University of São Paulo
"Luiz de Queiroz" College of Agriculture

Genomic dissection of a tropical maize diversity panel: a study on molecular characterization and resistance to the corn stunt disease complex

Fernando Garcia Espolador

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2020

Fernando Garcia Espolador
Agronomic Engineer

Genomic dissection of a tropical maize diversity panel: a study on molecular characterization and resistance to the corn stunt disease complex

Advisor:
Prof. Dr. **ROBERTO FRITSCHE NETO**

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Espolador, Fernando Garcia

Genomic dissection of a tropical maize diversity panel: a study on molecular characterization and resistance to the corn stunt disease complex / Fernando Garcia Espolador. - - Piracicaba, 2020.

71 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. *Zea mays* 2. Painel de diversidade 3. Estrutura 4. *Spiroplasma kunkelii* 5. Maize bushy stunt phytoplasma 6. GWAS I. Título

To my parents, Carlos and Maria de Fátima,
and my fianceé, Júlia,
who always supported me

ACKNOWLEDGMENTS

I would like to address my acknowledgments:

To God, for all the blessings, the wisdom, and for guiding my steps on my journey.

To my parents, Carlos Maschio Espolador and Maria de Fátima Garcia Espolador, for their love and support, for being examples for me and the source of my personal values.

To my brother, Murilo Garcia Espolador, for his friendship.

To my fiancée, Julia Silva Morosini, for her love, support, confidence, strength, and for being an example of professional and person to me.

To my parents-in-law, Carlos Alberto Morosini and Olivia Maria B. S. Morosini, and my sister-in-law, Natalia Silva Morosini, for their constant support and friendship.

To the University of São Paulo and College of Agriculture 'Luiz de Queiroz', for the opportunities and for my professional qualification.

To Prof. Dr. Roberto Fritsche Neto and the Allogamous team, for the opportunity to work with maize breeding and for the lessons.

To the University of Wisconsin-Madison, for receiving me as exchange student.

To Prof. Dr. Natalia de Leon, for the opportunity of internship abroad, her confidence, knowledge and for being an example of professionalism.

To the funding agencies. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, National Council for Scientific and Technological Development (CNPq), and São Paulo Research Foundation – FAPESP (Grant 2017/24327-0).

Finally, to all those who directly or indirectly helped in this work, in my personal training and professional qualification, my sincere THANK YOU!

TABLE OF CONTENTS

RESUMO 6
ABSTRACT 7
1. INTRODUCTION 9
REFERENCES 10
2. CHARACTERIZING A TROPICAL MAIZE DIVERSITY PANEL: ENHANCING KNOWLEDGE ABOUT BRAZILIAN MAIZE GERMPLASM 13
ABSTRACT 13
2.1. Introduction 13
2.2. Materials and methods 14
2.3. Results 17
2.4. Discussion 20
2.5. Conclusion 23
References 23
Figures 28
Tables 37
3. DISSECTING THE RESISTANCE TO THE CORN STUNT DISEASE COMPLEX IN A TROPICAL MAIZE DIVERSITY PANEL THROUGH A GENOME WIDE ASSOCIATION STUDY 45
ABSTRACT 45
3.1. Introduction 45
3.2. Materials and methods 46
3.3. Results 49
3.4. Discussion 52
3.5. Conclusion 54
References 55
Figures 61
Tables 65
APPENDIX 67

RESUMO

Análise genômica de um painel de diversidade de milho tropical: um estudo sobre caracterização molecular e resistência ao complexo do enfezamento do milho

O milho tropical é uma importante fonte de diversidade genética a ser explorada por programas de melhoramento para lidar com novos desafios agrícolas, mas isto demanda caracterização genética. Por exemplo, o complexo do enfezamento do milho (enfezamento pálido e enfezamento vermelho) tem causado perdas significativas de produtividade nas últimas safras, contudo o controle genético e o germoplasma resistente para estas doenças são pouco compreendidos. Neste trabalho, um painel de diversidade de milho tropical foi construído e caracterizado geneticamente visando delinear um pool representativo de germoplasma tropical para estudos genéticos e investigar os genes associados à resistência do complexo do enfezamento do milho, assim como as potenciais fontes de resistência. Para isso, 360 linhagens altamente diversas geneticamente foram genotipadas usando a abordagem de genotipagem por sequenciamento com as enzimas de restrição PstI e MseI e alinhadas com a versão 5 do genoma de referência B73. Para as avaliações genéticas, foram considerados dois conjuntos de dados: um com dados brutos e o segundo com dados imputados e filtrados para controle de qualidade, mantendo apenas marcadores bialélicos com frequência alelo raro maior que 0,05, valor de chamada de SNPs maior que 0,95 e desequilíbrio de ligação (r^2) menor que 0,99. O estudo de associação ampla do genoma (GWAS) foi realizado incorporando a matriz de relacionamento genômico e três componentes principais para lidar com a estrutura do painel. Usando o conjunto de dados filtrado, foram analisadas as características proporção de plantas sobreviventes (PSP), nota de sanidade das plantas sobreviventes (SSSP) e nota de sanidade total (WSS), cujas mensurações ocorreram em dois locais no estado de São Paulo, Brasil. O conjunto de dados brutos incluiu 196.803 SNPs bem distribuídos pelos cromossomos. A proporção de dados faltantes foi de 0,391 e a heterozigosidade média observada foi de 0,036. O conjunto de dados filtrados, contendo 14.655 SNPs, apresentou estimativas semelhantes de parâmetros genéticos populacionais em comparação ao primeiro conjunto. A análise da estrutura do painel apontou a existência de nove subpopulações. Através da GWAS, 13 marcadores foram significativos para as características avaliadas e apresentaram funções relacionadas principalmente ao metabolismo da celulose, via da auxina, genes de defesa e resposta associados à atividade de fagócito oxidase e produção de antocianinas. Os genes candidatos foram associados, por exemplo, à resposta do acúmulo de glicose nas folhas, redução do teor de auxina na planta e ataque direto aos patógenos. Para cada característica, foi possível identificar linhagens com todos os alelos favoráveis em homozigose, o que facilitaria a transferência de genes de resistência para outros genótipos. Além disso, as linhagens advindas do ancestral PF-41X05-33-05B apresentaram alta concentração de alelos favoráveis a todas as características simultaneamente. Os presentes resultados revelam mecanismos genéticos subjacentes desencadeados por plantas em resposta ao complexo do enfezamento do milho e permitiram a identificação de potenciais linhagens resistentes. Tais inferências podem melhorar substancialmente os ganhos genéticos quando incorporadas em programas de melhoramento e constituem uma importante contribuição para a compreensão genética do germoplasma de milho tropical.

Palavras-chave: *Zea mays*, Painel de diversidade, *Spiroplasma kunkelii*, Maize bushy stunt phytoplasma

ABSTRACT

Genomic dissection of a tropical maize diversity panel: a study on molecular characterization and resistance to the corn stunt disease complex

The tropical maize is an important pool of genetic diversity to be explored in breeding programs to face new agricultural challenges, nevertheless it requires genetic characterization. For instance, the corn stunt disease complex (corn stunt disease and maize bushy stunt disease) caused significative losses in the last crop seasons, but its genetic control and resistant germplasm are poorly comprehended. In this study, we assembled and genetically characterized a tropical maize diversity panel in order to construct a representative pool of tropical germoplasm for genetic studies and to investigate the genes associated with the resistance of corn stunt disease complex, as well as the potential sources of resistance. For that, 360 inbred lines highly diverse were genotyped using a genotyping-by-sequencing approach with restriction enzymes PstI and MseI and aligned to the version 5 of the B73 reference genome. For genetic assessments, two datasets were considered: one with the raw data, and the second one with data imputed and filtered for quality control, retaining only biallelic markers with minor allelic frequency higher than 0.05, call rate higher than 0.95, and linkage disequilibrium (r2) lower than 0.99. Genome wide association study (GWAS) was performed incorporating the genomic relationship matrix and 3 main principal components to deal with panel structure. Using the filtered dataset, we analyzed the traits proportion of survivor plants (PSP), sanity score in survivor plants (SSSP), and whole sanity score (WSS) which were evaluated in two sites in the State of São Paulo, Brazil. The raw dataset contained 196,803 SNPs well distributed across the chromosomes. The proportion of missing data was 0.391 and the average observed heterozygosity was 0.036. The filtered dataset, containing 14,655 SNPs, showed similar estimates of populational genetic parameters compared to the first one. The structure analysis indicated that this panel comprises nine subpopulations. Through GWAS, 13 markers were significant to the traits and presented functions mainly related to cellulose metabolism, auxin pathway, genes of defense and response to phagocyte oxidase activity, and anthocyanin production. The candidate genes were associated, for example, with the response of glucose accumulation in leaves, the reduction of auxin content, and the direct defense-attack against the pathogens. For each of the traits, we found lines with the totality of favorable alleles in homozygosis, which would facilitate transfering resistance genes to other genotypes. Additionally, the lines from the ancestral PF-41X05-33-05B exhibited high content of favorable alleles to all the traits simultaneously. Our findings reveal underlying genetic mechanisms triggered by plants in response to the corn stunt disease complex and allowed the identification of potential resistant inbred lines. These results can substantially improve the genetic gains when incorporated in breeding programs and constitute an important contribution to the genetic comprehension of the tropical maize germplasm.

Keywords: *Zea mays*, Diversity panel, Structure, *Spiroplasma kunkelii*, Maize bushy stunt phytoplasma

1. INTRODUCTION

Maize (*Zea mays*) is a crop domesticated from its ancestral species teosinte (*Zea teosinte*) by ancient civilizations near Rio Balsas region, Mexico, 8700 years ago (Mir et al., 2013). The natives domesticated this crop by selecting plants that presented only one stalk, bigger ears with more grains, absence of shell around kernels, and indehiscence (Doebley, 2004; Szabó & Burr, 1996). As result of such intense selection process, the derived species, *Zea mays*, had a wide spread across the pre-Columbian societies in America, leading to its dispersion to different environments: from low to high altitudes, ranging from temperate regions of North America, passing by Central American plains until areas in high altitude in Andes (Edmeades et al., 2017; Matsuoka et al., 2002). Consequently, the maize germplasm presents adaptation to cultivation in regions at up to 4000 m of altitude, in latitudes from 0 to 57°, on acid to basic soils, and in a broad range of annual rainfalls regimes (400–2500 mm) (Edmeades et al., 2017).

When the European arrived in the American continent and discovered maize, they dispersed it around the Europe and the world (Mir et al., 2013). Nowadays, this is the cereal most cultivated worldwide, with annual production of 1.45 billion tons of grain (FAOSTAT, 2020; Leff et al., 2004). This cereal is used as an essential food staple for several societies, raw material for industry, feed for livestock, and as renewable biofuel (Ranum et al., 2014).

Despite the great variability presented in the species, the germplasm pools are not fully explored. For instance, Hallauer & Miranda (1988) reported the existence of 130 distinct races in maize germplasm, from which 71% were originated in South America according Serratos-Hernández (2009). However, this diversity is underexplored, since studies on tropical maize are scarce and comprehend only part of the genetic variability existent (Laborda et al., 2005; Lanes et al., 2014; Wu et al., 2016). The exploitation of the tropical germplasm can contribute to meet current and future demands, to identify hybrids adapted to marginal crop conditions, and to find sources of resistance to abiotic and biotic stresses.

The temperate maize presents a better characterization, heterotic groups clearer defined and higher potential yield compared to the tropical maize (Edmeades et al., 2017; Stevenson & Goodman, 1972). Conversely, the latter presents a greater genetic variability.

Phytopathogenic mollicutes represent a threat to the current maize production, especially *Spiroplasma kunkelii* and the *maize bush stunt phytoplasma* (MBSP, also known as *Candidatus Phytoplasma asteris*). They are responsible for the corn stunt disease and maize bush stunt disease, respectively, and together are called the corn stunt disease complex (Bergamin Filho et al., 1995; Kimati et al., 2005). The disease complex is spread across the American continent and is transmitted by a hemipteran vector, e.g. *Dalbulus maidis*. Previously considered a secondary disease, it became greatly

important with the changes in the production system, such as the cultivation in a fall-winter second season ('safrinha'), leading to impressive losses (Kimati et al., 2005).

The exploitation of the genetic variability that exists in tropical maize for this set of diseases can lead to a better comprehension of the pathosystem and be a valuable resource for farmers and plant breeding programs. In this context, genome wide association studies (GWAS) can be applied to identify resistant genes associated with the complex, to assist understanding the metabolic pathways involved, and to find genotypes containing high proportion of favorable alleles.

Given that the construction of a representative diversity panel of tropical maize is a powerful tool to comprehend the genetic variability available, to discover genes and genotypes to improve the tropical agriculture and to meet future human and environmental demands, this study aimed to assemble and genotypically characterize a tropical maize diversity panel and to investigate the genes associated with the resistance of corn stunt disease complex, as well as the potential sources of resistance.

References

Bergamin Filho, A., Kimati, H., & Amorim, L. (1995). Manual de Fitopatologia: Princípios e Conceitos (A. Bergamin Filho, H. Kimati, & L. Amorim (eds.); 3rd ed.). Agronômica Ceres.

Doebley, J. (2004). The Genetics of Maize Evolution. Annual Review of Genetics, 38(1), 37–59. https://doi.org/10.1146/annurev.genet.38.072902.092425

Edmeades, G. O., Trevisan, W., Prasanna, B. M., & Campos, H. (2017). Tropical Maize (*Zea mays* L.). In H. Campos & P. D. S. Caligari (Eds.), Genetic improvement of tropical crops (pp. 57–109). https://doi.org/10.1007/978-3-319-59819-2

FAOSTAT. (2020). FAOSTAT. http://www.fao.org/faostat/en/?#data/QC

Hallauer, A. R., & Miranda, J. B. (1988). Quantitative Genetics in Maize Breeding (2nd ed.). Iowa State University Press.

Kimati, H., Amorim, L., Rezende, J. A. M., Bergamin Filho, A., & Camargo, L. E. A. (2005). Manual de Fitopatologia, Vol II. In Doenças das plantas cultivadas.

Laborda, P. R., Oliveira, K. M., Garcia, A. A. F., Paterniani, M. E. A. G. Z., & De Souza, A. P. (2005). Tropical maize germplasm: What can we say about its genetic diversity in the light of molecular markers? Theoretical and Applied Genetics, 111(7), 1288–1299. https://doi.org/10.1007/s00122-005-0055-7

Lanes, E. C. M., Viana, J. M. S., Paes, G. P., Paula, M. F. B., Maia, C., Caixeta, E. T., & Miranda, G. V. (2014). Population structure and genetic diversity of maize inbreds derived from tropical hybrids. Genetics and Molecular Research, 13(3), 7365–7376. https://doi.org/10.4238/2014.September.12.2

Leff, B., Ramankutty, N., & Foley, J. A. (2004). Geographic distribution of major crops across the world. Global Biogeochemical Cycles, 18(1), n/a-n/a. https://doi.org/10.1029/2003gb002108

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J. G., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences of the United States of America, 99(9), 6080–6084. https://doi.org/10.1073/pnas.052125199

Mir, C., Zerjal, T., Combes, V., Dumas, F., Madur, D., Bedoya, C., Dreisigacker, S., Franco, J., Grudloyma, P., Hao, P. X., Hearne, S., Jampatong, C., Laloë, D., Muthamia, Z., Nguyen, T., Prasanna, B. M., Taba, S., Xie, C. X., Yunus, M., … Charcosset, A. (2013). Out of America: tracing the genetic footprints of the global diffusion of maize. Theoretical and Applied Genetics, 126(11), 2671–2682. https://doi.org/10.1007/s00122-013-2164-z

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. Annals of the New York Academy of Sciences, 1312(1), 105–112. https://doi.org/10.1111/nyas.12396

Serratos-Hernández, J. A. (2009). The origin and diversity of maize in the american continent. https://doi.org/10.2166/wst.2009.007

Stevenson, J. C., & Goodman, M. M. (1972). Ecology of Exotic Races of Maize. I. Leaf Number and Tillering of 16 Races Under Four Temperatures and Two Photoperiods 1 . Crop Science, 12(6), 864–868. https://doi.org/10.2135/cropsci1972.0011183x001200060045x

Szabó, V. M., & Burr, B. (1996). Simple inheritance of key traits distinguishing maize and teosinte. Molecular and General Genetics, 252(1–2), 33–41. https://doi.org/10.1007/BF02173202

Wu, Y., San Vicente, F., Huang, K., Dhliwayo, T., Costich, D. E., Semagn, K., Sudha, N., Olsen, M., Prasanna, B. M., Zhang, X., & Babu, R. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theoretical and Applied Genetics, 129(4), 753–765. https://doi.org/10.1007/s00122-016-2664-8

2. CHARACTERIZING A TROPICAL MAIZE DIVERSITY PANEL: ENHANCING KNOWLEDGE ABOUT BRAZILIAN MAIZE GERMPLASM

ABSTRACT

Tropical maize genotypes exhibit high genetic diversity and are an important source of potential alleles for breeding programs. Despite their usefulness as an active genetic resource, their utilization is limited due to the lack of widespread genetic characterization of such collections. In this context, this study seeks to assemble and genetically characterize a tropical maize diversity panel. For that, 360 highly diverse inbred lines were genotyped using a genotyping-by-sequencing approach with restriction enzymes PstI and MseI and aligned to the version 5 of the B73 reference genome. For genetic assessments, two datasets were considered: one with the raw data, and the second one with data imputed and filtered for quality control, retaining only biallelic markers with minor allelic frequency higher than 0.05, call rate higher than 0.95, and linkage disequilibrium (r^2) lower than 0.99. The raw dataset contained 196,803 SNPs well distributed across the chromosomes. The proportion of missing data was 0.391 and the average observed heterozygosity was 0.036. The filtered dataset, containing 14,655 SNPs, showed similar estimates of populational genetic parameters compared to the first one. The structure analysis appointed that this panel presents nine subpopulations. The information derived from this study assist to comprehend the tropical maize germplasm and to support breeding programs.

Keywords: *Zea mays*, Structure, Genomic characterization, Brazil, SNP

2.1. Introduction

Maize (*Zea mays* L.) is a basic component in human and animal nutrition, in addition to being a source of raw material for various industrial products (Romay et al., 2013). Although this crop is among the most studied and corresponds to 37.2% of the cereals produced worldwide (FAOSTAT, 2020), there is still a great potential in regard of plant breeding to be achieved in the fields. This scenario is due to the wide diversity inherent to maize pools from different geographical locations and genetic backgrounds (van Heerwaarden et al., 2011), as well as the exploration of heterotic groups in terms of number and composition to generate new hybrids.

Maize populations can be divided into two main clusters, temperate and tropical. The latter covers genotypes that are sensitive to the photoperiod and poorly adapted to latitudes above 30 degrees (Edmeades et al., 2017). According to FISHER, BYERLEE and EDMEADES (2014) it is possible to classify the locations of maize cultivation into eight mega-environments according to temperature, altitude, precipitation, and sensitivity to photoperiod. Six of them are correspondent to tropical maize, which reinforces the importance of understanding the genetic basis of this germplasm and incorporate it into breeding programs.

In order to better understand the existing genetic diversity in maize, some studies on molecular characterization and formation of maize diversity panels have been carried out (Andorf

et al., 2016; Flint-Garcia et al., 2005; McMullen et al., 2009). They approached different collections of germplasm such as the Nested Association Mapping (NAM) population, Iowa, Wisconsin Diversity Panel (WIDIV), and United Stated Department of Agriculture (USDA), and employed various genotyping platforms. However, these studies sought to understand the genetic variability predominantly using tempered maize or approaching a generalist view, with a limited number of tropical lines.

Therefore, there are well stablished pools that comprehend the genetic diversity existing in the temperate maize germplasm, mainly from North America and Europe (Romay et al., 2013). On the other hand, studies focused on tropical maize are scarce and address only a small portion of the total genetic variability (LABORDA et al., 2005, WU et al., 2016, LANES et al., 2014) . In addition, since the tropical maize pool is larger than the temperate one because of the many rare alleles, great number of markers and the wide diversity of haplotypes (Liu et al. 2003), there is a significant amount of research to be conducted yet. Consequently, the current classification of the tropical germplasm lacks accuracy Another important aspect is the fact that Brazilian commercial maize hybrids are often grouped with the B73 line in population studies, which indicates that our tropical materials are genetically related to this reference genotype of the temperate germplasm (Andrade et al., 2016). In this context, there is an urgent appeal for tropical germplasm to be studied and implemented in commercial breeding programs in an efficient manner, aiming the true exploration of its genetic potential.

The quality and amount of information obtained using a diversity panel depend directly on its genotypic and phenotypic characterization. Genotypic variability can be explored with the aid of molecular markers through a great number of approaches, such as principal component analysis (Zhang et al., 2016), kinship coefficients (Yu & Buckler, 2006), allelic diversity (Buckler et al., 2009), mixed models (Riedelsheimer et al., 2012; X. Wu et al., 2015), and genome wide association studies (GWAS - WU et al., 2016).

