: Phenotypic variance explained

Table 6. Putative additive and dominances candidate genes associated with kernel oil content and fatty acid composition traits in testcross hybrids across three environments.

Trait	SNP	B73 Gene ID [1/]	Zm Gene ID[1/]	Function[2/]
		Additive candidate genes		
C18:0	S1_238342155	GRMZM2G167438	Zm00001eb046050	ad1 - adherent1 - 3-ketoacyl-CoA synthase (EC 2.3.1.-); very-long-chain 3-oxoacyl-CoA synthase (EC 2.3.1.199)
C20:0	S10_148116133	GRMZM2G143389	Zm00001eb432380	fax2 - fatty acid export2
C24:0	S10_148107187	GRMZM2G143389	Zm00001eb432380	fax2 - fatty acid export2
		Dominance candidate genes		
C18:0	S7_70916715	GRMZM2G480516	Zm00001eb308150	Fatty acyl-CoA reductase 1 (EC 1.2.1.84)
C20:1	S7_141545772	GRMZM2G700221	Zm00001eb317340	Acyl-[acyl-carrier-protein] hydrolase (EC 3.1.2.14)

[1/] Based on B73 RefGen_v5. [2/] Obtained from MaizeGDB and NCBI.

Table 7. SNPs significantly associated with kernel oil content and fatty acid composition traits using additive-dominance GWAS model in testcross hybrids across three environments.

Trait	SNP	Chr.	Position	P-value	MAF[1/]	PVE[2/]
Oil	S3_170538433	3	170,538,433	3.09E-07	0.430	34.97
C16:1	S1_19295971	1	19,295,971	2.75E-07	0.409	3.72
	S1_35213738	1	35,213,738	5.83E-08	0.374	1.98
	S3_194470582	3	194,470,582	3.76E-08	0.052	11.88
	S4_237682064	4	237,682,064	6.21E-13	0.061	7.24
	S7_161424941	7	161,424,941	6.35E-08	0.390	1.93
	S7_164205920	7	164,205,920	3.21E-11	0.398	9.24
	S8_152557100	8	152,557,100	1.05E-08	0.184	7.60
C18:0	S1_72805198	1	72,805,198	4.28E-10	0.048	18.59
	S5_173805505	5	173,805,505	4.96E-07	0.099	6.57
	S6_56233930	6	56,233,930	1.63E-07	0.050	5.18
	S7_70916715	7	70,916,715	1.09E-08	0.078	3.03
	S7_173272605	7	173,272,605	6.62E-07	0.091	2.26
	S8_165995519	8	165,995,519	7.17E-07	0.080	4.95
C18:1	S1_195488486	1	195,488,486	5.59E-10	0.406	3.70
	S1_298075461	1	298,075,461	5.74E-07	0.414	7.80
	S4_165841527	4	165,841,527	2.62E-07	0.417	2.68
	S6_104865219	6	104,865,219	1.66E-09	0.040	34.21
C18:2	S1_174669769	1	174,669,769	6.58E-08	0.104	2.28
	S2_188534051	2	188,534,051	7.84E-07	0.064	0.87
	S6_27803456	6	27,803,456	3.55E-08	0.064	0.99
	S6_104865219	6	104,865,219	7.53E-09	0.050	24.26
	S10_79408978	10	79,408,978	1.08E-18	0.118	10.72
	S10_82025026	10	82,025,026	2.08E-07	0.091	8.30
C18:3	S1_2803229	1	2,803,229	2.55E-07	0.388	1.09
	S2_3003244	2	3,003,244	1.13E-07	0.160	0.23
	S5_199739716	5	199,739,716	1.47E-08	0.045	1.88
	S9_127208636	9	127,208,636	4.85E-18	0.050	54.90
	S10_145097505	10	145,097,505	3.41E-13	0.050	33.42
C20:1	S1_42853941	1	42,853,941	1.09E-10	0.123	6.36
	S5_215380245	5	215,380,245	5.23E-08	0.401	5.23
	S6_51447615	6	51,447,615	7.36E-11	0.430	12.92
	S6_93496589	6	93,496,589	1.47E-10	0.396	0.17
	S7_141545772	7	141,545,772	1.90E-08	0.420	0.00
	S9_151733235	9	151,733,235	1.45E-07	0.102	4.30

[1/]MAF: Minor allele frequency; [2/]PVE: Phenotypic variance explained

Table 8. Estimates of variance components due to general combining ability of NSS lines ($\sigma^2_{GCA\_NSS}$), general combining ability of SS lines ($\sigma^2_{GCA\_SS}$) and specific combining ability (σ^2_{SCA}), estimates of residual variance (σ^2), and genomic heritability associated to general ($h^2_{GCA\_NSS}$ and $h^2_{GCA\_SS}$) and specific combining ability (h^2_{SCA}) for genomic prediction using additive and additive-dominance models for kernel oil content and fatty acid composition traits in testcross hybrids across three environments.

