



**LANDER SANTOS DE OLIVEIRA**

**AGRUPAMENTO HETERÓTICO DE MILHO USANDO  
MARCADORES SNPS**

**LAVRAS-MG**

**2020**

**LANDER SANTOS DE OLIVEIRA**

**AGRUPAMENTO HETERÓTICO DE MILHO USANDO MARCADORES SNPS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação do Mestrado Profissional em Genética e Melhoramento de Plantas, área de concentração em Genética Molecular e de Fitopatógenos, para a obtenção do título de Mestre.

Prof. Dr. Welison Andrade Pereira  
Orientador  
Prof. Dr. Ivan Schuster  
Coorientador

**LAVRAS-MG  
2020**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Oliveira, Lander Santos de.  
Agrupamento heterótico de milho usando marcadores SNPs /  
Lander Santos de Oliveira. - 2020.  
40 p. : il.

Orientador(a): Welison Pereira Andrade.  
Coorientador(a): Ivan Schuster.  
Dissertação (mestrado profissional) - Universidade Federal de  
Lavras, 2020.  
Bibliografia.

1. Diversidade genética. 2. Polimorfismos de nucleotídeo  
único. 3. Zea mays. I. Andrade, Welison Pereira. II. Schuster, Ivan.  
III. Título.

**LANDER SANTOS DE OLIVEIRA**

**AGRUPAMENTO HETERÓTICO DE MILHO USANDO MARCADORES SNPS**

**HETEROTIC GROUPING OF MAIZE USING SNP MARKERS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação do Mestrado Profissional em Genética e Melhoramento de Plantas, área de concentração em Genética Molecular e de Fitopatógenos, para a obtenção do título de Mestre.

APROVADA em 31 de julho de 2020.

Prof. Dr. Welison Andrade Pereira (UFLA)

Prof. Dr. Carlos Eduardo Pulcinelli (UFLA)

Prof Dr. Ivan Schuster (LONGPING HIGH-TECH)

Prof. Dr. Welison Andrade Pereira  
Orientador

Prof. Dr. Ivan Schuster  
Coorientador

**LAVRAS-MG**  
**2020**

## AGRADECIMENTOS

À LongPing High-Tech, empresa que apoiou financeiramente a realização deste trabalho.

À Universidade Federal de Lavras, que proporcionou a continuidade dos meus estudos de forma gratuita através do curso de pós-graduação na modalidade de Mestrado Profissional.

Ao Prof. Dr. Ivan Schuster, que acreditou no meu potencial, guiou meus passos compartilhando seus conhecimentos com maestria e pelo seu exemplo inspirador de ética e profissionalismo.

Ao Prof. Dr. Welison Pereira, pela paciência na orientação deste trabalho e brilhante coordenação do curso de Mestrado Profissional em Genética e Melhoramento de Plantas.

Ao Prof. Dr. Evandro Novaes, que dedicou seu tempo ensinando diversas análises com eficiência e educação admirável.

Aos meus pais, que me apoiaram durante toda esta jornada.

Aos colegas da turma de mestrado 2018/1, por tornarem os momentos de tensão em alegria, em especial ao Jorcélio Cabral, pelas inúmeras horas de dedicação conjunta aos estudos virtuais.

Aos professores do curso, pela dedicação e qualidade das aulas ministradas.

A todos os membros do departamento de Pesquisa e Desenvolvimento da LongPing High-Tech, que instigaram o desejo de buscar mais conhecimento pela área.

Ao membro da banca examinadora, Prof. Dr. Carlos Eduardo Pulcinelli, que aceitou gentilmente o convite para avaliar este trabalho.

E a todos os outros que participaram direta ou indiretamente desta conquista, o meu muito obrigado!