In light of the exposed, this study focused on the construction of a representative Brazilian tropical maize panel and the genotypically characterization of the lines via genetic population parameters, structure and relationship information.

2.2. Materials and methods

2.2.1. Genotypic Material

In an effort to represent the genotypic variability existing in the tropical maize germplasm, 360 inbred lines were used in this study (Table 1). They were originated from different regions in Brazil and were obtained from two active germplasm banks: 179 lines were from the Department of Genetics, "Luiz de Queiroz" College of Agriculture – University of São Paulo (ESALQ-USP), and 181 from the Agronomic Institute of Paraná (IAPAR).

2.2.2. Genomic data

Healthy leaf samples in the V3 stage were used for DNA extraction using the protocol of INGLIS et al. (2018). The inbred lines were genotyped using the POLAND et al. (2012) genotyping-by-sequencing (GBS) protocol modified by using the endonucleases PstI and MseI. The identification of single nucleotide polymorphism (SNPs) markers was performed through TASSEL 5 standalone V5.2.54 (Bradbury et al., 2007a). The alignment was performed with the B73 genome reference version 5.0 (released in 2020) using the software Bowtie2 V2.3.3.1 and the parameter "very-sensitive" from it (Langmead & Salzberg, 2013). This step yielded a raw marker dataset with 196,234 SNPs.

Sequentially, we filtered the raw dataset retaining only chromosomal markers and imputed the missing data in the lines using Beagle 5.0 (S. R. Browning & Browning, 2007). Complementarily, the software VCFTools (Danecek et al., 2011) and the R packages synbreed (Wimmer et al., 2012) and snpReady (Granato et al., 2018) were adopted to filter for biallelic loci, minor allele frequency (MAF) higher than 0.05, call rate (CR) of 0.95, and linkage disequilibrium (LD) pruning with r^2 lower than 0.99. Finally, a filtered dataset with 14,639 SNPs was obtained (Figure 1).

2.2.3. Panel genomic characterization

Both datasets, raw and filtered, were used to characterize the maize panel. The populational genetic parameters adopted were: expected and observed heterozygosity (He and Ho, respectively), number of alleles per marker locus, proportion of missing data, nitrogenous base ratios, endogamy coefficient (F_{st}), polymorphism information content (PIC), effective population size (Ne) and Roger genetic distance (D_a) (Rogers, 1972). The F_{st} was calculated using Resende et al. (2014) approach, as follows:

$$F_{st} = diag(G) - 1 = diag\left(\frac{WW'}{\sum_{i=1}^{n} 2p_i(1-p_i)}\right) - 1$$

where G is the genomic relationship matrix, W is the incidence matrix for the fixed effect of biallelic markers, p_i is the allelic frequencies of one allele of the biallelic markers for the loci I. The Ne was obtained by $N_e = 1 * (1 + F_{st})^{-1}$. Furthermore, we contrast the distribution of the markers along the chromosomes to the maize karyotype for identification of centromeres and telomeres (Wolfgruber et al., 2009).

2.2.4. Genetic structure

To determine and to visualize the genetic structure of the diversity panel, we used the filtered dataset to perform principal component analysis (PCA), construction of the populational dendrogram, and structure analysis using Roger's genetic distance matrix. We also analyzed the clustering via K-means method incorporating all eigenvectors. For that, we used the R package adegenet (Jombart, 2008). The most likely number of groups was identified as the one with lower average Bayesian Information Content (BIC) over 100 replicates.

The software Structure (Pritchard et al., 2000) was used to analyze the structuration of the panel and the possible occurrence of subpopulations from the filtered dataset through a admixture model. The number of simulated subpopulations (K) varied from 1 to 15, considering 10 replicates for each K value. For each run, we utilized a burn-in of 10^6 Markov Chain Monte Carlo (MCMC) iterations and kept the data from 10^6 iterations. To identify the most likely K value, we calculated the |L"(K)| and ΔK parameters proposed by EVANNO, REGNAUT and GOUDET (2005) and used the BIC criteria from the K-means analysis. Afterwards, the genotypes were classified according to the proportion of resemblance in each subpopulation (Q). Individuals with Q value higher than 0.5 in a given subpopulation were clustered as belonging to it, otherwise the individual was classified as mixed.

The dendrogram construction was fashioned using Unweighted Pair Group Method Average (UPGMA), and it was compared to the available pedigree, the K-means clusters, and to the groups resulted from the STRUCTURE analysis.

2.2.5. Data availability

The raw and filtered datasets used in the characterization of this tropical maize diversity panel are available to the scientific community in an online database hosted at https://data.mendeley.com/datasets/5gvznd2b3n/2. Furthermore, all the field evaluation data

(traditional characterization and high-throughput phenotyping) of this panel are made available every year at the same database.

2.3. Results

2.3.1. Raw data

Analyzing the raw dataset, the frequency of missing data by genotype had a normal distribution, with mean of 0.391 (Figure 2a). Only 07 genotypes out of the 360 (1.94%) presented a proportion of missing data higher than 0.6, which are: L147, L183, L211, L227, L269, L273, and L309. By marker, the distribution of missing data showed a peak of values below 0.05 (14.9% of markers), and a small proportion of markers with more than 0.80 (4.7% of markers – Figure 2b).

The total number of markers for the raw dataset was 196,803, from which 98.7% were aligned to chromosomal regions. The number of markers per chromosome was proportional to the chromosome length (except for chromosome 5, which presented more markers than chromosomes 3 and 4, that are longer).

The overall density of the markers was 1 by 10,967 base pairs (bp). This ratio was also found within chromosomes (Chr), ranging from 1:9,917 (Chr 2) to 1:12,625 (Chr 4). In general, we observed a satisfactory coverage of markers in each chromosome (Figure 3). The regions with low marker density can be classified in two categories: (1) long sparse markers regions, associated to the proximity of the centromere (Wolfgruber et al., 2009), and (2) short absent markers regions, associated to regions with low amount or absence of genes (Andorf et al., 2016).

As expected, the proportion of the nitrogenous basis between C and G (0.290 and 0.287) and between A and T (0.211 and 0.212) was approximately 1.0 (Figure 4a). The ratio of [CG] and [AT] was 1.366, indicating the prevalence of nucleic acids with three hydrogen bonds.

The average observed heterozygosis was 0.036 (Figure 4b), which is expected for lines in S4 (4 successive generations of selfing). Eight lines (2.2%) presented Ho higher than 0.07, indicating they undergone a lower number of selfing generations. The higher values were found in the lines L7 (0.100), L93 (0.104), and L147 (0.138).

The diversity parameters expected heterozygosity and polymorphism information content (PIC) were 0.289 and 0.236, respectively (Figure 4d, e). The F_{it} value distribution regarding the markers showed an important peak near 1.0, which indicates that most of the markers are in homozygosity (Figure 4f). By this parameter, we could also observe there was a small number of markers in Hardy-Weinberg equilibrium, and only a few markers apparently presented some degree

of heterozygosity. The coefficient of endogamy (F) for the whole population was 0.837, generating an effective population size (Ne) of 215.0 individuals.

The great majority of markers from the dataset were biallelic (96.8%). The presence of monomorphic (fixed), triallelic and tetrallelic markers was 0.4%, 2.8% and 0.1%, respectively (Figure 5). The majority of heterozygosity observed in the markers was below 0.05.

2.3.2. Filtered data

The filtered data set has 14,655 markers. Their overall distribution and parameters were similar to the raw dataset. The average MAF was 0.219, with a higher frequency of markers with MAF lower than 0.20 (Figure 6a). The expected heterozygosity was 0.307 and the observed was 0.029 (Figure 6b), which relates to lines in S4, as previously pinpointed. The majority of markers (91.6%) presented observed heterozygosity below 0.05. The average PIC was 0.250, ranging from 0.09 to 0.38 (Figure 6c).

The mean Fit for the markers was 0.890, and the trends observed here were similar to the ones raised for the raw dataset (Figure 6d). The fact that a small number of markers were heterozygous for the majority of genotypes is probably related to the residual heterozygosity from an event of selection pressure or the occurrence of a duplicate region absent in the temperate genome of reference, B73 (Brandenburg et al., 2017). In addition, six markers showed Fit lower than -0.90.

The coefficient of endogamy (F) for the whole population was 0.907, corresponding to a Ne of 198.4 individuals. The distribution of the markers presented the same pattern than the raw data.

2.3.3. Dendrogram and clustering

The genetic distance matrix for the raw dataset showed a genetically divergent population (Figure 7). The distance between individuals off the diagonal ranged from 0.023 to 0.294, with average of 0.253. Furthermore, we could observe some clusters in the panel, which suggests the existence of populational structure (Figure 7).

The cumulative PCA eigenvalues shows that 95% of the information is explained by 265 of higher principal components (PC) and 50% with 46 PC (Figure 8). The most informative PC's, PC1 and PC2, represent 3.87% and 3.58% of the total information, respectively. From these outcomes, it is possible to infer that the panel exhibits a great genetic diversity. PC1 and PC2

(Figure 9) are substantially correlated to the observed heterozygosity and the inbreeding coefficient (correlation between PC1 and Ho = 0.669, and between PC2 and Fi = 0.632), and individuals with higher values of He (L093, L145, and L154) were allocated at the extremes of both PC's.

By plotting the three main PCs, there is no clear structuration of the panel in clusters, which is probably due to the strong effect of heterozygosity on them. Conversely, the components PC3, PC4, and PC5 allowed to well distinguish the genotypes into subsets. The K-means analysis on the 359 PCs according to the BIC criterion indicated that the most likely number of groups that compose the whole panel is between 8 and 14 groups, being the strongest evidence to 10.

Using the delta K criterion for the structure analysis, the panel is most likely divided into 2 groups (Figure 10). In this case (K=2), the Structure analysis indicated a major group with individuals (black subgroup in Figure 11) and a second group with other genotypes (cyan group in Figure 11). This minor cluster, clearly separated, is composed by individuals with a high level of endogamy due to the fact they are from a pool of five parents: 'D 480', 'D 505', D 523', 'D 603' and 'D549'. In addition, this group is clustered under the same long branch in the dendrogram, which corroborates the high divergence from the panel population (Figure 12).

The first peak on the delta K analysis (K=2) endorsed a highly endogamic pool of individuals, which is also shown by the distance matrix. Next, the BIC analysis indicated a higher number of groups, and the L"(K) criterion indicated peaks for K=4, 6, and 9. Since the division of the population in a greater number of groups may increase the information gain in accordance to the BIC criterion, we adopted K=9 as the definitive number of clusters within our panel.

To assist the references of the groups, we standardized their names according to their color in the structure plot with K=9 (Figure 11d). It possible to observe that the panel structuration is in accordance to the population dendrogram since the majority of individuals belonging to the same group in the structure analysis is assigned in close branches and are monophyletic. There are a few exceptions, such as the purple group, whose individuals are out of its main branch. One possible explanation for that is the large number of individuals in it with contribution from other groups.

274 inbred lines were classified as from a specific group in the structure analysis. On the other hand, 23.3% of the panel were classified as derived from an admixture population. The group sizes ranged from 17 individuals (cyan group) to 64 (purple one).

2.4. Discussion

2.4.1. Panel diversity

The present population present elevated population effective size (Ne) because the genotypes evaluated were lines and some of them share the same progenitors. This value decreased when the filtered dataset was used, since the genetic diversity found in rare alleles (MAF<0.05) and the triallelic and tetrallelic markers was removed. The MAF distribution in the filtered dataset showed a higher density of markers yielding lower than 0.20 for this parameter, which is due to the dispersion of less frequent alleles in the panel. Similar values of MAF, PIC and expected heterozygosity were found by Andrade et al. (2016) when evaluating 20 Brazilian hybrids. On the other hand, Chen et al. (2016) assessed 561 maize inbred lines from CIMMYT and found similar values of PIC, but smaller MAF (0.16) and higher observed heterozygosity. Yang et al. (2010), evaluating a pool of 527 lines from tropical and temperate germplasms, found higher MAF (0.30), PIC (0.31) and similar observed heterozygosity. The similar values of PIC in all these populations suggest a sharing genetic diversity within the tropical populations. On the other hand, our panel exhibited rare alleles in higher frequency compared to the others.

Several lines from our population are derived from commercial breeding lines, which have shaped the genetic base of maize breeding in Brazil. In addition to that, the similarity between the genetic parameters of our panel and the Brazilian commercial hybrids from Andrade et al. (2016) leads to the inference that our population contains the genetic diversity of the active germplasm of breeding companies in Brazil. In this context, the use of lines with expired plant variety protection certificate (ex-PVPs) for North American germplasm would allow to exploit the genetic variability of current elite lines, even if they have smaller concentration of favorable alleles than commercial germplasms due to constant selection throughout generations (White et al., 2020).

The accumulated principal components analysis required a high number of lines to explain the genomic variation in the panel. In addition, the use of the raw dataset is an important tool to identify promising rare alleles. For example, using genotypes from different subpopulations in a nested mapping population can increase the frequency of rare alleles and assist their comprehension and application in breeding programs.

The presence of markers with Fit close to -1 may indicate a high pressure of selection against both homozygous genotypes or duplicate genes absent in the temperate reference line B73 (Brandenburg et al., 2017). Andrade et al. (2016) demonstrated that Brazilian hybrids clustered close to the line B73. However, even though the genetic parameters between both populations

were similar, there is a set of genes that differ between their genetic pools and reinforce the high genetic diversity available in this tropical panel.

2.4.2. Panel structure

The criteria to define the ideal number of subpopulations varied. In here, the delta K criterion indicated the most probable division in two groups. However, this division is due to a specific 'diallel' subpopulation present in the panel, that is, does not represent the real division of the panel. The second, third, and fourth peaks of delta K were much more modest (4.58 for K=6, 2.50 for K=4 and 1.57 for K=9) in relation to K=2 peak (2,189.36). The |L"(K)|parameter was more parsimonious, being K=2 the most likely number to assign the population division, followed by K=9. The BIC criterion diverged from the Structure analysis output. According to it, the highest peak occurs in K=10. Then, the use only of the first pick of delta K plot identify the most probable division, but do not necessary the division that explore better the genetic information available.

The parameters |L"(K)|, secondary delta K peaks, and BIC from K-means analysis should be considered together to define the number of subpopulation that better contemplates the information. Another potential approach is to use the subdivisions to understand the evolutionary breeding process. For example, the group cyan kept apart from the others in all peaks (K=2, 4, 6, and 9), while the blue-navy group was divided in four subgroups in K=9 compared to K=2.

The existence of individuals with high levels of heterozygosis affected the PCA. This can be observed by the extreme values in PC1 and PC2 for individuals that had a greater observed heterozygosity frequency, and by the dendrogram cluster formation, where these individuals were marginal to the subpopulations established.

The performance of backcrosses with the hybrid donors 2B710H and P30F53H did not affect the group formation and did not lead the groups to be close in the dendrogram. This can be explained by the fact that the expected genetic contribution of both these hybrids for the lines is 0.25, which is lower than the value defined as threshold to determine the clusters (0.50).

The purple group covers parental lines such as C-701, C-808, C-505, DKB C901, AG 9010, DKB 350, DKB 909, and AG 6. The companies that originate these lines were Cargill (responsible for the first three lines), Dekalb (responsible for the lines with 'DKB' codification), and Agroceres (responsible for the lines with 'AG'). They share genetic resemblance since Cargill, Agroceres and Dekalb were acquired for Monsanto company, which was recently acquired by Bayer. Furthermore, the genotypes C-505, C-701, C-808 and DKB 350 present the same type of grain (semi-flint).

In the green group, the most recurrent parent was ICI-8452 (or Zeneca 8452), a semi-dent double-cross hybrid. The company Zeneca, owner of this material, was bought by Syngenta, nowadays part of the ChimChem Chinese company.

The black group comprises as genitors the genotypes ST and LG8, both from the IAPAR germplasm bank. ST was backcrossed with the hybrid P30F53H from the company Pionner (nowadays Corteva), while LG8 was backcrossed with the hybrid 2B710H from DowAgroscience (which was recently split into LongPing High-Tech and Corteva). Similarly, the red group contains as parental individuals the IAPAR genotypes RSM and MC 04, which were respectively backcrossed with 2B710H (DowAgroscience/LongPing-Corteva) and P30F53H (Pionner/Corteva). Both black and red groups present long branches in the dendrogram, indicating a high genetic distance from rest of the panel, likely due to the backcross origin.

The gray group presented a high number of different ancestors. Among them, there are the CO-32 and C-333. The former is a three-way cross from the company Colorado, later acquired by DowAgroscience (now LongPing and Corteva), and the latter is a single hybrid from Cargill (now Bayer).

The orange population comprises as main ancestors the flint genotypes BR-201, HS-1, IG-2 and CMS-05. BR 201 is a double-cross hybrid from EMBRAPA, HS-1 is the female hybrid that crossed with HS-2 to produce BR-201, and IG-2 is from the ESALQ population, which was derived from a three-way cross between CMS-06 (tropical genotype from CIMMYT) and EPB-4, and then with HS-2. EPB-4 is a compose obtained from the tropical races Tuxpeño Crema I, MEB I, and Antigua Gpo 2. The genotype CMS-05 is related to the compose Suwan 1, a Thai tropical germplasm, which present good ability to combine with stiff stalk and non-stiff stalk materials (Fan et al., 2015).

The navy group has among the genitors eight hybrids: Z8420, C-791151, P30F88 and PF-41X05-33-05B, which respectively are a flint single-cross hybrid from DowAgroscience, a semi-flint three-way cross from Zeneca, a flint single-cross hybrid from Pioneer, and a genitor from ESALQ; and MC08, F948, F978 and F983, all four from IAPAR.

Progenitors and commercial lines lacking pedigree information, the use of non-commercial lines in panels and the specific schema of codification from each source may difficult the historic connection of maize germplasm, but that can be clarified using molecular markers (White et al., 2020). The absence of heterotic groups well assigned in the tropical germplasm may have contributed to the occurrence of nine different subpopulations in this panel (Edmeades et al., 2017). Nevertheless, it is recurrent the assumption of three main heterotic patterns: Tuxpeño, Non-Tuxpeño and Suwan1 (Edmeades et al., 2017; Lanes et al., 2014; Reif et al., 2003). In this scenario,

the use of BIC from K-means method |L''(K)| and delta K criteria represent parsimonious approaches in the establishment of subpopulations.

2.5. Conclusion

The tropical maize diversity panel evaluated in this study presented a considerable genetic variability. The filtered data yielded estimations similarly to the raw dataset, but with reduction in the population effective size due to the exclusion of the non-biallelic markers and the diversity present in rare alleles. The structure analysis designated nine groups as the most parsimonious clustering of the panel. They provided a consistent classification of genotypes according to their origin. For instance, the purple group comprised individuals from Cargill's germplasm. The similarity of genetic parameters obtained from this panel and from current commercial germplasm indicates a substantial resemblance between them, which must be explored better in future studies. Therefore, these finds play an important role on the direct assistance of tropical maize breeding programs.