Parameters	Oil	C16:0	C18:1	C18:2	C20:0	C24:0
			Additive genomic model			
$\sigma^2_{GCA\_NSS}$	0.0361**[1/]	0.4014**	2.0858**	3.5793**	0.0004**	0.0004**
$\sigma^2_{GCA\_SS}$	0.0044	0.2226**	0.4011**	0.7045**	0.0001	0.0001
σ^2	0.0560	0.1438	3.2727	3.2463	0.0006	0.0002
$h^2_{GCA\_NSS}$	0.37	0.52	0.36	0.48	0.36	0.54
$h^2_{GCA\_SS}$	0.05	0.29	0.07	0.09	0.11	0.09
			Additive – dominance genomic model			
$\sigma^2_{GCA\_NSS}$	0.0308*	0.2481*	1.9125**	3.5332**	0.0004**	0.0003**
$\sigma^2_{GCA\_SS}$	0.0000	0.1067	0.2567	0.6658	0.0001	0.0000
σ^2_{SCA}	0.0095	0.4044	0.4338	0.1091	0.0000	0.0001
σ^2	0.0560	0.0783	3.1927	3.2301	0.0006	0.0002
$h^2_{GCA\_NSS}$	0.32	0.30	0.33	0.47	0.36	0.50
$h^2_{GCA\_SS}$	0.00	0.13	0.04	0.09	0.11	0.05
h^2_{SCA}	0.10	0.48	0.07	0.01	0.00	0.08

[1/]** and * significant at the 0.01 and 0.05 by the likelihood ratio test (LRT), respectively.

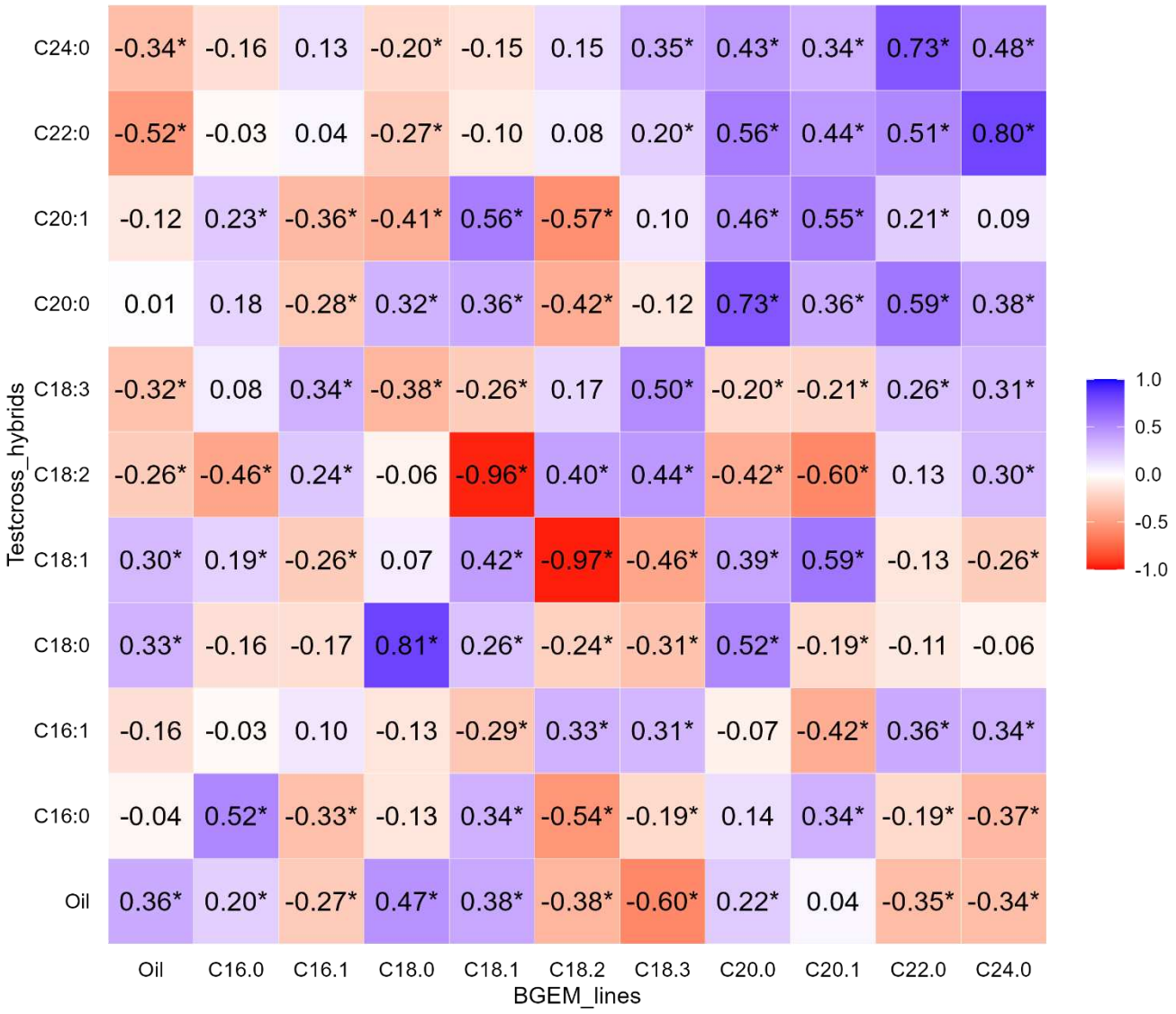

Figure 1. Heat map of estimates of Pearson correlation coefficients between pairs of vectors of BLUE values of the traits measured in BGEM lines (below diagonal) testcross hybrids (above diagonal), and estimates of Spearman correlation coefficients between same traits in BGEM lines and testcross hybrids (diagonal).