## RESUMO

Os avanços das tecnologias de genotipagem têm transformado a forma com a qual os programas de melhoramento gerenciam seus recursos genéticos. A identificação de polimorfismos de nucleotídeo único (SNPs) pode beneficiar o entendimento da diversidade genética entre linhagens de milho (*Zea mays*) e sua subsequente classificação em grupos heteróticos, contribuindo na orientação de cruzamentos para obtenção de híbridos com elevado desempenho. Neste contexto, o objetivo deste trabalho consistiu em agrupar linhagens de milho em grupos, tendo como base os dados oriundos de genotipagem por SNPs. Assim, neste estudo, a diversidade genética de 293 genótipos foi investigada com 5252 SNPs. A partir dos resultados, verificou-se que a distribuição média dos SNPs foi de 525 por cromossomo. O conteúdo de informação polimórfica (PIC) apresentou média de 0,297. Com base na matriz de distâncias genéticas, os indivíduos foram agrupados conforme o método hierárquico das médias aritméticas não ponderadas (UPGMA) e a análise dos componentes principais (PCA), revelando grupos similares, com alto índice de correlação cofenética (0.953 and 0.863, respectivamente). Os resultados demonstraram consistência quando comparados com grupos heteróticos previamente estabelecidos com informações de genealogia e *topcrosses*. Além disso, subgrupos foram revelados dentro de cada grupo, permitindo uma exploração mais aprofundada da variabilidade genética dentre os materiais.

**Palavras-chave:** Diversidade genética. Polimorfismos de nucleotídeo único. *Zea mays*.

## ABSTRACT

The advances in genotyping technologies have transformed the way breeding programs manage their genetic resources. The identification of single nucleotide polymorphisms (SNPs) can improve the understanding of genetic diversity among maize (*Zea mays*) inbred lines and their classification into heterotic groups, which is useful to guide certain crosses to generate hybrids with higher performance. The objective of this study was to classify maize inbred lines into groups with genotyping data. Genetic diversity of 293 genotypes was investigated with 5252 SNPs. The results showed that the SNPs distribution averaged 525 per chromosome. Polymorphism information content (PIC) averaged 0.297. The unweighted pair group method with arithmetic mean analysis (UPGMA) and principal component analysis (PCA) based on a genetic distance matrix revealed similar clusters with high cophenetic correlation coefficients (0.953 and 0.863, respectively). The results demonstrated consistency compared to previously established heterotic groups using pedigree and breeding information, and also allowed further exploration of genetic variability by revealing subgroups within each group.

**Keywords:** Genetic diversity. Single nucleotide polymorphism. *Zea mays*.

## SUMÁRIO

	<b>PRIMEIRA PARTE</b> .....	8
<b>1</b>	<b>APRESENTAÇÃO</b> .....	8
	<b>REFERÊNCIAS</b> .....	12
	<b>SEGUNDA PARTE – ARTIGO</b> .....	14
	<b>ARTIGO - Heterotic grouping of maize inbred lines based on SNP genotyping data</b> .....	12
<b>1</b>	<b>Introduction</b> .....	16
<b>2</b>	<b>Materials and Methods</b> .....	18
<b>3</b>	<b>Results and Discussion</b> .....	20
<b>4</b>	<b>Conclusion</b> .....	26
<b>5</b>	<b>Acknowledgments</b> .....	27
<b>6</b>	<b>References</b> .....	28
	<b>ANEXOS</b> .....	31



## PRIMEIRA PARTE

### 1 APRESENTAÇÃO

O milho (*Zea mays*) é um dos cereais mais cultivados no mundo, superando a marca de 1 bilhão de toneladas no últimos anos. O Brasil ocupa o terceiro lugar entre os países com maior produção do grão, atrás somente dos EUA e da China (FAOSTAT, 2020). Segundo dados da Companhia Nacional de Abastecimento (CONAB, 2020), a safra brasileira de 2018/19 atingiu a produção de 100 milhões de toneladas. O milho produzido no Brasil tem como principais destinos a exportação, produção de biocombustível e consumo animal.