References

Andorf, C. M., Cannon, E. K., Portwood, J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Braun, B. L., Campbell, D. A., Vinnakota, A. G., Sribalusu, V. V., Huerta, M., Cho, K. T., Wimalanathan, K., Richter, J. D., Mauch, E. D., Rao, B. S., Birkett, S. M., Sen, T. Z., & Lawrence-Dill, C. J. (2016). MaizeGDB update: New tools, data and interface for the maize model organism database. Nucleic Acids Research, 44(D1), D1195–D1201. https://doi.org/10.1093/nar/gkv1007

Andrade, L. R. B. de, Fritsche Neto, R., Granato, Í. S. C., Sant'Ana, G. C., Morais, P. P. P., & Borém, A. (2016). Genetic Vulnerability and the Relationship of Commercial Germplasms of Maize in Brazil with the Nested Association Mapping Parents. PLOS ONE, 11(10), 1–14. https://doi.org/10.1371/journal.pone.0163739

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. Bioinformatics, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brandenburg, J. T., Mary-Huard, T., Rigaill, G., Hearne, S. J., Corti, H., Joets, J., Vitte, C., Charcosset, A., Nicolas, S. D., & Tenaillon, M. I. (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genetics, 13(3), 1–30. https://doi.org/10.1371/journal.pgen.1006666

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics, 81(5), 1084–1097. https://doi.org/10.1086/521987

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., Goodman, M. M., Harjes, C., Guill, K., Kroon, D. E., Larsson, S., Lepak, N. K., Li, H., Mitchell, S. E., Pressoir, G., … McMullen, M. D. (2009). The Genetic Architecture of Maize Flowering Time. Science, 325(5941), 714–718. https://doi.org/10.1126/science.1174276

Chen, J., Zavala, C., Ortega, N., Petroli, C., Franco, J., Burgueño, J., Costich, D. E., & Hearne, S. J. (2016). The development of quality control genotyping approaches: A case study using elite maize lines. PLoS ONE, 11(6), 1–17. https://doi.org/10.1371/journal.pone.0157236

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics, 27 (15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Edmeades, G. O., Trevisan, W., Prasanna, B. M., & Campos, H. (2017). Tropical Maize (*Zea mays* L.). In Genetic Improvement of Tropical Crops (pp. 57–109). Springer International Publishing. https://doi.org/10.1007/978-3-319-59819-2_3

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fan, X. M., Bi, Y. Q., Zhang, Y. D., Jeffers, D. P., Yao, W. H., Chen, H. M., Zhao, L. Q., & Kang, M. S. (2015). Use of the Suwanl heterotic group in maize breeding programs in Southwestern China. Agronomy Journal, 107(6), 2353–2362. https://doi.org/10.2134/agronj15.0122

FAOSTAT. (2020). FAOSTAT. http://www.fao.org/faostat/en/?#data/QC

Fisher, T., Byerlee, D., & Edmeades, G. (2014). Crop yields and global food security: will yield increase continue to feed the world? Australian Centre for International Agricultural Research. https://doi.org/ISBN 978 1 925133 06 6 (PDF)

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant Journal, 44(6), 1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Granato, I. S. C., Galli, G., de Oliveira Couto, E. G., e Souza, M. B., Mendonça, L. F., & Fritsche-Neto, R. (2018). snpReady: a tool to assist breeders in genomic analysis. Molecular Breeding, 38(102), 1–7. https://doi.org/10.1007/s11032-018-0844-8

Inglis, P. W., Marilia de Castro, R. P., Resende, L. V., & Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLoS ONE, 13(10), 1–14. https://doi.org/10.1371/journal.pone.0206085

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Laborda, P. R., Oliveira, K. M., Garcia, A. A. F., Paterniani, M. E. A. G. Z., & De Souza, A. P. (2005). Tropical maize germplasm: What can we say about its genetic diversity in the light of molecular markers? Theoretical and Applied Genetics, 111(7), 1288–1299. https://doi.org/10.1007/s00122-005-0055-7

Lanes, E. C. M., Viana, J. M. S., Paes, G. P., Paula, M. F. B., Maia, C., Caixeta, E. T., & Miranda, G. V. (2014). Population structure and genetic diversity of maize inbreds derived from tropical hybrids. Genetics and Molecular Research, 13(3), 7365–7376. https://doi.org/10.4238/2014.September.12.2

Langmead, B., & Salzberg, S. (2013). Bowtie2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923.Fast

Liu, K., Goodman, M., Muse, S., Smith, J. S., Buckler, E., & Doebley, J. (2003). Genetic Structure and Diversity Among Maize Inbred Lines as Inferred From DNA Microsatellites. Genetics, 2128(December), 2117–2128.

McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S. E., Peterson, B., … Buckler, E. S. (2009). Genetic Properties of the Maize Nested Association Mapping Population. Science, 325(5941), 737–740. https://doi.org/10.1126/science.1174320

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE, 7(2). https://doi.org/10.1371/journal.pone.0032253

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics, 155, 945–959.

Reif, J. C., Melchinger, A. E., Xia, X., Warburton, M. L., Hoisington, D. A., Vasal, S. K., Srinivasan, G., Bohn, M., & Frisch, M. (2003). Genetic Distance Based on Simple Sequence Repeats and Heterosis in Tropical Maize Populations. June 2014, 1275–1282. https://doi.org/10.2135/cropsci2003.1275

Resende, M.D.V., Silva, F.F., Azevedo, C. F. (2014). Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência (1st ed.). Suprema.

Riedelsheimer, C., Technow, F., & Melchinger, A. E. (2012). Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-452

Rogers, J. S. (1972). Measures of genetic similarity and genetic distances. Studies in Genetics, 7213, 145–153.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., Elshire, R. J., Acharya, C. B., Mitchell, S. E., Flint-Garcia, S. A., McMullen, M. D., Holland, J. B., Buckler, E. S., & Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology, 14(6), R55. https://doi.org/10.1186/gb-2013-14-6-r55

van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., de Jesus Sanchez Gonzalez, J., & Ross-Ibarra, J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proceedings of the National Academy of Sciences, 108(3), 1088–1092. https://doi.org/10.1073/pnas.1013011108

White, M. R., Mikel, M. A., de Leon, N., & Kaeppler, S. M. (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. Crop Science, 60(1), 100–114. https://doi.org/10.1002/csc2.20050

Wimmer, V., Albrecht, T., Auinger, H.-J., & Schön, C.-C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics, 28(15), 2086–2087. https://doi.org/10.1093/bioinformatics/bts335

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D. H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., Allison, J., Birchler, J. A., Jiang, J., Dawe, R. K., & Presting, G. G. (2009). Maize centromere structure and evolution: Sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genetics, 5(11), 13–16. https://doi.org/10.1371/journal.pgen.1000743

Wu, X., Li, Y., Li, X., Li, C., Shi, Y., Song, Y., Zheng, Z., Li, Y., & Wang, T. (2015). Analysis of genetic differentiation and genomic variation to reveal potential regions of importance during maize improvement. BMC Plant Biology, 15(1), 1–13. https://doi.org/10.1186/s12870-015-0646-7

Wu, Y., San Vicente, F., Huang, K., Dhliwayo, T., Costich, D. E., Semagn, K., Sudha, N., Olsen, M., Prasanna, B. M., Zhang, X., & Babu, R. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theoretical and Applied Genetics, 129(4), 753–765. https://doi.org/10.1007/s00122-016-2664-8

Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., Gao, Y., Chai, Y., Fu, Z., Zhou, Y., Xu, S., Bai, G., Meng, Y., Zheng, Y., & Li, J. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theoretical and Applied Genetics, 121(3), 417–431. https://doi.org/10.1007/s00122-010-1320-y

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology, 17(2), 155–160. https://doi.org/10.1016/j.copbio.2006.02.003

Zhang, X., Zhang, H., Li, L., Lan, H., Ren, Z., Liu, D., Wu, L., Liu, H., Jaqueth, J., Li, B., Pan, G., & Gao, S. (2016). Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. BMC Genomics, 17(1), 1–16. https://doi.org/10.1186/s12864-016-3041-3

Figures

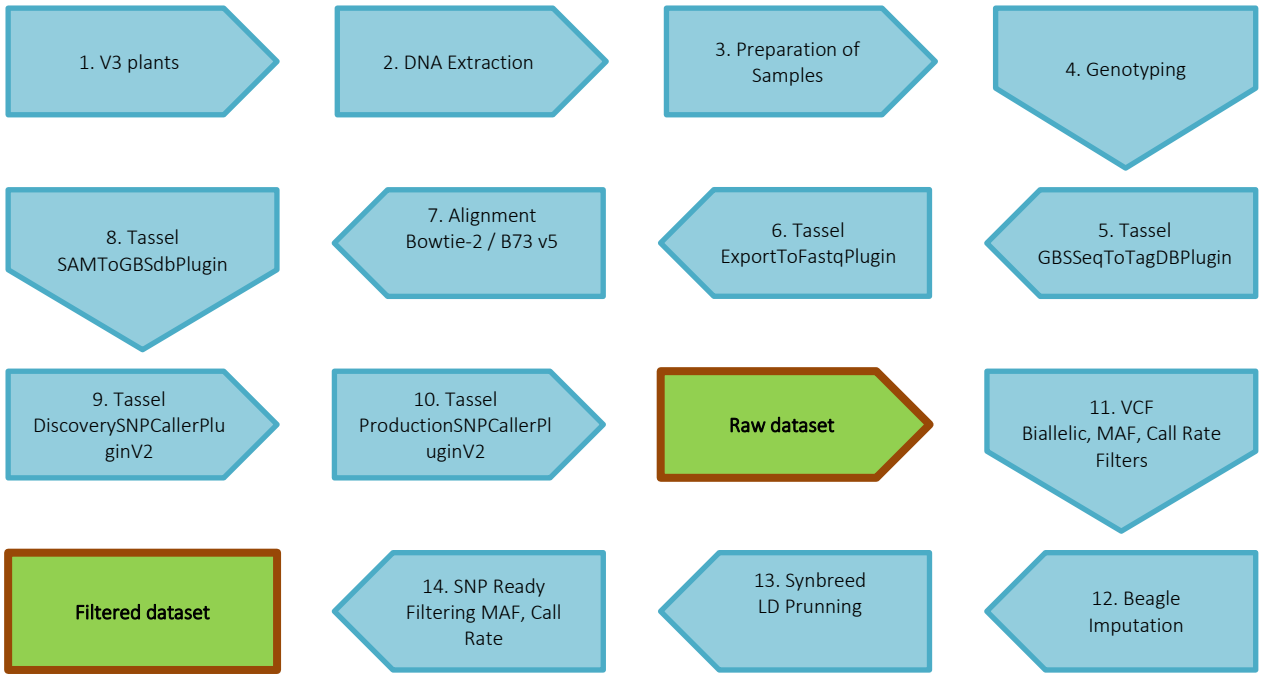

Figure 1. Molecular markers acquisition pipeline. Blue Arrows indicate processes and green arrows indicate the markers dataset.

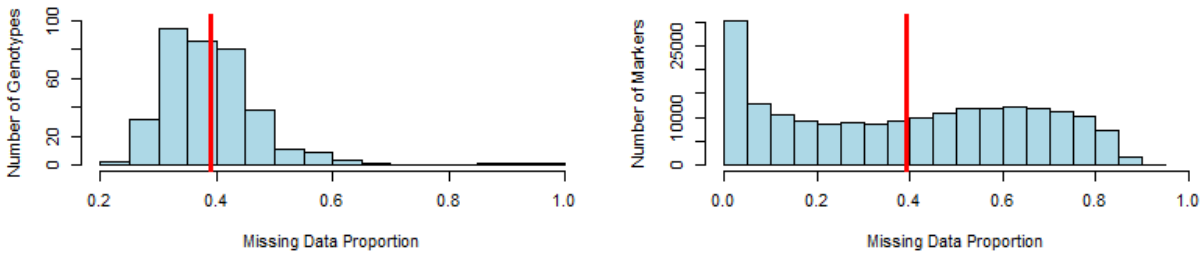

Figure 2. Distribution of proportion of missing data of tropical maize diversity panel in the raw dataset by genotype (left) and marker (right).

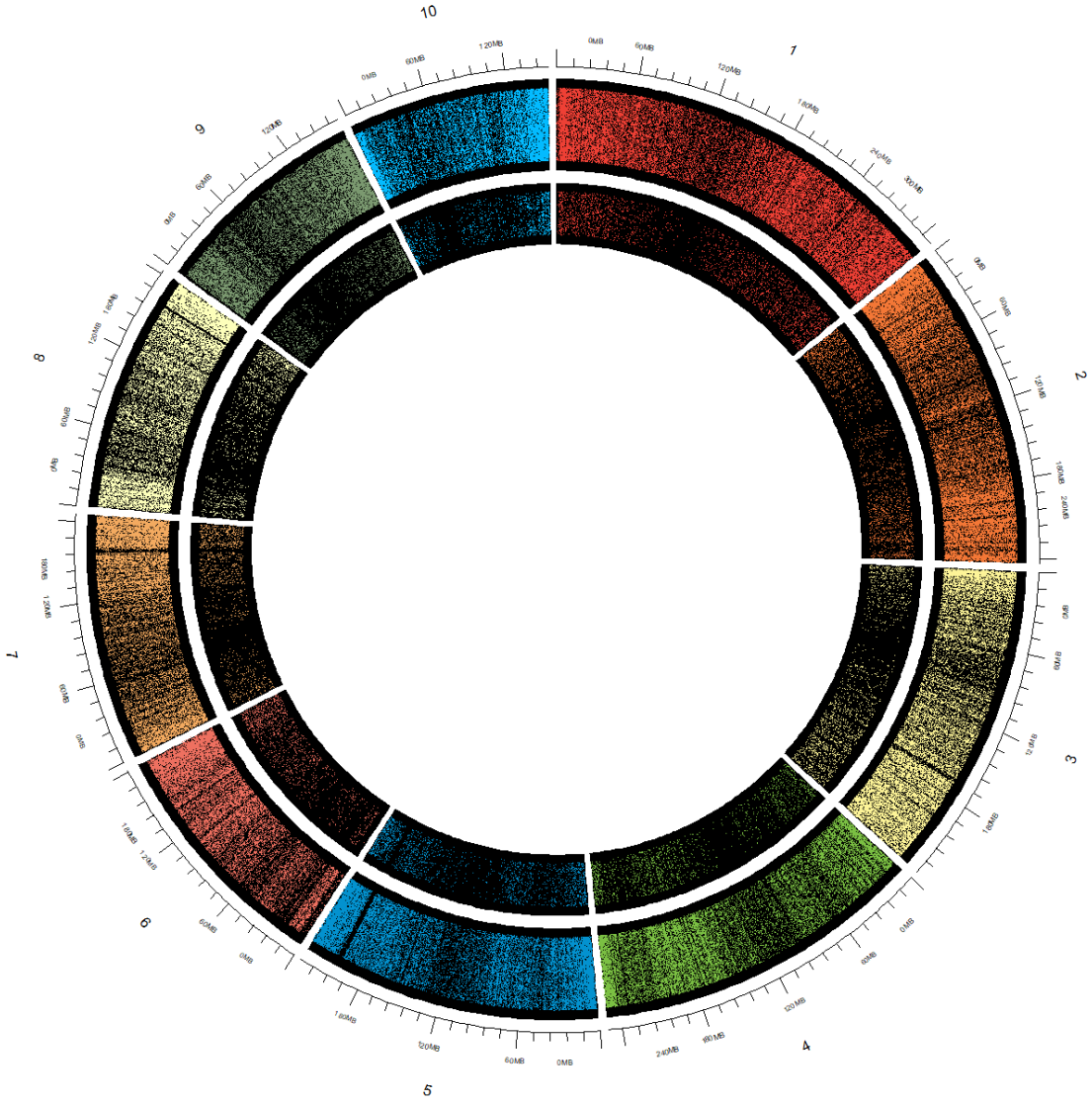

Figure 3. Distribution of raw and filtered dataset SNPs along the maize chromosomes of version 5 of B73 genome of reference. Colorful dots represent SNPs, outer circle represents raw dataset and internal one the filtered dataset.

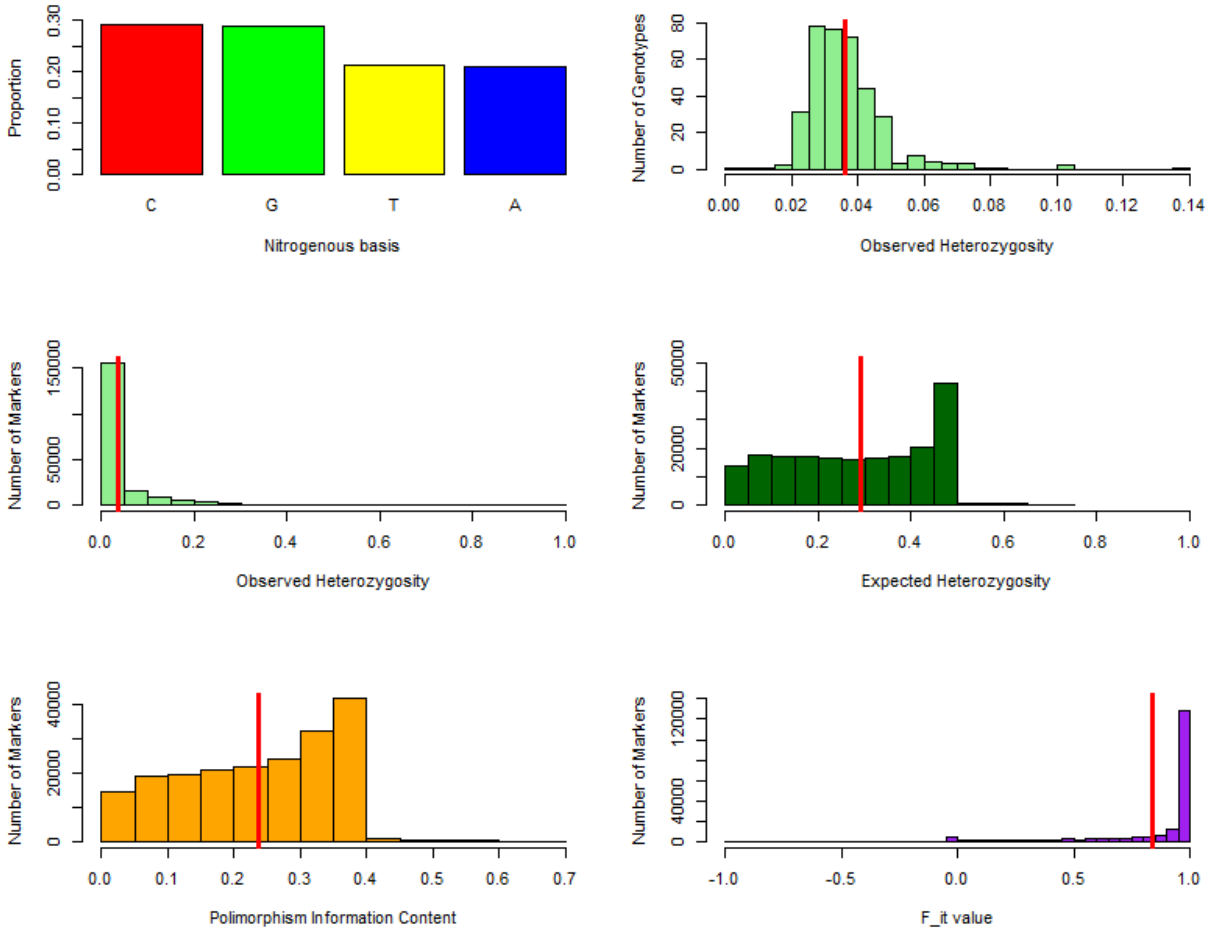

Figure 4. Genetic parameters for of tropical maize diversity panel in the raw dataset. From left to right and downwards, (a) proportion of nitrogenous basis, (b) frequency of observed heterozygosity by genotype and (c) by marker, (d) frequency of expected heterozygosity by marker, (e) polymorphism information content distribution by marker, and (f) Fit distribution by marker.

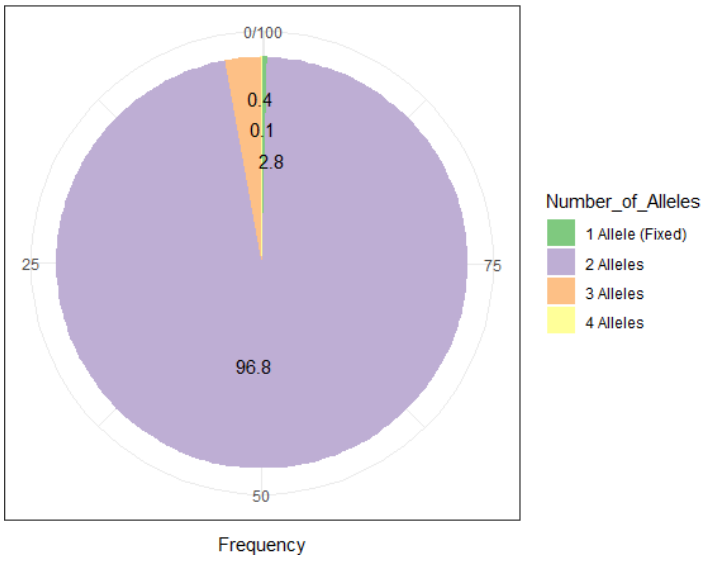

Figure 5. Proportion of polymorphism for of tropical maize diversity panel in the raw dataset.

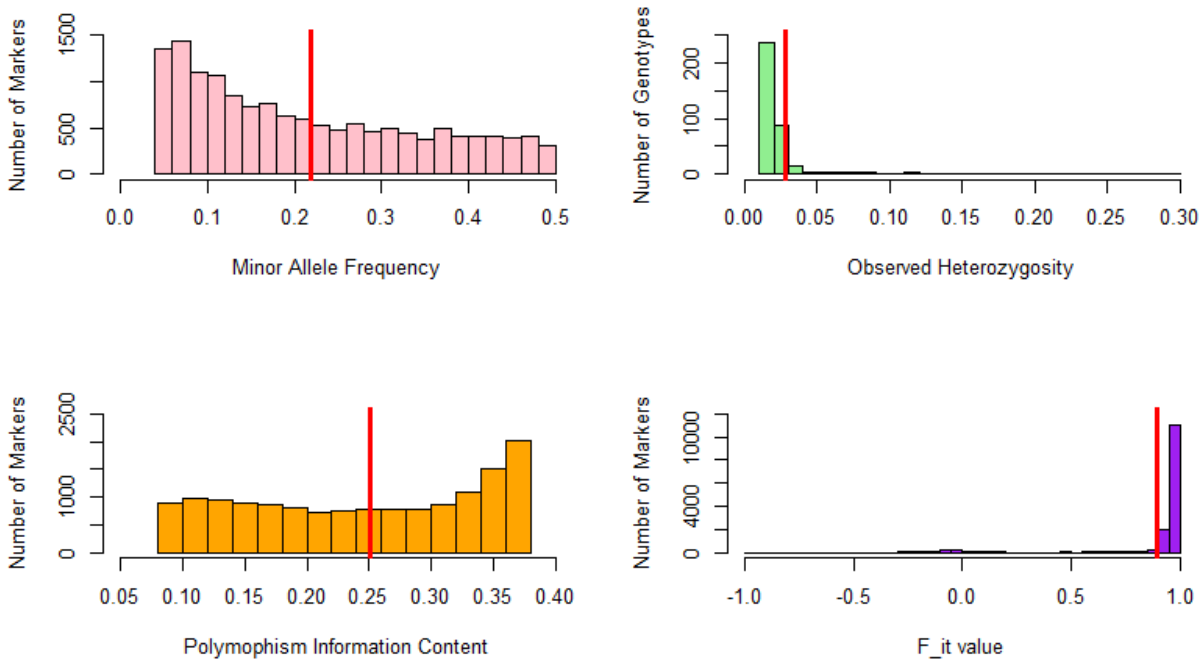

Figure 6. Genetic parameters for of tropical maize diversity panel in the filtered dataset. From left to right and downwards, (a) minor allele frequency distribution, (b) frequency of observed heterozygosity by genotype, (c) polymorphism information content distribution by marker, and (d) Fit distribution by marker.