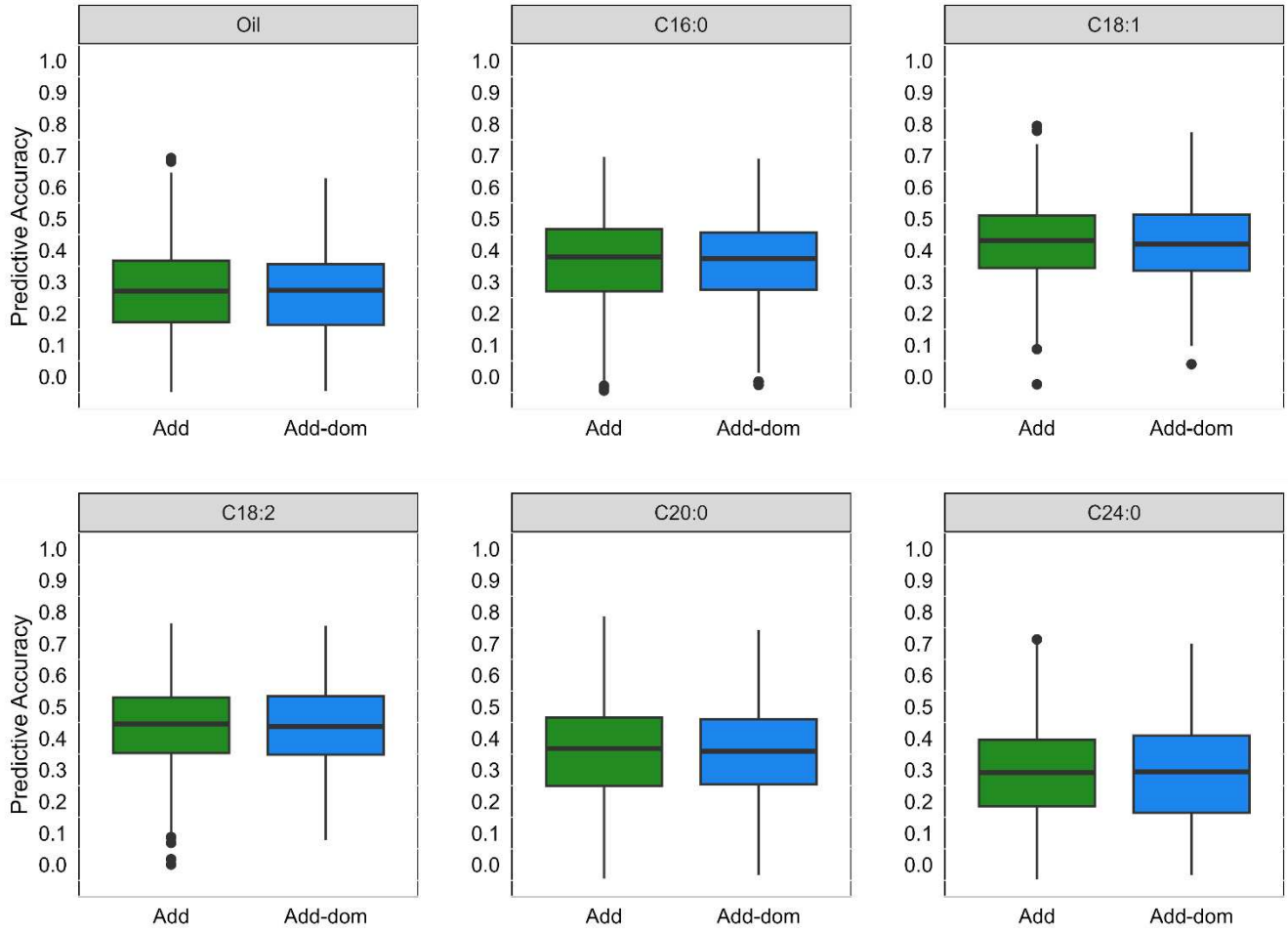

Figure 2. Predictive accuracy by additive and additive-dominance GBLUP model estimated by Pearson's correlation between predicted and observed genotypic values of the validation set for oil content (Oil) and five fatty acids: palmitic acid (C16:0), oleic acid (C18:1), linoleic acid (C18:2), arachidic acid (C20:0) and tetracosanoic acid (C24:0).