Para alcançar estes números, o cereal é produzido em todas as regiões do país e em duas épocas do ano. A necessidade de se incrementar a produção para atender à demanda crescente por alimentos e energia exige renovação dos produtos comercializados, de forma que estes apresentem elevada produtividade, com adaptabilidade a diferentes condições ambientais e estresses bióticos e abióticos (CONTINI et al., 2019).

Com isso, o desafio dos programas de melhoramento genético de milho são enormes. Felizmente, as oportunidades de aumento da produtividade também são grandes, pois o milho apresenta grande diversidade genética a ser explorada. Portanto, é essencial que os programas investiguem a diversidade de seus recursos genéticos, estando aptos para mudanças nos objetivos do melhoramento (SMITH; GARDNER; COSTICH, 2017).

Desde o início do século XX, linhagens endogâmicas de milho adaptadas a diversas condições ambientais são desenvolvidas por programas de melhoramento públicos e privados, as quais são utilizadas como parentais na produção de híbridos (SMITH; GARDNER; COSTICH, 2017). Com o crescimento dramático no número de linhagens produzidas por estes programas, as avaliações fenotípicas de desempenho de todas as combinações híbridas possíveis tornaram-se inviáveis. Isso fez com que se tornasse comum classificar as linhagens em grupos heteróticos (ANDORF et al., 2019).

O conceito de grupo heterótico baseia-se no princípio da heterose, isto é, manifestação do vigor híbrido resultante do cruzamento entre dois indivíduos geneticamente divergentes (SHULL, 1948). É possível identificar genitores adequados à obtenção de híbridos com maior efeito heterótico, avaliando-se a diversidade genética entre indivíduos de uma população (CRUZ; FERREIRA; PESSONI, 2011).

Diversidade genética refere-se às variações nas frequências gênicas de uma população ou espécie. Seu estudo permite entender a evolução e adaptabilidade de populações em diferentes condições ambientais (CRUZ; FERREIRA; PESSONI, 2011). A redução da diversidade pode tornar uma população vulnerável a situações de estresses, como variações climáticas e surgimento de novas pragas e doenças. Ela pode ser analisada com informações fenotípicas, bioquímicas ou moleculares (SCHERLOSKY et al., 2018; SARAVANAN et al., 2019).

A avaliação de diversidade genética para classificar linhagens de milho em grupos heteróticos pode ser feita através de métodos convencionais, como dialelos e *topcrosses*. No entanto, estas metodologias demandam tempo, grandes áreas de plantio e trabalho manual de polinização ou despendoamento para obtenção de todos cruzamentos híbridos desejados. Assim, o número de populações a serem avaliadas fica restringido (FERNANDES et al., 2015; LENG et al., 2019). Da mesma forma, a classificação em grupos por meio de informações de parentesco, apesar de simples, exige conhecimento profundo da genealogia dos genótipos, o que nem sempre está disponível (LENG et al., 2019). Portanto, o uso de informações moleculares constituem a maneira mais adequada para quantificação da diversidade genética, uma vez que não são influenciados pelo ambiente (SCHERLOSKY et al., 2018; SARAVANAN et al., 2019).

Com o advento do melhoramento assistido por marcadores moleculares na década de 90 e, mais recentemente, com as tecnologias de genotipagem em larga escala, os estudos moleculares de diversidade genética de milho possibilitaram uma melhor classificação das linhagens, permitindo que melhoristas concentrassem seus esforços em um menor número de combinações híbridas no campo (ANDORF et al., 2019), o que aumentou a eficiência dos programas.

Uma maneira de expressar o grau de diversidade genética entre genótipos com informações moleculares é através de distâncias genéticas entre pares de genótipos. Quanto maior for a distância entre os indivíduos, isto é, quanto mais alelos distintos entre eles, maior a probabilidade de que o cruzamento entre eles resultem em um híbrido com elevada heterose. Portanto, é interessante dispor de uma grande quantidade de locos polimórficos para inferir na diversidade de uma população (CRUZ; FERREIRA; PESSONI, 2011; SARAVANAN et al., 2019).