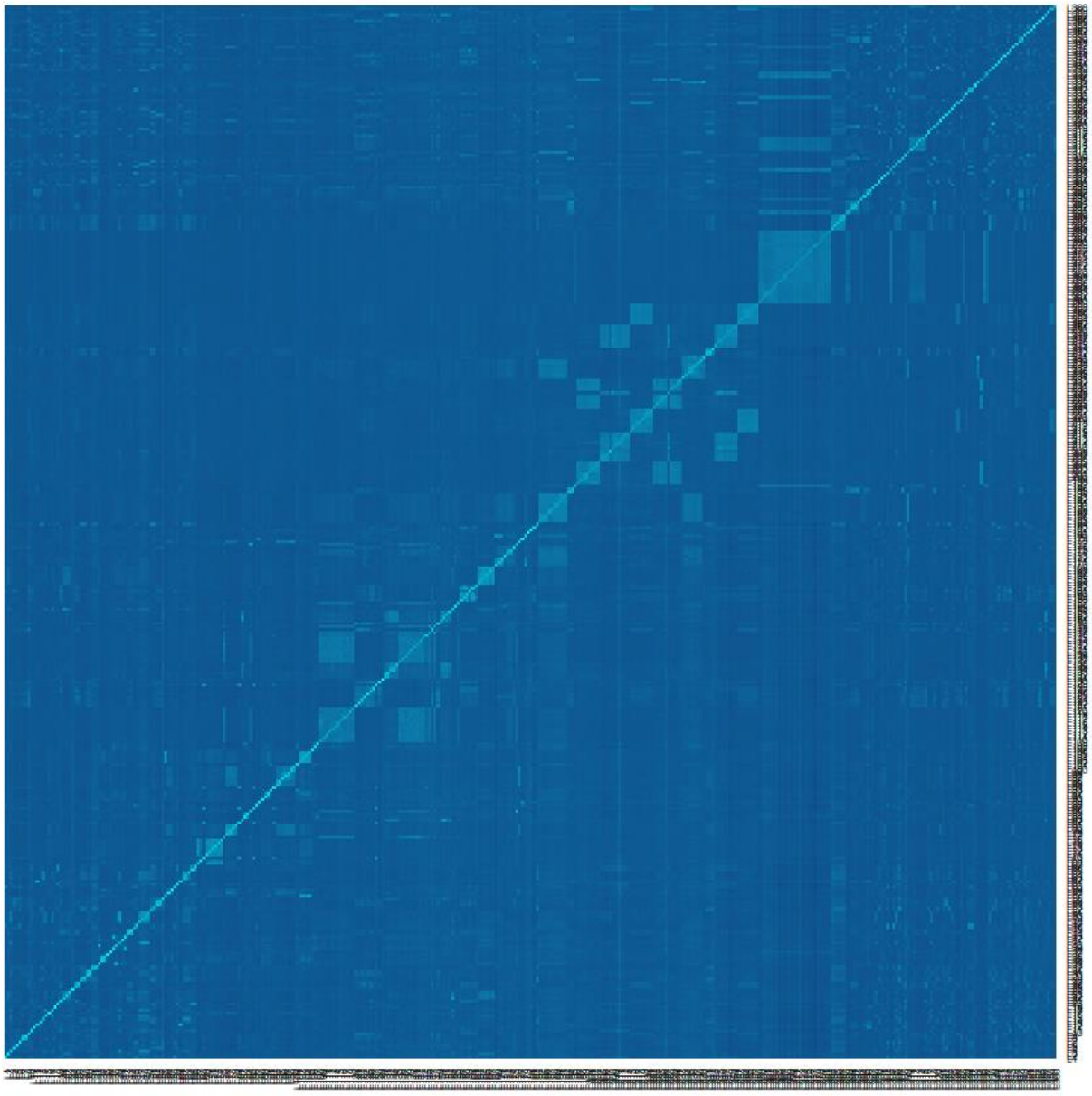

Figure 7. Rogers Genetic distance matrix for the tropical maize diversity panel using the filtered dataset. Darker blue pixels indicate high genetic distance (maximum of 0.290), and light cyan pixels indicate values low genetic distance (minimum of 0.000). Lines are ordered from left to right and from bottom to top (L1 to L360)

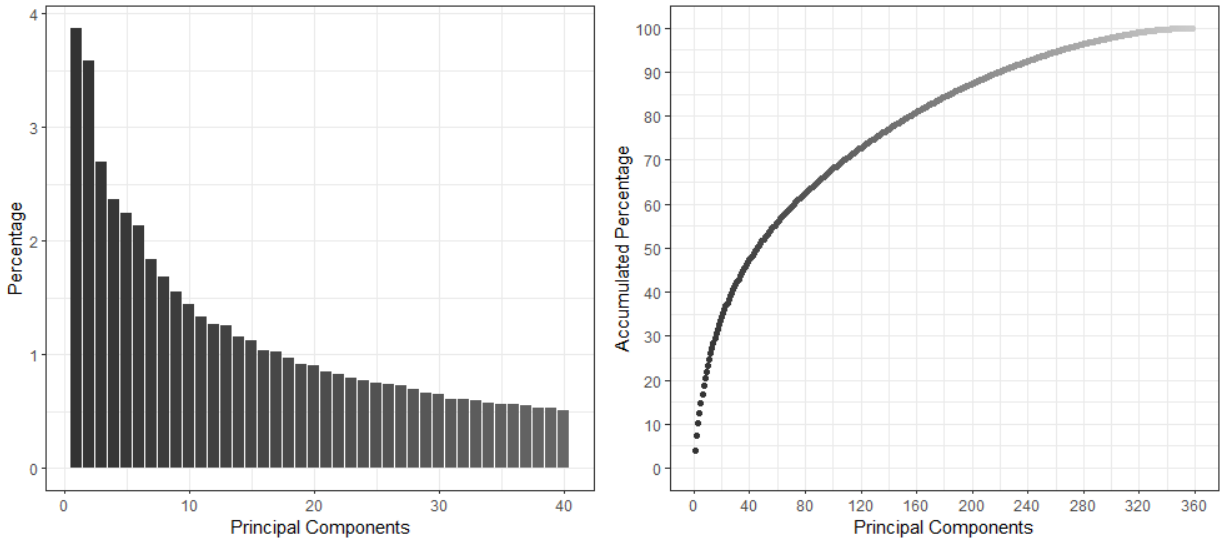

Figure 8. Principal component analysis for a tropical maize diversity panel of 360 inbred lines. A) percentage of genetic variation explained by the 40 first principal components; B) accumulated percentage of genetic variation explained by the principal components.

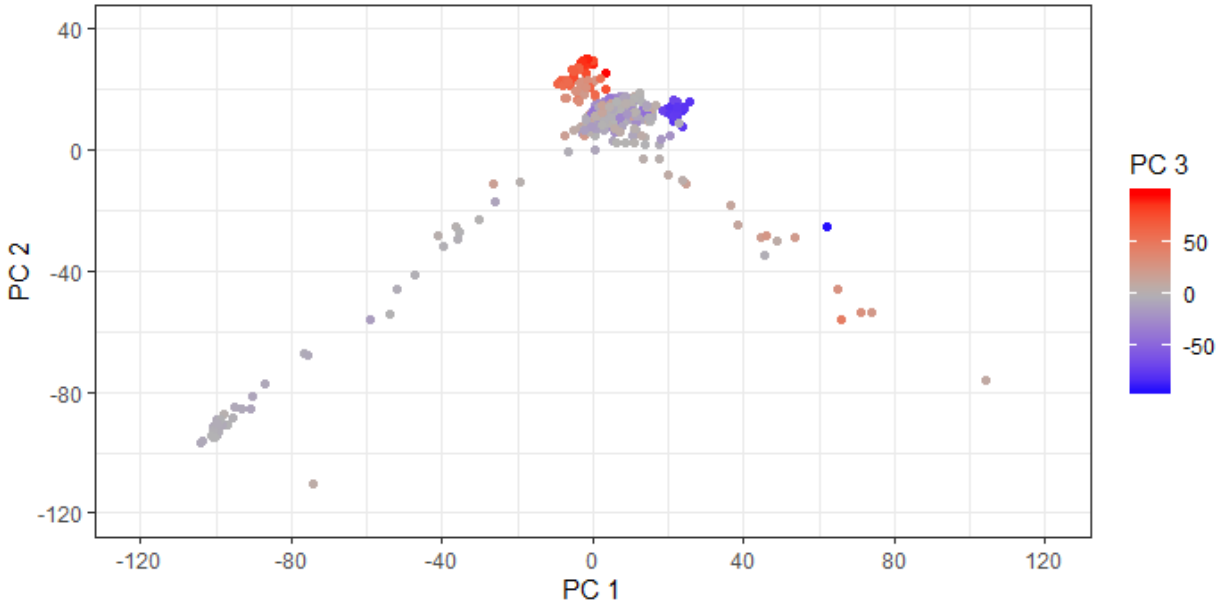

Figure 9. Distribution of the three main principal components (PC) for a tropical maize diversity panel of 360 lines. The genotypes L145 (PC 1 = 449.4, PC 2 = -433.1, PC 3 = 43.1) and L93 (PC 1 = 200.1 , PC 2 = -191.0, PC 3 = 83.2) are not present in the plot.

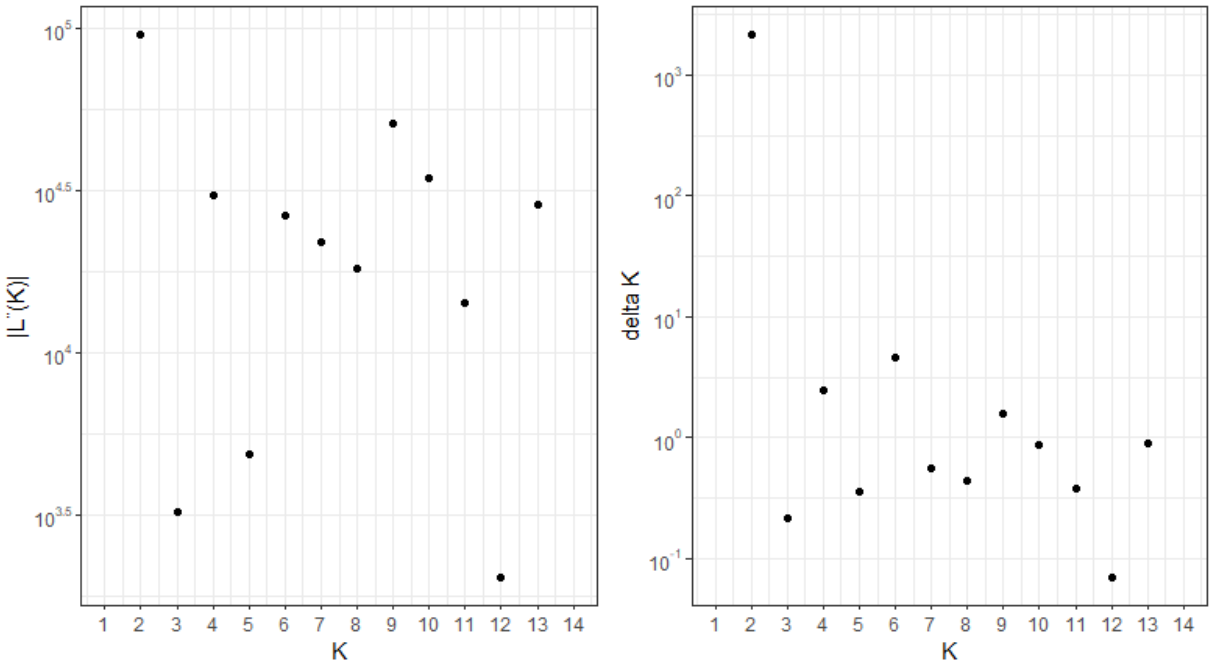

Figure 10. Evanno`s criteria for the indication of the best number of subpopulations (K) to be considered in the following step (Structure analysis) in the tropical maize diverse panel of 360 inbred lines. Left: criterion |L''(K)|; right: delta K

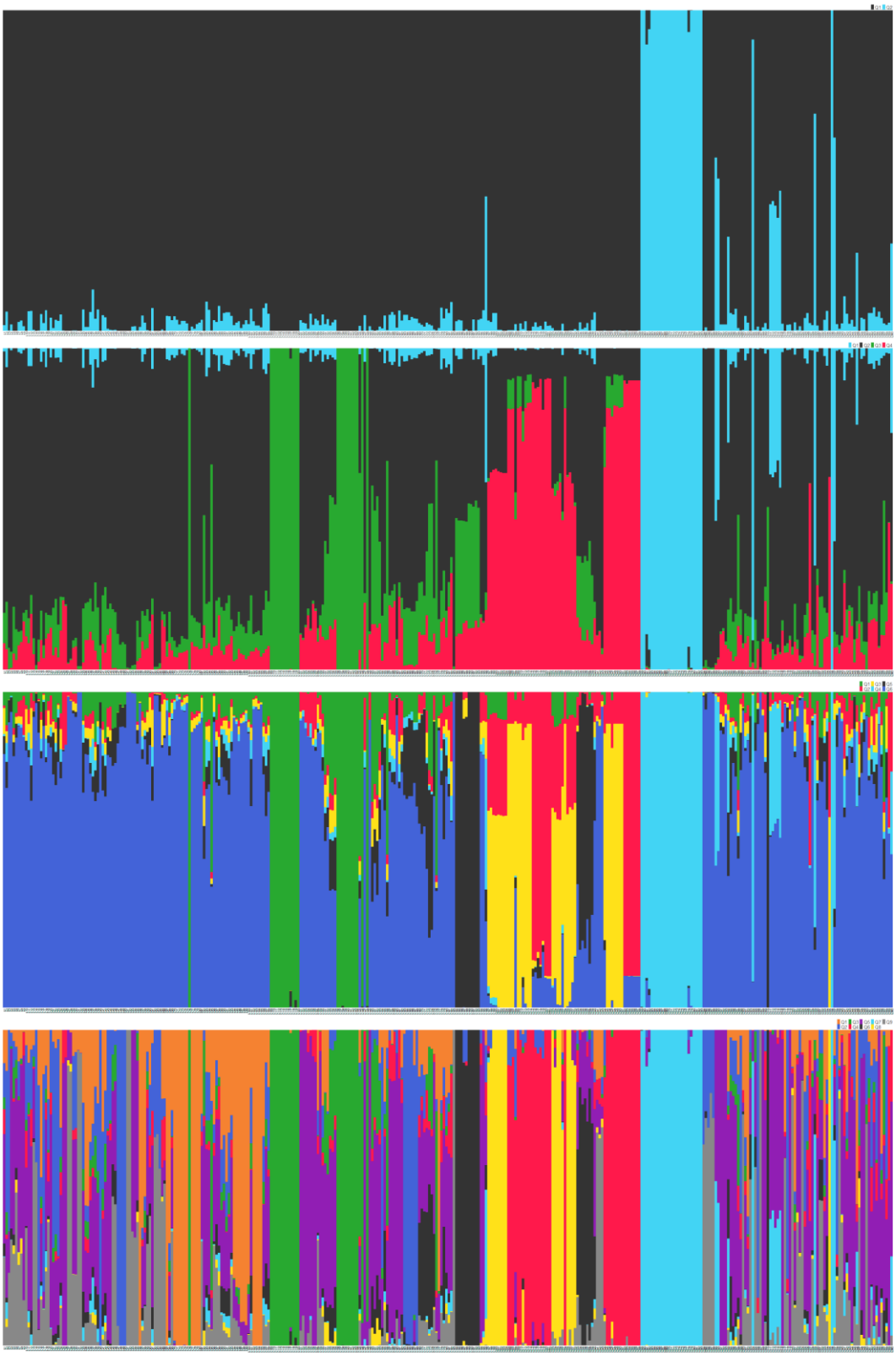

Figure 11. Structure analysis for the tropical maize diversity panel of 360 inbred lines. From top to bottom: structure division for 2, 4, 6 and 9 subpopulations (K). Lines are ordered from left to right from L1 to L360.

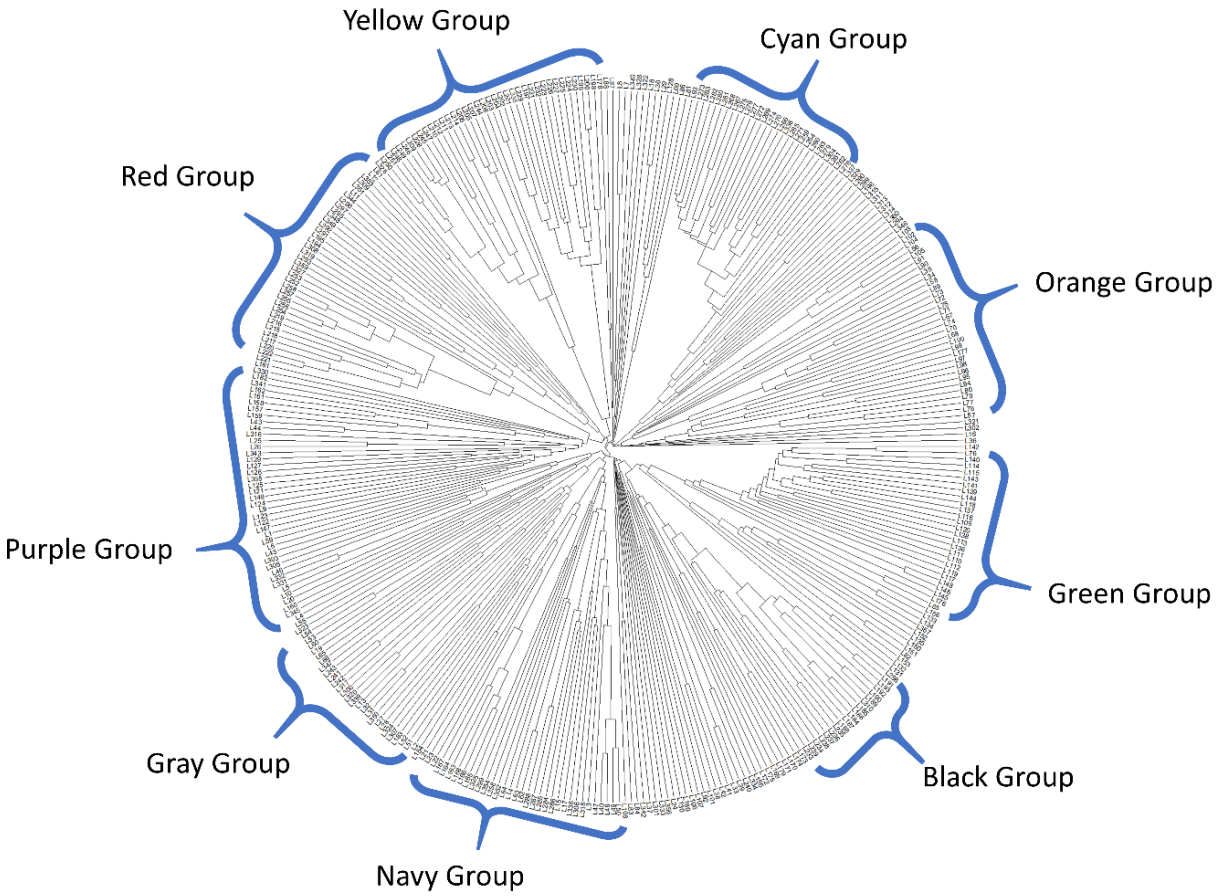

Figure 12. Dendrogram of the tropical maize diversity of 360 inbred lines. The 09 subpopulations indicated by the blue arrows were designated according to the Structure analysis (K=9).

Tables

Table 1. Identification and origin for the 360 inbred lines that compose the tropical maize diversity panel. Grain color and type were obtained through seed samples. Group was obtained by Structure analysis with K=9, value of Q higher than 0.5 attributed the genotype to the given population.

Line	Passport ID	Origin	Origin (alt.)	Grain Type	Grain Color	Group
L1	L-001	LP1 (08-841-3)	34M11	Flint	Red	M
L2	L-002	LP2 (08-1263-1)		Dent	Yellow	M
L3	L-003	LP3 (08-1400-2)		Flint	Orange	2 - Navy
L4	L-004	LP4 (08-1424-2)		Flint	Yellow	5 - Purple
L5	L-005	LP5 (08-1460)		Flint	Orange	M
L6	L-006	LP6 (08-1807-2)	Forte	SemiFlint	Orange	M
L7	L-007	LP7 (08-1816-2)	XB 8028	Flint	Orange	M
L8	L-008	LN1 (08-820-1)		SemiDent	Yellow	5 - Purple
L9	L-009	LN2 (08-840-2)		Flint	Orange	5 - Purple
L10	L-010	LN3 (08-894-3)		Flint	Orange	M
L11	L-011	LN4 (08-982)		Flint	Orange	5 - Purple
L12	L-012	LN6 (08-1029-1)		Flint	Orange	5 - Purple
L13	L-013	LN7 (08-1191-2)		SemiFlint	Orange	9 - Gray
L14	L-014	1378-1	CO 32	SemiFlint	Orange	9 - Gray
L15	L-015	1309-1	Z8420	SemiFlint	Orange	2 - Navy
L16	L-016	974		Flint	Yellow	M
L17	L-017	1309-1	Z8420	Flint	Orange	2 - Navy
L18	L-018	1134-1		Flint	Orange	M
L19	L-019	09-935-1	DKB350	Flint	Yellow	M
L20	L-020	4580-2	DKB C901	Flint	Yellow	5 - Purple
L21	L-021	4585-3		Flint	OrangeYellow	2 - Navy
L22	L-022	4586-2		SemiFlint	Orange	2 - Navy
L23	L-023	4587-2		Flint	Orange	2 - Navy
L24	L-024	4590-3		Flint	Orange	M
L25	L-025	4593-1	AG 9010	Flint	Orange	5 - Purple
L26	L-026	4593-3	AG 9010	SemiFlint	Orange	5 - Purple
L27	L-027	4596-1	CO 32	SemiDent	Yellow	9 - Gray
L28	L-028	4596-2	CO 32	SemiDent	Orange	9 - Gray
L29	L-029	4598-1		SemiFlint	Orange	M
L30	L-030	4604-2		Flint	Orange	M
L31	L-031	4605-1	Z8420	Flint	Orange	9 - Gray
L32	L-032	4606-2		Flint	OrangeYellow	9 - Gray
L33	L-033	4610-2		SemiDent	Yellow	M
L34	L-034	4621-1	BR201	SemiDent	Orange	1 - Orange
L35	L-035	4622-2		Dent	Orange	1 - Orange
L36	L-036	4626-2	AL 30	Flint	Orange	M
L37	L-037	4631-1		SemiDent	Yellow	M
L38	L-038	4634-2	AL 30	Flint	Orange	M
L39	L-039	4649-2	BRCMS28	Flint	Yellow	M
L40	L-040	4653-3		Flint	Orange	2 - Navy
L41	L-041	4654-2		Flint	Orange	M
L42	L-042	4657-5		Flint	Orange	M
L43	L-043	4667-2		Flint	Orange	5 - Purple
L44	L-044	4669-5	Z8480	Flint	Orange	5 - Purple
L45	L-045	4670-2	Z8480	Flint	Orange	M
L46	L-046	4672-2	P 30F88	SemiFlint	Orange	M
L47	L-047	4673-2	P 30F88	Flint	Orange	2 - Navy
L48	L-048	4674-2		Flint	Orange	2 - Navy
L49	L-049	4675-2		Flint	Orange	2 - Navy
L50	L-050	4678-2	DKB C333	Flint	Orange	2 - Navy
L51	L-051	4680-3	DKB C333	SemiDent	Yellow	9 - Gray

L52	L-052	4681-2		Flint	Orange	9 - Gray
L53	L-053	4687-1		Flint	Orange	5 - Purple
L54	L-054	4687-2		Flint	Orange	5 - Purple
L55	L-055	4690-2		SemiFlint	Orange	M
L56	L-056	5071-2		SemiDent	Yellow	1 - Orange
L57	L-057	5391-2		SemiFlint	Yellow	1 - Orange
L58	L-058	5399-2		Flint	Orange	1 - Orange
L59	L-001B	LP1 (08-841-3)	34M11	SemiFlint	Orange	M
L60	L-007B	LP7 (08-1816-2)	XB 8028	Flint	Orange	M
L61	L-012B	LN6 (08-1029-1)		SemiDent	Orange	5 - Purple
L62	L-014B	1378-1	CO 32	Flint	Orange	9 - Gray
L63	L-014C	1378-1	CO 32	Flint	Orange	9 - Gray
L64	L-016B	974		SemiFlint	Orange	9 - Gray
L65	L-019B	09-935-1	DKB350	SemiFlint	Orange	M
L66	L-019C	09-935-1	DKB350	Flint	LightYellow	9 - Gray
L67	16-02D	IG-2		Flint	Orange	1 - Orange
L68	25-04D	IG-2		Flint	Orange	1 - Orange
L69	39-05D	IG-2		Dent	Orange	M
L70	55-02D	IG-2		Flint	Orange	1 - Orange
L71	66-08D	IG-2		SemiFlint	Orange	1 - Orange
L72	94-02D	IG-2		SemiDent	Yellow	1 - Orange
L73	102-2D	IG-2		Flint	Orange	1 - Orange
L74	120-04F	IG-2		Flint	LightYellow	1 - Orange
L75	149-05D	IG-2		SemiFlint	Orange	1 - Orange
L76	168-D	IG-2		Flint	Orange	3 - Green
L77	20-02R	HS-1		Flint	Orange	1 - Orange
L78	30-07F	HS-1		Flint	Orange	1 - Orange
L79	31-01F	HS-1		Flint	Orange	1 - Orange
L80	33-04D	HS-1		SemiFlint	Orange	1 - Orange
L81	36-07F	HS-1		Dent	Yellow	M
L82	04-05F	IG-1		Flint	Orange	5 - Purple
L83	08-05F	IG-1		SemiFlint	Orange	M
L84	88-05F	IG-1		Flint	Orange	M
L85	8F	L-560		SemiFlint	Orange	3 - Green
L86	14D	L-560		Flint	Orange	M
L87	56D	L-560		Flint	Orange	M
L88	128D	L-560		SemiDent	Orange	M
L89	45-03D	TOPCROSS		Flint	Yellow	M
L90	49-02D	TOPCROSS		Flint	Orange	M
L91	53-01F	TOPCROSS		Flint	Orange	M
L92	82-01D	TOPCROSS		Flint	Orange	M
L93	131-01F	TOPCROSS		Flint	Orange	M
L94	16-04R	BR-201		Flint	Red	1 - Orange
L95	16-07R	BR-201		Flint	Orange	1 - Orange
L96	22-02D	BR-201		Flint	Yellow	1 - Orange
L97	23-05D	BR-201		Flint	LightYellow	1 - Orange
L98	24-03D	BR-201		Flint	Yellow	1 - Orange
L99	29-03D	BR-201		Flint	Orange	1 - Orange
L100	35-04F	BR-201		SemiFlint	Orange	1 - Orange
L101	18-08AF	CMS-05		Flint	Orange	M
L102	37-02BD	CMS-05		Flint	Orange	1 - Orange
L103	37-03BD	CMS-05		Flint	Orange	1 - Orange
L104	37-04BF	CMS-05		Flint	Orange	1 - Orange
L105	37-07BD	CMS-05		Flint	Orange	1 - Orange
L106	84-03F	TOPCROSS		Flint	Orange	M
L107	08-04BD	CMS-05		Flint	Orange	M
L108	20-01F	IG-1		Flint	Orange	M
L109	02-03D	IG-1		SemiFlint	Orange	3 - Green
L110	6	ICI-8452		Flint	Orange	3 - Green

L111	69	ICI-8452		Flint	Orange	3 - Green
L112	94-01	ICI-8452		Flint	Yellow	3 - Green
L113	112-02	ICI-8452		Flint	Orange	3 - Green
L114	123-02	ICI-8452		SemiFlint	Orange	3 - Green
L115	142	ICI-8452		SemiFlint	Orange	3 - Green
L116	172S5	ICI-8452		Flint	Orange	3 - Green
L117	179	ICI-8452		SemiFlint	Orange	3 - Green
L118	180-02	ICI-8452		Flint	Orange	3 - Green
L119	192S5	ICI-8452		Flint	Orange	3 - Green
L120	194-02D	ICI-8452		SemiFlint	Orange	3 - Green
L121	15	C-701		Flint	Orange	5 - Purple
L122	31	C-701		Flint	Orange	5 - Purple
L123	37	C-701		SemiFlint	Orange	5 - Purple
L124	38	C-701		Flint	Orange	5 - Purple
L125	38-01	C-701		Flint	Orange	5 - Purple
L126	34	C-701		Flint	Orange	5 - Purple
L127	75-01	C-701		Dent	LightYellow	5 - Purple
L128	99	C-701		SemiFlint	Orange	M
L129	131	C-701		Flint	OrangeYellow	5 - Purple
L130	143	C-701		SemiFlint	Orange	M
L131	23	ICI-8392FLINT		Flint	Orange	M
L132	77-01	ICI-8392FLINT		Flint	Orange	M
L133	154-02	ICI-8392FLINT		Flint	Orange	M
L134	155-02S5	ICI-8392FLINT		Flint	Orange	M
L135	175-02S5	ICI-8392FLINT		SemiFlint	Orange	M
L136	4	ICI-8452		Flint	Orange	3 - Green
L137	17-02	ICI-8452		SemiFlint	Orange	3 - Green
L138	52-01	ICI-8452		Flint	Orange	3 - Green
L139	83	ICI-8452		Flint	Orange	3 - Green
L140	103	ICI-8452		Flint	Orange	3 - Green
L141	141	ICI-8452		SemiFlint	Orange	3 - Green
L142	174	ICI-8452		SemiFlint	Orange	3 - Green
L143	176-01	ICI-8452		Flint	Orange	3 - Green
L144	180-01	ICI-8452		Flint	Orange	3 - Green
L145	186-01	ICI-8452		Flint	Orange	3 - Green
L146	204	ICI-8452		Flint	Orange	3 - Green
L147	41-01	C-701		Flint	Orange	5 - Purple
L148	43-01	C-701		Flint	Orange	3 - Green
L149	51-02	C-701		Flint	Orange	5 - Purple
L150	84	ICI-8392		Flint	Orange	M
L151	131	ICI-8392		Flint	Orange	5 - Purple
L152	155-01	ICI-8392		Flint	Orange	M
L153	159-02	ICI-8392		Flint	Orange	M
L154	61-02	C-791151		Flint	Orange	2 - Navy
L155	89	C-505		Flint	Orange	M
L156	60	ICI-8447		Flint	Orange	M
L157	2	C-808		SemiFlint	Orange	5 - Purple
L158	3	C-808		Flint	Orange	5 - Purple
L159	4	C-808		Flint	Orange	5 - Purple
L160	5	C-808		Flint	Orange	5 - Purple
L161	6S6	C-808		SemiFlint	Orange	5 - Purple
L162	9	C-808		Flint	OrangeYellow	5 - Purple
L163	2	PF-41X05-33-05B		Flint	Orange	2 - Navy
L164	3	PF-41X05-33-05B		Flint	Orange	2 - Navy
L165	5	PF-41X05-33-05B		Flint	Orange	2 - Navy
L166	6	PF-41X05-33-05B		Flint	Orange	2 - Navy
L167	7	PF-41X05-33-05B		Flint	Orange	2 - Navy
L168	8	PF-41X05-33-05B		Flint	Orange	2 - Navy
L169	2	C-505		Flint	Orange	M

L170	3	C-505		SemiFlint	Orange	M
L171	4	C-505		Flint	Orange	M
L172	5	C-505		Flint	Orange	M
L173	6	C-505		Flint	Orange	M
L174	7	C-505		Flint	Orange	M
L175	8	C-505		Flint	Orange	5 - Purple
L176	L-ASO			Flint	Orange	3 - Green
L177	L-46-10D	BR-201		Flint	Orange	1 - Orange
L178	61-02F	TOPCROSS		SemiDent	OrangeYellow	M
L179	1	C-505		Flint	Orange	M
L180	2024	L 102		Flint	Orange	M
L181	2018	L 63		Flint	Orange	M
L182	2373	C 113		Flint	LightYellow	M
L183	8501	C 333		SemiFlint	Orange	9 - Gray
L184	LT53-001	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L185	LT53-002	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L186	LT53-003	ST X (P 30 F 53H X ST)		SemiFlint	Orange	6 - Black
L187	LT53-006	ST X (P 30 F 53H X ST)		Flint	OrangeYellow	6 - Black
L188	LT53-010	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L189	LT53-013	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L190	LT53-014	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L191	LT53-018	ST X (P 30 F 53H X ST)		SemiFlint	Orange	6 - Black
L192	LT53-019	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L193	LT53-024	ST X (P 30 F 53H X ST)		Flint	Orange	6 - Black
L194	LT6F-005	T1056 (P 30 F 53 H x T 1056)		Flint	Yellow	5 - Purple
L195	LT6F-007	T1056 (P 30 F 53 H x T 1056)		Flint	Orange	5 - Purple
L196	LD83-002	D 480 X (P 30 F 53H X D 480)		Dent	Orange	M
L197	LM2P-005	MC 02 X (P 30 F 53H X MC 02)		Flint	Yellow	8 - Yellow
L198	LM2P-010	MC 02 X (P 30 F 53H X MC 02)		Flint	Orange	8 - Yellow
L199	LM2P-013	MC 02 X (P 30 F 53H X MC 02)		Flint	Orange	8 - Yellow
L200	LM2P-014	MC 02 X (P 30 F 53H X MC 02)		Flint	Orange	8 - Yellow
L201	LM2P-021	MC 02 X (P 30 F 53H X MC 02)		Flint	Orange	8 - Yellow
L202	LM2P-022	MC 02 X (P 30 F 53H X MC 02)		Dent	OrangeYellow	8 - Yellow
L203	LM2P-023	MC 02 X (P 30 F 53H X MC 02)		Flint	Yellow	8 - Yellow
L204	LM2P-024	MC 02 X (P 30 F 53H X MC 02)		Flint	Orange	8 - Yellow
L205	LM4P-003	MC 04 X (P 30 F 53H X MC 04)		Flint	Orange	4 - Red
L206	LM4P-004	MC 04 X (P 30 F 53H X MC 04)		Flint	Orange	4 - Red
L207	LM4P-005	MC 04 X (P 30 F 53H X MC 04)		Flint	Orange	4 - Red
L208	LM4P-011	MC 04 X (P 30 F 53H X MC 04)		Flint	Orange	M
L209	LM4P-012	MC 04 X (P 30 F 53H X MC 04)		SemiFlint	Orange	4 - Red
L210	LM4P-014	MC 04 X (P 30 F 53H X MC 04)		Dent	Yellow	4 - Red
L211	LM4P-015	MC 04 X (P 30 F 53H X MC 04)		SemiFlint	Orange	4 - Red
L212	LM4P-017	MC 04 X (P 30 F 53H X MC 04)		Flint	Orange	4 - Red
L213	LM4P-021	MC 04 X (P 30 F 53H X MC 04)		Dent	Orange	4 - Red
L214	LM4P-022	MC 04 X (P 30 F 53H X MC 04)		Dent	Orange	4 - Red
L215	LR3F-003	RSM X (P 30 F 53H X RSM)		Dent	Yellow	4 - Red
L216	LR3F-008	RSM X (P 30 F 53H X RSM)		Dent	Yellow	4 - Red
L217	LR3F-012	RSM X (P 30 F 53H X RSM)		Dent	Yellow	4 - Red
L218	LR3F-013	RSM X (P 30 F 53H X RSM)		Dent	Orange	4 - Red
L219	LR3F-021	RSM X (P 30 F 53H X RSM)		Dent	Yellow	4 - Red
L220	LR3F-025	RSM X (P 30 F 53H X RSM)		Dent	Yellow	4 - Red
L221	LR3F-027	RSM X (P 30 F 53H X RSM)		Dent	Orange	4 - Red
L222	LR3F-028	RSM X (P 30 F 53H X RSM)		Dent	Orange	4 - Red
L223	LM2B-001	MC 02 X (2B 710 H X MC 02)		SemiDent	Yellow	8 - Yellow
L224	LM2B-004	MC 02 X (2B 710 H X MC 02)		SemiFlint	Yellow	8 - Yellow
L225	LM2B-005	MC 02 X (2B 710 H X MC 02)		SemiFlint	Yellow	8 - Yellow
L226	LM2B-008	MC 02 X (2B 710 H X MC 02)		Dent	Yellow	8 - Yellow
L227	LM2B-009	MC 02 X (2B 710 H X MC 02)		SemiDent	Yellow	8 - Yellow
L228	LM2B-011	MC 02 X (2B 710 H X MC 02)		Flint	Orange	4 - Red

L229	LM2B-013	MC 02 X (2B 710 H X MC 02)		Flint	Orange	8 - Yellow
L230	LM2B-019	MC 02 X (2B 710 H X MC 02)		Dent	Orange	8 - Yellow
L231	LM2B-031	MC 02 X (2B 710 H X MC 02)		Dent	Yellow	8 - Yellow
L232	LM2B-034	MC 02 X (2B 710 H X MC 02)		Flint	Yellow	8 - Yellow
L233	LG8B-008	LG 8 X (2B 710 H X LG 8)		Flint	Orange	M
L234	LG8B-011	LG 8 X (2B 710 H X LG 8)		SemiFlint	Orange	6 - Black
L235	LG8B-014	LG 8 X (2B 710 H X LG 8)		Dent	Orange	6 - Black
L236	LG8B-015	LG 8 X (2B 710 H X LG 8)		Flint	Orange	6 - Black
L237	LG8B-017	LG 8 X (2B 710 H X LG 8)		Dent	Yellow	6 - Black
L238	LG8B-019	LG 8 X (2B 710 H X LG 8)		Flint	Orange	6 - Black
L239	LG8B-022	LG 8 X (2B 710 H X LG 8)		Flint	Orange	6 - Black
L240	LG8B-028	LG 8 X (2B 710 H X LG 8)		Flint	Orange	M
L241	LG3B-001	LG 13 X (2B 710 H X LG 13)		Flint	Orange	9 - Gray
L242	LG3B-002	LG 13 X (2B 710 H X LG 13)		SemiFlint	Orange	9 - Gray
L243	LG3B-006	LG 13 X (2B 710 H X LG 13)		SemiFlint	Orange	9 - Gray
L244	LM4B-001	MC 04 X (2B 710 H X MC 04)		Dent	Orange	4 - Red
L245	LM4B-011	MC 04 X (2B 710 H X MC 04)		Flint	Orange	4 - Red
L246	LM4B-012	MC 04 X (2B 710 H X MC 04)		SemiFlint	Orange	4 - Red
L247	LM4B-017	MC 04 X (2B 710 H X MC 04)		Flint	Orange	4 - Red
L248	LM4B-025	MC 04 X (2B 710 H X MC 04)		Dent	Orange	4 - Red
L249	LM4B-027	MC 04 X (2B 710 H X MC 04)		SemiFlint	Orange	4 - Red
L250	LM4B-028	MC 04 X (2B 710 H X MC 04)		SemiFlint	Orange	4 - Red
L251	LM4B-030	MC 04 X (2B 710 H X MC 04)		SemiFlint	Orange	4 - Red
L252	LR2B-003	RSM X (2B 710 H X RSM)		Dent	Yellow	4 - Red
L253	LR2B-004	RSM X (2B 710 H X RSM)		SemiFlint	Yellow	4 - Red
L254	LR2B-006	RSM X (2B 710 H X RSM)		Dent	Yellow	4 - Red
L255	LR2B-007	RSM X (2B 710 H X RSM)		Dent	Yellow	4 - Red
L256	LR2B-008	RSM X (2B 710 H X RSM)		Dent	Yellow	4 - Red
L257	LR2B-009	RSM X (2B 710 H X RSM)		SemiFlint	Yellow	4 - Red
L258	LR2B-010	RSM X (2B 710 H X RSM)		Dent	Yellow	4 - Red
L259	LD85-002	D 480 X D 505		Dent	Orange	7 - Cyan
L260	LD85-004	D 480 X D 505		Dent	Orange	7 - Cyan
L261	LD83-004	D 480 X D 523		Flint	Orange	7 - Cyan
L262	LD83-011	D 480 X D 523		SemiFlint	Orange	7 - Cyan
L263	LD58-002	D 505 X D 480		Dent	LightYellow	7 - Cyan
L264	LD58-010	D 505 X D 480		SemiDent	Orange	7 - Cyan
L265	LD58-012	D 505 X D 480		SemiDent	Orange	7 - Cyan
L266	LD56-001	D 505 X D 603		Dent	Yellow	7 - Cyan
L267	LD56-003	D 505 X D 603		SemiFlint	Yellow	7 - Cyan
L268	LD56-004	D 505 X D 603		Flint	Yellow	7 - Cyan
L269	LD56-005	D 505 X D 603		Dent	Orange	7 - Cyan
L270	LD56-009	D 505 X D 603		Flint	Yellow	7 - Cyan
L271	LD56-012	D 505 X D 603		Flint	Orange	7 - Cyan
L272	LD56-013	D 505 X D 603		Dent	Orange	7 - Cyan
L273	LD56-014	D 505 X D 603		Flint	Yellow	7 - Cyan
L274	LD56-020	D 505 X D 603		Dent	Orange	7 - Cyan
L275	LD56-025	D 505 X D 603		SemiDent	Orange	7 - Cyan
L276	LD56-029	D 505 X D 603		Flint	Orange	7 - Cyan
L277	LD56-031	D 505 X D 603		Flint	Yellow	7 - Cyan
L278	LD25-002	D 523 X D 505		Dent	Yellow	7 - Cyan
L279	LD25-013	D 523 X D 505		Flint	Orange	7 - Cyan
L280	LD65-001	D 603 X D 505		Flint	Yellow	7 - Cyan
L281	LD65-003	D 603 X D 505		Flint	Orange	7 - Cyan
L282	LD65-006	D 603 X D 505		Flint	Yellow	7 - Cyan
L283	LD65-008	D 603 X D 505		Flint	Yellow	7 - Cyan
L284	LF38-003	F 932 X F 983		Flint	Yellow	9 - Gray
L285	LF38-009	F 932 X F 983		Flint	Yellow	9 - Gray
L286	LF38-011	F 932 X F 983		Flint	Yellow	9 - Gray
L287	LF38-012	F 932 X F 983		Flint	Yellow	9 - Gray