O conteúdo de informação de polimorfismo (PIC) descrito por Botstein et al. (1980) é um bom indicador do grau de polimorfismo de marcadores. Seu valor fornece uma estimativa do poder discriminatório dos marcadores, pois considera o número de alelos por loco e a

frequência relativa desses alelos (OLIVEIRA, 2015). Assim, o índice permite verificar a diversidade genética geral de uma população através da distribuição dos marcadores informativos (ELTAHER et al., 2018).

Os marcadores SNPs (polimorfismos de nucleotídeo único) são ótimos marcadores moleculares para esta finalidade, uma vez que são as fontes de variação com maior densidade nos genomas. Isso aumenta a probabilidade de que pelo menos um SNP seja polimórfico nos genes de interesse. Além disso, sua aplicação tem oferecido custo reduzido e possibilidade de automação, o que se aplica apropriadamente às necessidades dos programas de melhoramento (RASHEED et al., 2017; GUO et al., 2019).

Com as informações de distância genética, métodos multivariados podem ser aplicados para examinar as relações genéticas entre os indivíduos, ordenando-os entre os mais similares e dissimilares geneticamente, como em análises de agrupamento e análises de dispersão gráfica (MOHHAMADI; PRASANA, 2003). O método de agrupamento hierárquico baseado nas médias das distâncias (UPGMA) é a técnica de agrupamento mais utilizada para análise de diversidade em plantas. Ela fundamenta-se na identificação de dois indivíduos similares e agrupamento deles em um dendrograma. Após o primeiro agrupamento, as distâncias são recalculadas usando a média das distâncias do grupo formado. Isto repete-se gradualmente até que o dendrograma seja formado (MOHHAMADI; PRASANA, 2003). Os ramos do dendrograma identificam as distâncias entre os indivíduos conectados (BERTAN et al., 2006). Já nos métodos de dispersão gráfica, como a análise de componentes principais (PCA), os genótipos são dispersos em um gráfico bidimensional ou tridimensional, em que as distâncias geométricas entre os indivíduos refletem a distância genética entre eles (SEMAGN et al., 2012). Trata-se de um método de ordenação interessante, pois permite detectar visualmente os indivíduos presentes entre dois grupos (MOHAMMADI; PRASANA, 2003).

O presente trabalho exemplifica algumas maneiras de se investigar a diversidade genética entre as linhagens de um programa de melhoramento genético de milho com o objetivo de classificá-las mais adequadamente. Além disso, o estudo demonstra que o uso de marcadores para agrupamento de linhagens é um processo muito mais rápido que o convencional. No esquema convencional, todas as linhagens precisariam ser cruzadas entre si em um esquema de dialelo e as sementes híbridas obtidas, plantadas. Somente após o atingimento da maturidade completa, haveria a avaliação da capacidade de combinação das linhagens que as formaram com base nos dados de produtividade. Trata-se de um processo

demorado, que demanda grandes áreas de plantio e trabalho manual. Além disso, os genótipos avaliados são fortemente influenciados pelo ambiente.

Utilizando marcadores SNP, os quais se encontram amplamente distribuídos por todo o genoma, é possível avaliar a diversidade genética através do DNA. O DNA pode ser extraído de tecidos das plantas ainda em seus estágios iniciais de desenvolvimento. Com isso, diversas linhagens podem ser genotipadas em um curto período de tempo e as distâncias genéticas estimadas entre elas. Assim, a diversidade entre os genótipos é refletida com precisão, visto que os dados genotípicos não são influenciados por interações genótipo-ambiente.

Aplicando métodos de agrupamento aos dados de distância genética, as linhagens podem ser classificadas em grupos, os quais podem auxiliar melhoristas na disposição adequada de quais genótipos deverão ser cruzados com os testadores utilizados em um programa de melhoramento para avaliação de desempenho. Assim, cruzamentos entre materiais do mesmo grupo, isto é, que não expressam heterose, podem ser evitados, reduzindo o número de cruzamentos desnecessários no campo.