L288	LF38-014	F 932 X F 983		Flint	Yellow	9 - Gray
L289	LT7T-002	T1147 X T 1056		Dent	Orange	7 - Cyan
L290	LT7T-003	T1147 X T 1056		SemiFlint	Orange	5 - Purple
L291	LT7T-004	T1147 X T 1056		Flint	Orange	5 - Purple
L292	LT7T-005	T1147 X T 1056		Dent	Orange	5 - Purple
L293	LT7T-008	T1147 X T 1056		Flint	Orange	5 - Purple
L294	5506	AS 1548		Flint	Red	M
L295	5132	FB 61		Flint	Orange	5 - Purple
L296	5292	FB 84		Flint	Orange	5 - Purple
L297	5041	MB 51		Flint	Orange	5 - Purple
L298	5211	MB 61		Flint	Orange	M
L299	5451	MC 08		Flint	Orange	2 - Navy
L300	1498	BRS 1001		Dent	Orange	9 - Gray
L301	1520	MC 06		Dent	Yellow	M
L302	1509	ND 970		Dent	Yellow	1 - Orange
L303	1525	Z 627		SemiDent	Orange	9 - Gray
L304	1533	D 603		Dent	Orange	7 - Cyan
L305	1562	POPZ 658		Dent	Orange	9 - Gray
L306	1629	F 948		Dent	Orange	2 - Navy
L307	1640	C 333 B82		SemiFlint	Orange	9 - Gray
L308	1643	Tork 234		Flint	Orange	5 - Purple
L309	1723	FB 51		Dent	Orange	5 - Purple
L310	1765	ST		Dent	Orange	6 - Black
L311	1830	DG 12		Flint	Orange	5 - Purple
L312	1836	DG 13		Flint	Orange	5 - Purple
L313	1834	DG 15		Flint	Orange	5 - Purple
L314	1826	DG 29		Dent	Orange	5 - Purple
L315	1828	DG 7		Flint	Orange	5 - Purple
L316	1837	POP A 113		Flint	Orange	5 - Purple
L317	1859	BRS 1010		Dent	Orange	9 - Gray
L318	1852	F 978		Dent	Yellow	2 - Navy
L319	1849	T1147		Dent	Orange	5 - Purple
L320	1851	TR3 170		Dent	Orange	9 - Gray
L321	1932	A 2555		Dent	Orange	1 - Orange
L322	1973	AG 8080		Dent	Orange	M
L323	1974	DKB 350		Flint	Orange	5 - Purple
L324	1970	DKB 390		Flint	Orange	9 - Gray
L325	2010	L 114		Flint	Orange	M
L326	2007	L 66		Dent	OrangeYellow	M
L327	2146	AG 7000		SemiFlint	Orange	4 - Red
L328	2253	AG 7575		Flint	Orange	5 - Purple
L329	2230	D 549		Flint	Orange	7 - Cyan
L330	2250	DKB 909		Flint	Orange	5 - Purple
L331	2256	DOW 8480		Dent	Orange	M
L332	2260	L 175 COOPANOR		Dent	Yellow	M
L333	2217	LG 8		Flint	Orange	M
L334	2188	M C02		SemiDent	Orange	M
L335	2297	AG 3		Flint	Orange	8 - Yellow
L336	2295	D 505		Dent	Orange	7 - Cyan
L337	2296	D 606		Flint	Orange	7 - Cyan
L338	2276	F 983		Flint	Yellow	2 - Navy
L339	2300	HSTR 3 170		Flint	Orange	9 - Gray
L340	2293	STR		Dent	Yellow	5 - Purple
L341	2372	C 85		Dent	Yellow	M
L342	2369	L 58		Flint	Orange	M
L343	2391	LG 02		Dent	Orange	5 - Purple
L344	2406	LG 09		Flint	Orange	5 - Purple
L345	2395	LG 04		Flint	Orange	5 - Purple
L346	2400	LG 06		SemiDent	Red	1 - Orange

L347	2401	LG 07		Dent	Yellow	M
L348	2408	LG 10		Flint	Orange	5 - Purple
L349	2410	LG 11		Flint	Orange	5 - Purple
L350	2411	LG 12		SemiDent	Orange	9 - Gray
L351	2413	LG 14		Dent	Orange	5 - Purple
L352	2382	NS 016		Dent	Orange	M
L353	2480	Status TL		SemiFlint	Orange	5 - Purple
L354	3088	AC 58		Flint	Yellow	M
L355	3093	BA 8		SemiDent	Orange	5 - Purple
L356	3090	FF 6		Dent	Orange	M
L357	3092	II 47		Flint	Orange	5 - Purple
L358	3094	MT 05/06		Flint	Orange	M
L359	694	AG 6		Dent	Orange	5 - Purple
L360	8499	AG 8		Flint	Orange	M

3. DISSECTING THE RESISTANCE TO THE CORN STUNT DISEASE COMPLEX IN A TROPICAL MAIZE DIVERSITY PANEL THROUGH A GENOME WIDE ASSOCIATION STUDY

ABSTRACT

The corn stunt disease and the maize bushy stunt disease, jointly called corn stunt disease complex, are increasing in agronomical importance due to the recent significative yield losses. In parallel, the tropical maize diversity is a valuable germplasm resource to deal with new agricultural challenges, but it is not fully explored or comprehended. Therefore, this work aimed to identify genes associated to the resistance of corn stunt disease complex and potential sources of resistance in a Brazilian tropical maize diversity panel. For that, we used 360 tropical maize inbred lines genotyped using 14,655 high-quality SNP markers and evaluate for the traits proportion of survivor plants (PSP), sanity score in survivor plants (SSSP), and whole sanity score (WSS) in two sites in São Paulo state, Brazil. Genome wide association study (GWAS) was performed incorporating a genomic relationship matrix and 3 main principal components to account for panel structure. In total, 13 markers were significant to the traits and presented functions mainly related to the cellulose metabolism, the auxin pathway, genes of defense and response associated to phagocyte oxidase activity, and anthocyanin production. The candidate genes were associated to response of glucose accumulation in leaves, reduction of auxin content in plant, and direct defense-attack against the pathogens. For each of the traits, we found lines with the totality of favorable alleles in homozygosis, which would facilitate transfering resistance genes to other genotypes. Additionally, the lines from the ancestral PF-41X05-33-05B exhibited high content of favorable alleles to all the traits simultaneously. Our findings reveal underlying genetic mechanisms triggered by plants in response to the corn stunt disease complex and allowed the identification of potential resistant inbred lines. These results can substantially improve the genetic gains when incorporated in breeding programs and constitute an important contribution to the genetic comprehension of the tropical maize germplasm.

Keywords: 1. *Zea mays*, 2. Resistance, 3. *Spiroplasma kunkelii*, 4. *Maize bushy stunt phytoplasma*, 5. Germplasm, 6. GWAS

3.1. Introduction

The Corn Stunt Disease (CSD) and Maize Bushy Stunt Disease (MBSD) are diseases that occur in maize due to mollicutes, microorganisms from a bacteria domain which comprises individuals without cell wall (Bergamin Filho et al., 1995; Kimati et al., 2005). The first disease is caused by the pathogen *Spiroplasma kunkelii*, a prokaryote with spiral form, and the second one by the *Maize Bushy Stunt Phytoplasma* (MBSP, also known as *Candidatus Phytoplasma asteris*), a phytoplasma with oval-spherical form. In average, this complex of diseases can imply losses of 50% and can reach complete production loss in some cases (Gordon et al., 1981; MASSOLA JÚNIOR, N. S.; BEDENDO & AMORIM, L.; LOPES, 1999).

These pathogens are transmitted by the corn leafhopper *Dalbulus maidis* (Delong &Wolcott), a yellow pale hemipteran, averaging between 3.7 mm to 4.3 mm long, very agile and usually habits the maize whorl (Chales Martins de Oliveira et al., 2017). The disease cycle starts

when the hemipteran feeds from a diseased plant, then the pathogens infects salivary glands of vector and stays latent multiplying for 3 to 4 weeks (Moya-Raygoza & Nault, 1998; Özbek et al., 2003) . After this period, when the insect feeds from a health plant, the pathogens are transmitted. Inside the plants, the mollicutes multiply in phloem vessels and spread along the all plant tissues (GUSSIE, J. S.; FLECHER, J.; CLAYPOOL, 1995).

Both diseases cause dwarfism, abnormal prolificity, spike and grain deformation, chlorosis of leaves and vegetative organs, smaller number of roots, premature plant death, and consequently yield loss (Chang, 1998; Gonzalez et al., 2018; Elizabeth De Oliveira et al., 2003). Usually the symptom of pallid chlorosis is associated to *S. kunkelii*, and chlorosis combined with anthocyanin in the leaves margin are related to MBSP. However, these differences are not sufficient to identify the causal pathogen in field conditions (Nault, 1980; E. Oliveira et al., 2002). In addition to that, both diseases can occur simultaneously, making them to be treated as a disease complex called the corn stunt disease complex.

The management of the complex is usually performed with insecticide (Charles Martins De Oliveira et al., 2007) and/or biological control (Querino et al., 2017) to control the vector, and the use of genetic resistant hybrids (Elizabeth de Oliveira et al., 2013). The genetic resistance is an interesting approach since it reduces costs with pesticides and copes directly with the pathogens.

The corn stunt resistance is characterized by a quantitative genetic control and presents predominance of additive effects (Elizabeth de Oliveira et al., 2013; MÁRQUEZ SÁNCHEZ, 1982; Silva et al., 2003). The molecular changes promoted by the corn stunt complex increases the protein concentration and reduces the sugar and phenolic content, consequently affecting the photosynthesis and the leaves senescence (Junqueira et al., 2004). Furthermore, it can potentially decrease the auxin levels in the plant (Chang, 1998; Elizabeth De Oliveira et al., 2002).

The genetic mechanisms of resistance to corn stunt complex and the location of potential resistance genes in the tropical germplasm is not fully elucidated. In this study, we aimed to identify potential genes in tropical maize associated to resistance to the stunt disease complex and identify sources of resistance through a genomic wide association study.

3.2. Materials and methods

3.2.1. Plant material and genotypic data

To understand the maize resistance to the corn stunt disease complex, we used 360 maize inbred lines from a Brazilian tropical diversity panel. They were genotyped using healthy leaf samples in the V3 stage according the DNA extraction protocol of INGLIS et al. (2018). The

inbred lines were genotyped using the POLAND et al. (2012) genotyping-by-sequencing modified protocol with the restriction enzymes PstI and MseI. The SNP (Single Nucleotide Polymorphism) calling was performed with the software TASSEL 5 standalone V5.2.54 (Bradbury et al., 2007b) and the alignment of the reads was performed using the version 5.0 (released in 2020) of maize line B73 as genome reference with the software Bowtie2 V2.3.3.1 (Langmead & Salzberg, 2013) in module 'very-sensitive'. The markers obtained were filtered only for chromosomal markers and the missing data was imputed using the software Beagle 5.0 (B. L. Browning & Browning, 2008). Complementarily, the software VCFTools (Danecek et al., 2011) and the R packages synbreed (Wimmer et al., 2012) and snpReady (Granato et al., 2018) were used to filter for biallelic loci, minor allele frequency (MAF) higher than 0.05, call rate of 0.95 and linkage disequilibrium (LD) with r^2 lower than 0.99, and to convert the files to adequate formats. As result, it was obtained a dataset with 14,655 high-informative SNP markers. In addition, the additive genomic relationship matrix (VanRaden, 2008) and the main principal components distribution were analyzed regarding panel relationship and structure.

3.2.2. Phenotypic data

Serologic analysis was performed and confirmed the occurrence of both pathogens in the area (Supplementary Material 1 and 2). Then, to characterize the disease complex, three traits were evaluated in phenological stage R3:

(1) proportion of survivor plants (PSP), that accounts for the ratio between the number of plants with any active photosynthetic area over the number of all plants of the plot;

(2) sanity score in survivor plants (SSSP), which measures the proportion of active photosynthetic (health) area for the survivor plants. For that, we used a scale with six grades: 0.00 (absence of health area), 0.13 (1 – 25% of health area), 0.38 (26 – 50%), 0.63 (51 – 75%), 0.88 (76 – 99%) and 1.00 (100%); and

(3) whole sanity score (WSS), that accounts for the proportion of photosynthetic area of all plants in the plot, calculated directly by multiplication of PSP and SSSP.

Each plot was evaluated by two different evaluators and the mean of their scores composed the phenotypic value for the given trait and plot. The scores of the three traits ranged from 0.00 to 1.00, being 1.00 the healthier status.

The genotypes were evaluated in two sites in the second season ('safrinha') 2019, ESALQ (22º42'22.6"S, 47º38'16.6"W) and Anhembi (22º50'52.0"S, 48º01'06.1"), São Paulo State, Brazil. Each site had two replicates with 20 blocks each and each block contained 18 plots. To estimate

the genotypic value, we used the following mixed model using the R package ASREML version 4 (Butler et al., 2017):

$$Y = \mu + Dg + Es + Fn + Lr + Mb + \varepsilon \quad (1)$$

where Y is the phenotypic value for the given trait, μ is the constant common to all treatments, g is the fixed effect of genotype, s is the fixed effect of site, n is the interaction between genotype and site effect, r is the fixed effect of replication within site, b is the random effect of block within environment and replication with $b \sim N(0, \sigma_b^2)$, and ε is the model error with $\varepsilon \sim N(0, \sigma^2)$. D, E, F, L and M are the incidence matrixes for the associate effects. The fixed effects were tested with Wald test at 5% of significance.

3.2.3. Genomic wide association study

To perform the genome wide association study (GWAS), we used the R-environment package FarmCPU (X. Liu et al., 2016). Aiming to minimize the occurrence of spurious associations, the population structure (matrix Q, from the three main principal components) and the relationship between the lines (matrix K) were incorporated in the analysis through the following mixed model:

$$y = X\beta + \mathrm{Qs} + Zv + \varepsilon \quad (2)$$

where y is the vector of genotypic values (BLUEs) of the interested trait; β is the vector containing the fixed effects including the constant common to all treatments and effect of the genetic markers; s is the fixed effect of three main principal components; v is the random effect of genotypes weighted by the kinship matrix (K); ε is the residual vector. The distribution of random effects was:

$$\binom{v}{\varepsilon} \sim N_2\left(\binom{0}{0}, \begin{pmatrix} 2K\sigma_a^2 & 0 \\ 0 & I\sigma^2 \end{pmatrix}\right) \quad (3)$$

where σ_a^2 is the additive variance and σ^2 is the residual variance. X, Q and Z are the incidence matrixes for the given effects.

The significance of marker effects (m_i) was tested using the Bonferroni multiple comparison method with α_{global} of 0.01, and consequently, $\alpha_{local} = 6.81\ 10^{-7}$. The significant markers were screened using a 100 kb sliding window in order to identify co-localizing or adjacent candidate genes based on the public maize genome data set B73 reference genome assembly version 5 (released in 2020). Functions, gene ontology terms and metabolic pathways for the significant markers were accessed via KEGG (Kanehisa & Goto, 2000), InterPro (Mitchell et al.,

2019), MaizeGDB (Portwood et al., 2019), and GeneOntology (Ashburner et al., 2000) database banks.

3.3. Results

3.3.1. Genomic data and structure

The genotypes that compose the diversity panel present high genetic diversity, which can be observed in VanRaden additive genomic relationship matrix (VanRaden, 2008) (Figure 13). It is possible to observe some clusters of endogamic individuals, which evidences the need to consider the structuration of the panel in the following analysis.

The panel structure according the three main principal components (PC) is displayed in Figure 14. The first to the third PC's explained respectively 3.87%, 3.58% and 2.70% of the data variation, summarizing 10.15% in total. The genotype dispersion was concentered on three axes, suggesting the high structuration of the panel. To account for it, the first three PC were incorporated into the GWAS model.

3.3.2. ANOVA and exploratory analysis

The exploratory analysis demonstrates all fixed effects in the model (1) were significant according to Wald test. The mean for PSP, SSSP, and WSS was 0.426, 0.362, and 0.216, respectively. However, the intensity of the disease was higher in ESALQ (PSP = 0.261, SSSP = 0.260 and WSS = 0.113) compared to Anhumas (PSP = 0.592, SSSP = 0.465 and WSS = 0.320).

For all the traits, the genotype had a significant effect, suggesting the existence of genetic variability in the panel. The interaction genotype x site was also significant for the three traits, which indicates the complexity of the response to the disease complex and the potential high influence of the environment in the maize response. The Spearman's correlation coefficient of the adjusted mean of the genotypes between sites were moderate (0.539, 0.449 and 0.555 for PSP, SSSP and WSS, respectively) suggesting the prevalence of simple interactions.

3.3.3. GWAS

The QQ-plots for the traits indicated a good fit (Figure 15). The Manhattan plot for PSP presented six significant markers, each one in a different chromosome. The trait SSSP presented four significant markers, three of them in chromosome 7, and one of them was also significant for the trait PSP. Finally, four markers were significant for WSS. The markers, their position, effect, MAF, genotypic frequencies and significance are present in Table 2.

The effect of significant markers for the trait PSP ranged from 0.046 to 0.163. The favorable alleles were equally distributed between the common alleles (negative effect) and the less frequent ones (positive effect) in population. In addition, two significant markers for PSP presented elevated frequency of heterozygotes and low frequency of homozygote for the less frequent allele. The adjustment of the regression for the markers (R^2) in relation the BLUEs varied from 0.004 and 0.091 (Figure 16). For SSSP, the effects ranged from 0.029 to 0.055 and three out of four markers presented the favorable allele as the less frequent in the population. The R^2 for this trait ranged from 0.009 to 0.090. The trait WSS presented effects ranging from 0.029 to 0.057 and three out of four markers presented the favorable allele as the common allele in population. One marker had elevated heterozygosity and low frequency of homozygote for the less frequent allele. The R^2 for this trait varied from 0.035 to 0.096.

We calculated a predicted value for the genotypes considering the sum of only the significant marker effects. Based on this estimation and applying a intensity selection of 20% for each trait, only 13 lines (L1, L2, L10, L13, L46, L122, L163, L164, L165, L166, L167, L168 and L303) are classified simultaneously for three traits, and six of them presented as ancestor the parent PF-41X05-33-05B. It was also possible to identify 04 lines (L160, L243, L331 and L332) in homozygosity for all favorable alleles for PSP trait, 43 lines for SSSP, and only one (L2) for WSS.

3.3.4. Candidate Genes

Based on the significant markers, 64 candidate genes were identified: 20 exclusively for PSP, 13 exclusively for SSSP, 26 exclusively for WSS, and 05 simultaneously for PSP and SSSP. Only 01 marker (mk_7) did not have any candidate gene associated within the screened framework.

49 gene ontology (GO) enriched terms were related to the candidate genes. The GO terms that appeared the most were ATP binding (7), protein kinase activity (4), protein phosphorylation (4), integral component of membrane (4), protein binding (4), and oxidation-reduction process (4). Other relevant terms identified were related to metal ion binding, membrane composition, and cellulose synthesis. The functional analysis based on the InterPro database identified 49 different terms related their protein family and domain similarity. Four terms presented repeated in the

analysis: Protein kinase domain (4), NAC domain (3), FAD linked oxidase, N-terminal (2) and COBRA, plant (2).

For PSP trait, the marker mk_6 related the candidate gene Zm00001e030592 (Chr6:112937090..112940915), a cellulose synthase gene, while the marker mk_9 indicated candidate genes related to the cellulose biosynthesis. The mk_6 had the candidate gene Zm00001e030592 (Chr6:112937090..112940915) a cellulose synthase gene and the mk_9 presented two COBRA genes, Zm00001e035504 (chr7:173290229..173293140) and Zm00001e035505 (chr7:173292498..173295913), overlapping in the different DNA strands. The function of COBRA gene is related to extracellular glycosylphosphatidylinositol (GPI)-anchored proteins, responsible for cell expansion through orientating microfibrils and directing cellulose deposition (Roudier et al., 2005). The cellulose synthase has the function of transfer glucose residuals and binding them in a cellulose polymer (Pear et al., 1996).

The marker mk_10 presented the gene Zm00001e026633 (chr8:76973351..76975713), which function is related to anion-transport pores in stomatal guard cells responsible for stomatal closure in response to stress and/or physiological stimulus (Vahisalu et al., 2008). The marker mk_13 referred to a candidate gene called Zm00001e039668 (chr10:22419086..22426364), which encodes the protein dehydration-induced-19 (Di19), a protein expressed in the leaves and roots correlated to the increase of plant sensitivity to environmental stress (Gosti et al., 1995). In addition, marker mk_12 was associated with the gene Zm00001e037636 (chr9:108652381..108653598) or DnaJ, a member of hsp40 (chaperone). Together with DnaK (hsp70), DnaJ is responsible for folding of nascent proteins, translocation of polypeptides across organelle membranes, coordinating responses to stress, and targeting selected proteins for degradation (Malinverni et al., 2017).

One of highlighted candidate genes for SSSP was the mk_9 marker, which presented two COBRA genes, responsible for regulation of cell expansion and cellulose deposition. Interestingly, both mk_2 and mk_8 markers were associated with genes with NAC domain (Zm00001e021721 in chr4:34191672..34195733 and Zm00001e035277 in chr7:167307455..167310486). The NAC proteins are involved in developmental processes, including formation of the shoot apical meristem, floral organs and lateral shoots, as well as in plant hormonal control and defense (Duval et al., 2002). Furthermore, NAC proteins have been implicated in responses to stress and viral infections (Xie et al, 1999; Ren et al, 2000; Collinge & Boller, 2001).

Between the significant markers of WSS trait is the mk_3. It implicated the gene Zm00001e023016 (chr4:147320508..147323259), that resembles a phosphatidylinositol N-acetylglucosaminyltransferase subunit-C (PIG-C). The PIG-C is responsible to add a N-

acetylglucosamine to the phosphatidylinositol (PI) molecule, which is the first step of GPI anchor biosynthesis (Inoue et al., 1996).

The mk_11 was associated to the gene Zm00001e028120 (chr8:156736301..156738932), a UDP-glucosyltransferase responsible for adding a glucosyl radical to small hydrophobic molecules. In plants, it is related to transfer a glucose to a flavonol as one of last steps of anthocyanin biosynthesis (Kleinehollenhorst et al., 1982; Sutter & Grisebach, 1973). Anthocyanin is responsible for the colors red, pink, purple and blue in plants (Grotewold, 2004; Landi et al., 2015), and the abnormal production of red-purple pigments in leaves is a common symptom of corn stunt disease.

The mk_5 was related to the gene Zm00001e030469 (chr6:106994699..107005151), a PX (phox) domain related to phosphatidylinositol (PI) binding. The PX domains are involved in several functions such as cell signaling, vesicular trafficking, protein sorting and lipid modification (Odorizzi et al., 2000; Worby & Dixon, 2002). The protein phosphorylation drives conformational change in protein p47phox allowing its PX domain to bind to PI what activates the phagocyte oxidase as a defense response (Karathanassis et al., 2002).

The mk_4 was associated to the candidate gene Zm00001e024744 (chr4:227211226..227216034), also called seven-in-absentia (SINA) gene. This gene is associate to plant development and responses to stress (Sharma et al., 2014). The SINA and NAC genes belong to the same metabolic pathway, since both are regulated by auxins, and apparently SINA ubiquitinates NAC and directs it to proteasomal degradation (Xie et al., 2002).

3.4. Discussion

3.4.1. Metabolic defense-resistance processes

Observing the functions of the candidate genes, it is possible to pinpoint three main metabolic processes associated to the corn stunt disease complex resistance: (1) NAC/ SINA auxin regulation, (2) cellulose biosynthesis, and (3) plant defense weaponry.

The NAC-SINA interaction is related to the auxin metabolism. Individally, NAC and SINA up and down regulate the hormone expression, respectively (Xie et al., 2002). Some symptoms common to stunt corn disease and maize bushy stunt disease as dwarfism, abnormal prolificity, and reduced number of roots (Nault, 1980) are associated to auxin metabolism since this phytohormone acts in plant growth through enabling cell expansion and promoting production of cell wall, in accordance to the acid-growth hypothesis (Rayle & Cleland, 1992).

In *S. citri*, a correlated pathosystem to the complex, it was observed that the infected plants undergo reduction of auxin and anthocyanin content, and present only a few secondary roots (Chang, 1998) due to lacking of NAC activity or overexpression of the SINA gene (Xie et al., 2002). Since our GWAS analysis indicated significant markers containing both NAC and SINA candidate genes, it is reasonable to infer that the disease complex is analogous to *S. citri*. Thus, these genes are important to maintain normal auxin levels or at least being less affected by the pathogen metabolites. This hypothesis is reinforced when we observe the high frequency of heterozygous for mk_4 (SINA), while the homozygous for less frequent allele is low. Probably, the less frequent allele promotes a weak ubiquitination of NAC gene, and consequently greater auxin concentration. However, in homozygosis, this gene may not be favorable for plant survivor in normal conditions.

Other metabolic pathway affected by auxins and here identified is the cellulose biosynthesis. Both functions overlap because the cell expansion is an auxin dependent process (Rayle & Cleland, 1992). In a mollicute infection, enzymes associated to cellulose synthesis that perform better under low concentration of auxin may produce mild symptoms and be more resistance to the disease complex.

The cellulose associated genes may also reduce the abnormal glucose accumulation in the leaves. This accumulation happens because both pathogenies use preferentially fructose as source of energy, consequently reducing fructose content and inhibiting the activity of invertase enzyme, which leads to the accumulation of glucose in the leaves and the repression of the photosynthetic process (André et al., 2005). Possibly, the resistant plants avoid the glucose accumulation and direct this sugar to other metabolic pathways, for example, for producing cellulose polymers of β-1,4 linked glucose residues. Finally, the conversion of the excess of glucose in dry matter allows the photosynthesis process to keep working.

Some candidate genes are related to signaling, homeostasis and defense against environmental stresses, demonstrating changes in plant machinery to deal with the pathogens. In this case, UDP-glucosyltransferase assist to produce anthocyanin, a compound responsible for reducing reactive oxygen species (ROS)(Landi et al., 2015) and participating in ROS-induced signaling cascades (Hatier & Gould, 2008). The action of DnaJ also helps to protect the plant and keep the homeostasis during the stress condition. On the other hand, the PX receptor can lead to recruitment of the phagocyte vacuole to kill pathogen microorganisms through ROS production (Segal, 2008). The gene Di19 has the function to assist plant to response to drought and regulating some pathogenesis-related (PR) genes, specially, in the scenario where the disease complex reduces the root development.

3.4.2. Plant breeding implications

From a plant breeding perspective, the evaluation of the three traits simultaneously is preferred compared to the marginal use of WSS, since different markers and candidate genes with different sets of molecular functions were identified using both. Furthermore, the quantification of WSS is refined due to the weighted between PSP and SSSP, and the differentiation of the disease symptoms.

The marker effects presented high dispersion in boxplots (Figure 16), especially when the allele is common in the population. Focusing only in the presence of the two most frequent genotypic classes, the allelic substitution effect is significant. On the other hand, considering the genotypic class less frequent (frequency lower than 0.02 for all markers in the population), it is reasonable to infer the existence of dominance (mk_1) and overdominance (mk_5) which may be exploited in breeding programs. It is important to highlight that this aspect must be investigated in-depth in panels with higher frequency of less frequent alleles, for example evaluating hybrid derived from these lines or using diallels.

The marker mk_1 presented the largest effect, with favorable allele as the most common and with MAF close to 0.06. These attributes indicate a potential selection pressure against the homozygous of its rare allele, which is emphasized under the occurrence of the disease complex, leading to the fixation of this marker. The same may not be happen for the mk_6 and mk_4 markers, since the former present the favorable allele as the less frequent and the latter has the heterozygous genotypic class as the most adapted. The use of these two markers in breeding programs, in a hybrid production context, is conditioned to the validation: if the homozygous genotypes for the less frequent allele is deleterious in absence of the disease complex pressure, their use for hybrid production may be compromised. Other hypothesis that may explain the existence of heterozygous markers in the inbred lines is that these set of markers actually correspond to duplicated genes (paralogous), absent in the temperate genome of reference B73 (Brandenburg et al., 2017).

3.5. Conclusion

It was possible to identify genetic variability for the traits PSP, SSSP and WSS concerning the corn stunt disease and bushy maize stunt disease in this tropical maize diversity panel. In total, we found 13 markers associated to the disease complex, and the candidate genes identified support and elucidates the metabolic pathways affected. The genes were associated mainly to auxin metabolism, cellulose biosynthesis and defense-protection against pathogens genes. The markers

identified were well distributed across the germplasm, being possible to identify lines with high content of favorable alleles for each trait. Inbred lines descendent of PF-41X05-33-05B presented high concentration of favorable alleles for the three traits simultaneously. Therefore, they may be used directly to produce hybrids or incorporated in recurrent-selection programs to enhance the favorable alleles in the population. Future researches must validate and verify the occurrence of non-additive effects.

References

André, A., Maucourt, M., Moing, A., Rolin, D., & Renaudin, J. (2005). Sugar Import and Phytopathogenicity of Spiroplasma citri : Glucose and Fructose Play Distinct Roles. 18(1), 33–42. https://doi.org/10.1094/MPMI-18-0033.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., & Rubin, Gerald M. Sherlock, G. (2000). The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J. Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556.Gene

Bergamin Filho, A., Kimati, H., & Amorim, L. (1995). Manual de Fitopatologia: Princípios e Conceitos (A. Bergamin Filho, H. Kimati, & L. Amorim (eds.); 3rd ed.). Agronômica Ceres.

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. Bioinformatics, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brandenburg, J. T., Mary-Huard, T., Rigaill, G., Hearne, S. J., Corti, H., Joets, J., Vitte, C., Charcosset, A., Nicolas, S. D., & Tenaillon, M. I. (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. PLoS Genetics, 13(3), 1–30. https://doi.org/10.1371/journal.pgen.1006666

Browning, B. L., & Browning, S. R. (2008). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics, 84(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R Reference Manual Version 4. 176. http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Chang, C. J. (1998). Pathogenicity of aster yellows phytoplasma and Spiroplasma citri on periwinkle. Phytopathology, 88(12), 1347–1350. https://doi.org/10.1094/PHYTO.1998.88.12.1347

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Oliveira, Charles Martins, De Oliveira, E., Canuto, M., & Cruz, I. (2007). Controle químico da cigarrinha-do-milho e incidência dos enfezamentos causados por molicutes. Pesquisa Agropecuaria Brasileira, 42(3), 297–303. https://doi.org/10.1590/s0100-204x2007000300001

de Oliveira, Elizabeth, Gomes e Gama, E. E., Teixeira, F. F., de Oliveira, A. C., & da Silva, A. R. (2013). Genetic control of maize resistance to corn stunt spiroplasma. Phytopathogenic Mollicutes, 3(2), 67. https://doi.org/10.5958/j.2249-4677.3.2.016

De Oliveira, Elizabeth, Magalhães, P. C., Gomide, R. L., Vasconcelos, C. A., Souza, I. R. P., Oliveira, C. M., Cruz, I., & Schaffert, R. E. (2002). Growth and nutrition of mollicute-infected maize. Plant Disease, 86(9), 945–949. https://doi.org/10.1094/PDIS.2002.86.9.945

Duval, M., Hsieh, T. F., Kim, S. Y., & Thomas, T. L. (2002). Molecular characterization of AtNAM: A member of the Arabidopsis NAC domain superfamily. Plant Molecular Biology, 50(2), 237–248. https://doi.org/10.1023/A:1016028530943

Gonzalez, J. G., Jaramillo, M. G., & Lopes, J. R. S. (2018). Undetected Infection by *Maize Bushy Stunt Phytoplasma* Enhances Host-Plant Preference to *Dalbulus maidis* (Hemiptera: Cicadellidae). Environmental Entomology, 47(2), 396–402. https://doi.org/10.1093/ee/nvy001

Gordon, D. T., Knobe, J. K., & Scott, G. E. (1981). Introduction: history, geographical distribution, pathogen characteristics, and economic importance. In D. T. Gordon (Ed.), Virus and viruslike diseases of maize in the United States (Southern C, pp. 1–12). Ohio Agric. Res. and Dev. Center.

Gosti, F., Bertauche, N., Vartanian, N., & Giraudat, J. (1995). Abscisic acid-dependent and -independent regulation of gene expression by progressive drought in Arabidopsis thaliana. Mgg Molecular & General Genetics, 246(1), 10–18. https://doi.org/10.1007/BF00290128

Granato, I. S. C., Galli, G., de Oliveira Couto, E. G., e Souza, M. B., Mendonça, L. F., & Fritsche-Neto, R. (2018). snpReady: a tool to assist breeders in genomic analysis. Molecular Breeding, 38(102), 1–7. https://doi.org/10.1007/s11032-018-0844-8

Grotewold, E. (2004). The challenges of moving chemicals within and out of cells: Insights into the transport of plant natural products. Planta, 219(5), 906–909. https://doi.org/10.1007/s00425-004-1336-0

GUSSIE, J. S.; FLECHER, J.; CLAYPOOL, P. L. (1995). Movement and multiplication of *Spiroplasma kunkelii* in corn. Ecology and Epidemiology, 85(10), 1093–1098.

Hatier, J.-H. B., & Gould, K. S. (2008). Foliar anthocyanins as modulators of stress signals. Journal of Theoretical Biology, 253, 625–627.

Inglis, P. W., Marilia de Castro, R. P., Resende, L. V., & Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLoS ONE, 13(10), 1–14. https://doi.org/10.1371/journal.pone.0206085

Inoue, N., Watanabe, R., Takeda, J., & Kinoshita, T. (1996). PIG-C, one of the three human genes involved in the first step of glycosylphosphatidylinositol biosynthesis is a homologue of Saccharomyces cerevisiae GPI2. Biochemical and Biophysical Research Communications, 226(1), 193–199. https://doi.org/10.1006/bbrc.1996.1332

Junqueira, A., Bedendo, I., & Pascholati, S. (2004). Biochemical changes in corn plants infected by the *maize bushy stunt phytoplasma*. Physiological and Molecular Plant Pathology, 65(4), 181–185. https://doi.org/10.1016/j.pmpp.2005.01.005

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Karathanassis, D., Stahelin, R. V., Bravo, J., Perisic, O., Pacold, C. M., Cho, W., & Williams, R. L. (2002). Binding of the PX domain of p47phox to phosphatidylinositol 3, 4-bisphosphate and phosphatidic acid is masked by an intramolecular interaction. EMBO Journal, 21(19), 5057–5068. https://doi.org/10.1093/emboj/cdf519

Kimati, H., Amorim, L., Rezende, J. A. M., Bergamin Filho, A., & Camargo, L. E. A. (2005). Manual de Fitopatologia, Vol II. In Doenças das plantas cultivadas.

Kleinehollenhorst, G., Behrens, H., Pegels, G., Srunk, N., & Wiermann, R. (1982). Formation of Flavonol 3-O-Diglycosides and Flavonol 3-O-Triglycosides by Enzyme Extracts from Anthers of Tulipa cv. Apeldoorn: Characterization and Activity of Three Different O-Glycosyltransferases during Anther Development. Zeitschrift Fur Naturforschung - Section C Journal of Biosciences, 37(7–8), 587–599. https://doi.org/10.1515/znc-1982-7-808

Landi, M., Tattini, M., & Gould, K. S. (2015). Multiple functional roles of anthocyanins in plant-environment interactions. Environmental and Experimental Botany, 119, 4–17. https://doi.org/10.1016/j.envexpbot.2015.05.012

Langmead, B., & Salzberg, S. (2013). Bowtie2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923.Fast

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome- Wide Association Studies. Mlm, 1–24. https://doi.org/10.1186/1471-2156-13-100

Malinverni, D., Lopez, A. J., De Los Rios, P., Hummer, G., & Barducci, A. (2017). Modeling Hsp70/Hsp40 interaction by multi-scale molecular simulations and coevolutionary sequence analysis. ELife, 6, 1–20. https://doi.org/10.7554/eLife.23471

MÁRQUEZ SÁNCHEZ, F. (1982). The genetic improvement of resistance to the maize disease stunt and downy mildew in Nicaragua. Revista Chapingo, México, 7(35/36), 26–27.

MASSOLA JÚNIOR, N. S.; BEDENDO, I. P. ., & AMORIM, L.; LOPES, J. R. S. (1999). Quantificação de danos causados pelo enfezamento vermelho e enfezamento pálido do milho em condições de campo. Fitopatologia Brasileira, 24(2), 136–142.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Research, 47(D1), D351–D360. https://doi.org/10.1093/nar/gky1100

Moya-Raygoza, G., & Nault, L. R. (1998). Transmission biology of *maize bushy stunt phytoplasma* by the corn leafhopper (Homoptera: Cicadellidae). Annals of the Entomological Society of America, 91(5), 668–676. https://doi.org/10.1093/aesa/91.5.668

Nault, L. R. (1980). Maize Bushy Stunt and Corn Stunt: A Comparison of Disease Symptoms, Pathogen Host Ranges, and Vectors. Phytopathology, 70(7), 659. https://doi.org/10.1094/Phyto-70-659

Odorizzi, G., Babst, M., & Emr, S. D. (2000). Phosphoinositide signaling and the regulation of membrane trafficking in yeast. Trends in Biochemical Sciences, 25(5), 229–235. https://doi.org/10.1016/S0968-0004(00)01543-7

Oliveira, Chales Martins de, Querino, R. B., & Frizzas, M. R. (2017). Leafhoppers and planthoppers in maize crop in Brazil. In Charles Martins de Oliveira & E. de O. Sabato (Eds.), Diseases in Maize: Insect vectors, mollicutes and viruses (1st ed., pp. 71–96). EMBRAPA.

Oliveira, Elizabeth De, Fernandes, F., Souza, I., Oliveira, C. De, & Cruz, I. (2003). Enfezamentos, Viroses e Insetos Vetores em Milho - Identificação e Controle. Circular Técnica Embrapa, 26, 1–10.

Oliveira, E., Oliveira, C. M., Souza, I. R. P., Magalhães, P. C., & Cruz, I. (2002). Enfezamentos em Milho: Expressão de Sintomas Foliares, Detecção dos Molicutes e Interações com Genótipos. Revista Brasileira de Milho e Sorgo, 1(1), 53–62. https://doi.org/10.18512/1980-6477/rbms.v1n1p53-62

Özbek, E., Miller, S. A., Meulia, T., & Hogenhout, S. A. (2003). Infection and replication sites of *Spiroplasma kunkelii* (Class: Mollicutes) in midgut and Malpighian tubules of the leafhopper *Dalbulus maidis*. Journal of Invertebrate Pathology, 82(3), 167–175. https://doi.org/10.1016/S0022-2011(03)00031-4

Pear, J. R., Kawagoe, Y., Schreckengost, W. E., Delmer, D. P., & Stalker, D. M. (1996). Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proceedings of the National Academy of Sciences of the United States of America, 93(22), 12637–12642. https://doi.org/10.1073/pnas.93.22.12637

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE, 7(2). https://doi.org/10.1371/journal.pone.0032253

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Walsh, J. R., Sen, T. Z., Cho, K. T., Schott, D. A., Braun, B. L., Dietze, M., Dunfee, B., Elsik, C. G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., … Andorf, C. M. (2019). Maizegdb 2018: The maize multi-genome genetics and genomics database. Nucleic Acids Research, 47(D1), D1146–D1154. https://doi.org/10.1093/nar/gky1046

Querino, R. B., Meneses, A. R., Albarracin, E. L., Oliveira, C. M. de, & Triapitsyn, S. (2017). Biological control of *Dalbulus maidis* in Brazil: An overview of the parasitoids. In Charles Martins de Oliveira & E. de O. Sabato (Eds.), Diseases in Maize: Insect vectors, mollicutes and viruses (1st ed., pp. 119–138). EMBRAPA.

Rayle, D. L., & Cleland, R. E. (1992). The acid growth theory of auxin-induced cell elongation is alive and well. Plant Physiology, 99(4), 1271–1274. https://doi.org/10.1104/pp.99.4.1271

Roudier, F., Fernandez, A. G., Fujita, M., Himmelspach, R., Borner, G. H. H., Schindelman, G., Song, S., Baskin, T. I., Dupree, P., Wasteneys, G. O., & Benfey, P. N. (2005). COBRA, an Arabidopsis extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. Plant Cell, 17(6), 1749–1763. https://doi.org/10.1105/tpc.105.031732

Segal, A. W. (2008). The function of the NADPH oxidase of phagocytes and its relationship to other NOXs in plants , invertebrates , and mammals. 40, 604–618. https://doi.org/10.1016/j.biocel.2007.10.003

Sharma, M., Pandey, A., & Pandey, G. K. (2014). β-catenin in plants and animals: Common players but different pathways. Frontiers in Plant Science, 5(APR), 1–5. https://doi.org/10.3389/fpls.2014.00143

Silva, R. G., Galvão, J. C. C., Miranda, G. V., & De Oliveira, E. (2003). Controle genético da resistência aos enfezamentos do milho. Pesquisa Agropecuaria Brasileira, 38(8), 921–928. https://doi.org/10.1590/s0100-204x2003000800004

Sutter, A., & Grisebach, H. (1973). UDP-glucose: Flavonol from cell suspension cultures of parsley. Biochimica et Biophysica Acta (BBA) - Enzymology, 309(2), 289–295. https://doi.org/10.1016/0005-2744(73)90027-2

Vahisalu, T., Kollist, H., Wang, Y. F., Nishimura, N., Chan, W. Y., Valerio, G., Lamminmäki, A., Brosché, M., Moldau, H., Desikan, R., Schroeder, J. I., & Kangasjärvi, J. (2008). SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. Nature, 452(7186), 487–491. https://doi.org/10.1038/nature06608

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

Wimmer, V., Albrecht, T., Auinger, H.-J., & Schön, C.-C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics, 28(15), 2086–2087. https://doi.org/10.1093/bioinformatics/bts335

Worby, C. A., & Dixon, J. E. (2002). Sorting out the cellular functions of sorting nexins. Nature Reviews Molecular Cell Biology, 3(12), 919–931. https://doi.org/10.1038/nrm974

Xie, Q., Guo, H. S., Dallman, G., Fang, S., Weissman, A. M., & Chua, N. H. (2002). SINAT5 promotes ubiquitin-related degradation of NAC1 to attenuate auxin signals. Nature, 419(6903), 167–170. https://doi.org/10.1038/nature00998

Figures

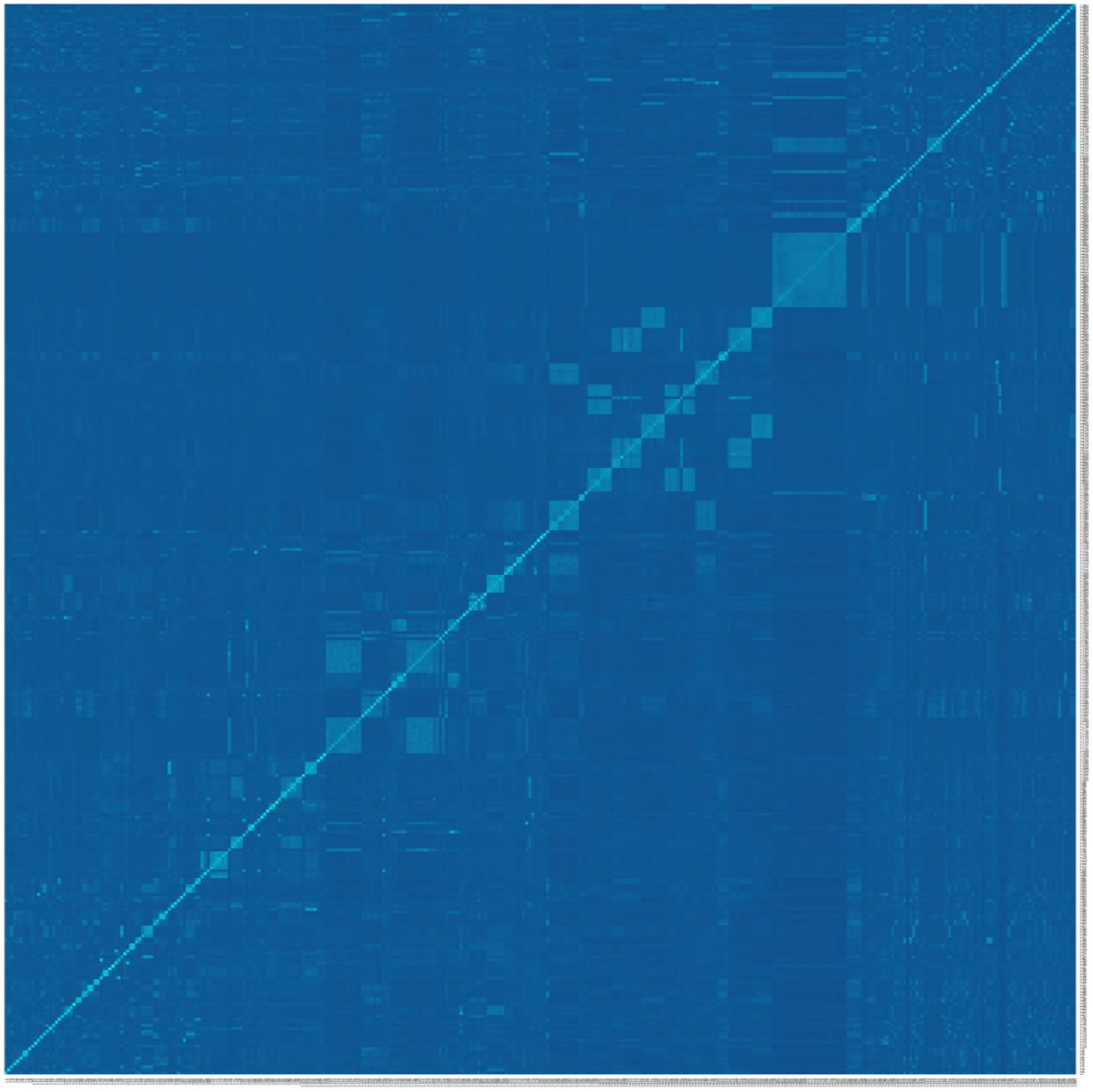

Figure 13. VanRaden additive genomic relationship matrix of 360 tropical maize inbred lines. Darker blue pixels indicates low endogamy (minimum of -0.127), and light cyan pixels indicate values high endogamy (maximum of 1.066). Lines are ordered from left to right and upwards from L1 to L360.

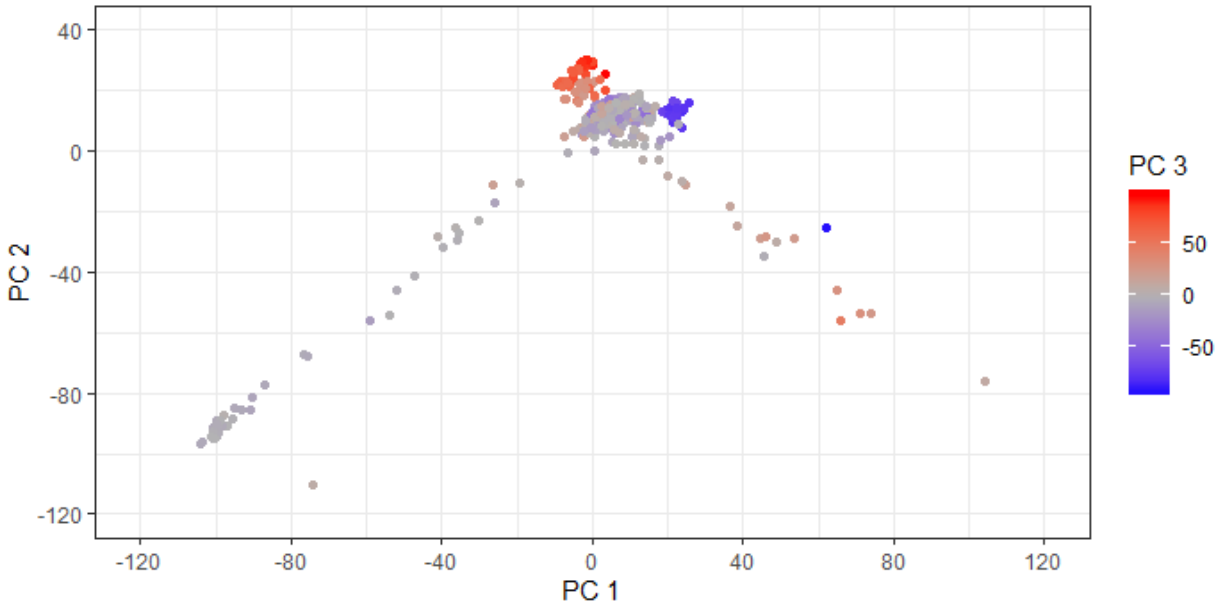

Figure 14. Distribution of three main principal components (PC) for a tropical maize diversity panel. The lines L145 (PC 1 = 449.4, PC 2 = -433.1, PC 3 = 43.1) and L93 (PC 1 = 200.1 , PC 2 = -191.0, PC 3 = 83.2) are not present in the plot.

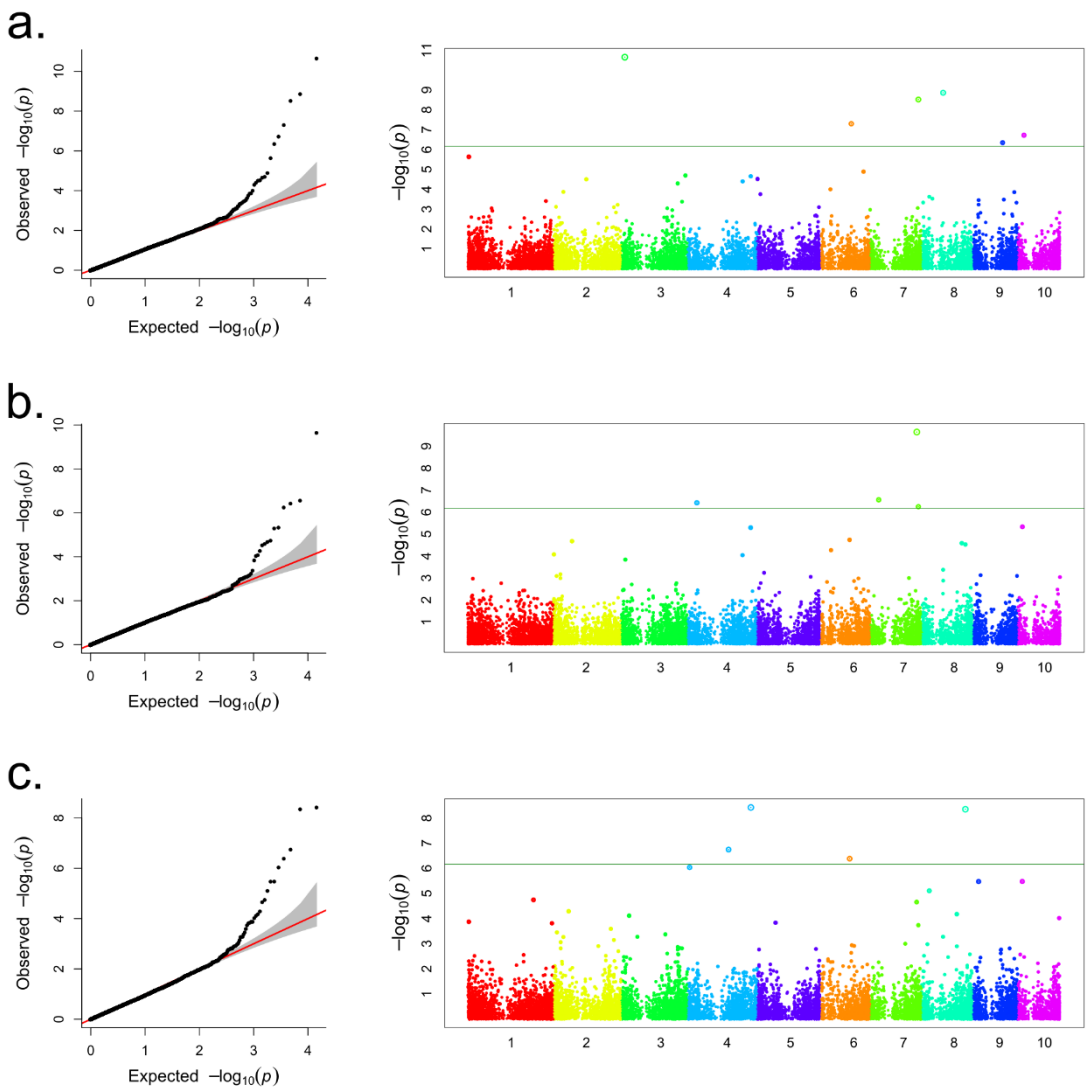

Figure 15. QQ plot (left) and Manhattan plots (right) for the traits (a) proportion of survivor plants (PSP), (b) sanity score in survivor plants (SSSP), and (c) whole sanity score (WSS) in a genome wide association study of a tropical maize diversity panel with 360 lines.

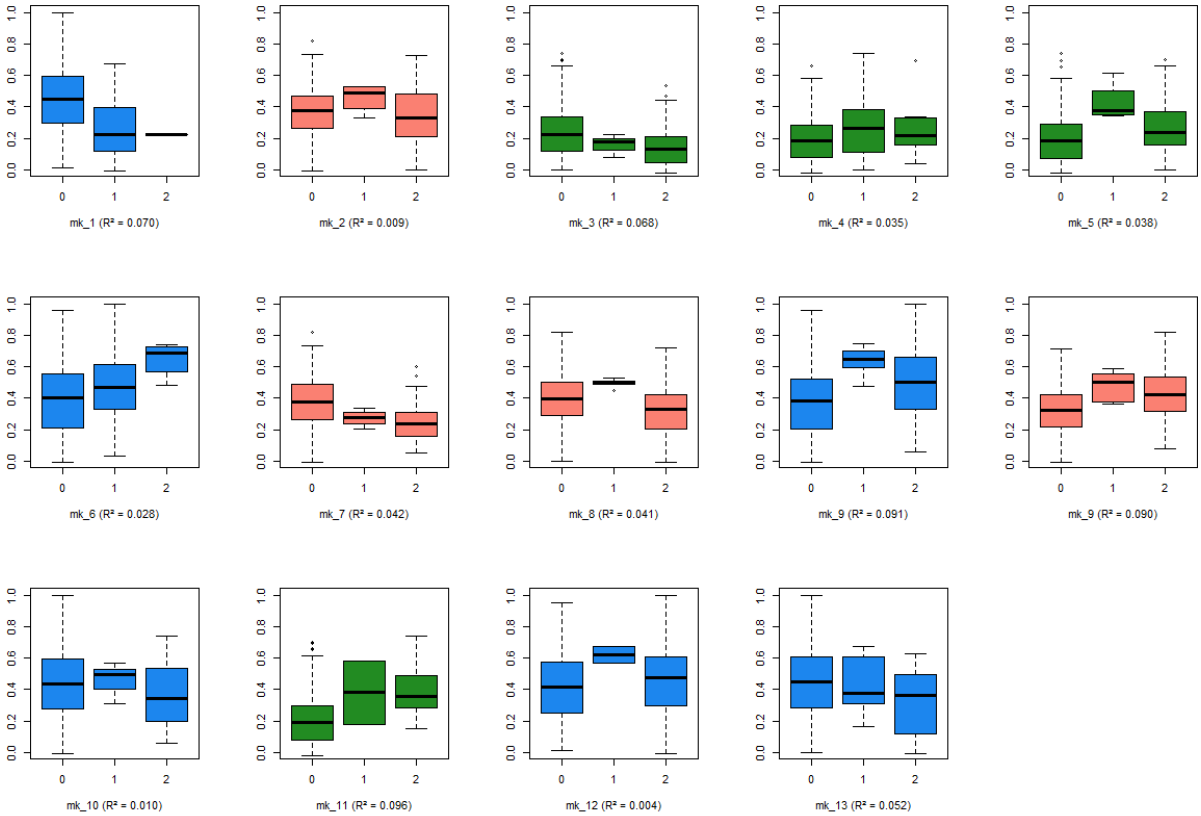

Figure 16. Boxplot BLUE for the traits in according of allelic dosage for the significant markers in GWAS analysis of a tropical maize diversity panel. PSP associated marker is in blue, SSSP in orange, and WSS in green. R^2 resembles for the model adjustment.

Tables

Table 2. Significant markers in genome wide association study for traits proportion of survivor plants (PSP), sanity score in survivor plants (SSSP) and whole survivor score (WSS) using a tropical maize inbred line population. MAF, P, H, Q and R^2 resembles for minor allele frequency, number of individuals homozygous for the common allele, heterozygous and homozygous for the less common allele, and R^2-adjust, respectively.

Marker	Chr.	Position	Trait	Effect	MAF	P	H	Q	P-value	R^2
mk_1	3	11,874,476	PSP	-0.163	0.06	318	41	1	$2.18\ 10^{-11}$	0.07
mk_2	4	34,191,165	SSSP	-0.029	0.39	218	4	138	$3.67\ 10^{-07}$	0.01
mk_3	4	147,394,225	WSS	-0.034	0.25	268	3	89	$1.75\ 10^{-07}$	0.07
mk_4	4	227,309,550	WSS	0.056	0.17	246	105	9	$3.71\ 10^{-09}$	0.04
mk_5	6	107,011,465	WSS	0.029	0.23	274	4	82	$4.03\ 10^{-07}$	0.04
mk_6	6	112,881,661	PSP	0.073	0.21	215	141	4	$4.88\ 10^{-08}$	0.03
mk_7	7	30,906,940	SSSP	-0.055	0.08	331	3	26	$2.69\ 10^{-07}$	0.04
mk_8	7	167,289,762	SSSP	-0.039	0.50	177	5	178	$2.20\ 10^{-10}$	0.04
mk_9	7	173,291,600	PSP	0.047	0.39	216	5	139	$2.95\ 10^{-09}$	0.09
			SSSP	0.032					$5.51\ 10^{-07}$	0.09
mk_10	8	76,877,319	PSP	-0.078	0.08	329	3	28	$1.35\ 10^{-09}$	0.01
mk_11	8	156,735,397	WSS	0.057	0.072	318	41	1	$4.36\ 10^{-09}$	0.10
mk_12	9	108,640,116	PSP	0.046	0.231	218	4	138	$4.38\ 10^{-07}$	0.01
mk_13	10	22,375,201	PSP	-0.056	0.167	268	3	89	$1.86\ 10^{-07}$	0.05

APPENDIX

APPENDIX A. Serologic test to attest the presence and identification of corn stunt disease complex pathogens in plant samples of the present experiments (In Portuguese).

CLÍNICA FITOPATOLÓGICA PROF. HIROSHI KIMATI
Departamento de Fitopatologia e Nematologia
Av. Pádua Dias, 11 Caixa Postal 09 CEP: 13.418-900 Piracicaba - SP
Fone: (19) 3429-4124 opção 3. Fax: (19) 3447-8647 Email: clinica.esalq@usp.br

Piracicaba, 02 de maio de 2019.

Ilmo. Sr.
Roberto Fritsche Neto

Prezado Senhor,

Com referência à sua consulta sobre doença na cultura de milho, proveniente de Piracicaba-SP, se tem a informar que através observações sintomatológicas e análises moleculares, foram constatadas nas amostras enviadas para análise:

Amostra 01 – Planta 01:
A presença de *Spiroplasma kunkelii*.

Amostra 02 – Planta 02:
A presença de *Candidatus Phytoplasma asteris* e *Spiroplasma kunkelii*.

Amostra 03 – Planta 03:
A presença de *Candidatus Phytoplasma asteris* e *Spiroplasma kunkelii*.

Candidatus Phytoplasma asteris ocasiona a doença Enfezamento Vermelho, também conhecida por "maize bushy stunt".

Spiroplasma kunkelii causa a doença Enfezamento Pálido, também conhecida por "corn stunt" ou Enfezamento Amarelo.

Essas doenças ocorrem com maior frequência e severidade principalmente no segundo plantio, a partir de janeiro. Os agentes causais são transmitidos e disseminados na lavoura por insetos (cigarrinhas – *Dalbulus maidis*).

Para o controle devem ser adotadas várias medidas preventivas conjuntamente, tais como: o plantio de cultivares resistentes; evitar plantios tardios (safrinha), época em que ocorre a maior população do inseto vetor; a eliminação das plantas afetadas e de restos da cultura anterior; a capina periódica da área, eliminando as plantas daninhas. O controle da cigarrinha pode proporcionar certo controle da doença, assim, é indicada a pulverização da cultura no início de seu desenvolvimento com inseticidas registrados.

Sem mais para o momento, colocamo-nos à disposição para eventuais esclarecimentos.

Atenciosamente

Enga. Agrônoma Liliane De Diana Teixeira
Doutora em Fitopatologia

OBS: O resultado da análise apenas representa a amostra entregue na Clínica Fitopatológica Prof. Hiroshi Kimati. A Clínica Fitopatológica não se responsabiliza pela coleta da amostra. Infestações muito baixas podem não ser detectadas.

APPENDIX B. Field images of diseased plants with corn stunt disease and maize bushy stunt disease.