Este estudo também abre portas para estudos futuros de orientação de cruzamentos híbridos e predição de híbridos, visto que os dados genotípicos estão disponíveis para avaliar a complementariedade entre os pares de indivíduos. Com isso, um número reduzido de cruzamentos a campo precisaria ser realizado para avaliação de desempenho.

## REFERÊNCIAS

- ANDORF, C. et al. Technological advances in maize breeding: past, present and future. **Theoretical and Applied Genetics**, v. 132, n. 3, p. 817-849, 2019.
- BERTAN, I. et al. Comparação de métodos de agrupamento na representação da distância morfológica entre genótipos de trigo. **Revista Brasileira de Agrociência**, v. 12, n. 3, p. 279–286, 2006.
- BOTSTEIN, D. et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **The American Journal of Human Genetics**, v. 32, n. 3, p. 314-331, 1980.
- CONAB. Companhia nacional de abastecimento. **Série histórica das safras**. Disponível em: <https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras>. Acesso em: 6 jun. 2020.
- CONTINI, E. et al. Milho: caracterização e desafios tecnológicos. Brasília: Embrapa. **Desafios do Agronegócio Brasileiro**. v. 2, p. 1-45, 2019.
- CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L.A. **Biometria aplicada ao estudo da diversidade genética**. 1. ed. Visconde do Rio Branco: Suprema, 2011.
- ELTAHER, S. et al. Genetic Diversity and Population Structure of F 3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. **Frontiers in genetics**, v. 9, p. 76, 2018.
- FAOSTAT. Food and agriculture organization of the United Nations. **Crop data**. FAO United Nations, Rome. Disponível em: <http://www.fao.org/faostat/en/#data/QC>. Acesso em: 6 jun. 2020.
- FERNANDES E. H., et al. Genetic diversity in elite inbred lines of maize and its association with heterosis. **Genetics and Molecular Research**, v. 14, n. 2, p. 6509–6517, 2015.
- GUO, Z. et al. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. **Molecular Breeding**, v. 39, n. 37, p. 1–12, 2019.
- LENG Y., et al. Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). **Molecular Breeding**, v. 39, n. 38, p. 1–19, 2019.
- MOHAMMADI, S. A.; PRASANNA, B. M. Analysis of genetic diversity in crop plants - Salient Statistical Tools. **Crop Science**, v. 43, n. 4, p. 1235–1248, 2003.
- OLIVEIRA, A. M. C. E. **Associação das abordagens de genética quantitativa e de genética de populações no melhoramento de plantas**. 2015. 80 p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa, 2015.

RASHEED, A. et al. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. **Molecular Plant**, v. 10, n. 8, p. 1047–1064, 2017.

SARAVANAN, K. A. et al. Advanced software programs for the analysis of genetic diversity in livestock genomics: a mini review. **Biological Rhythm Research**, p. 1–11, 2019.

SCHERLOSKY, A. et al. Genetic variability of Brazilian wheat germplasm obtained by high-density SNP genotyping. **Crop Breeding and Applied Biotechnology**, v. 18, p. 399–408, 2018.

SEMAGN, K. et al. Molecular characterization of diverse CIMMYT maize inbred lines from eastern and southern Africa using single nucleotide polymorphic markers. **BMC genomics**, v. 13, n. 1, p. 1-11, 2012.

SHULL, G. H. What is “heterosis”? **Genetics**, v. 33, n. 5, p. 439–446, 1948.

SMITH, J. S. C.; GARDNER, C. A. C.; COSTICH, D. E. Ensuring the genetic diversity of maize and its wild relatives. **Achieving sustainable cultivation of maize. Volume 1: from improved varieties to local applications**, p. 3-50, 2017.

**SEGUNDA PARTE – ARTIGO<sup>1</sup>****HETEROTIC GROUPING OF MAIZE INBRED LINES BASED ON SNP  
GENOTYPING DATA**

Lander Santos de Oliveira<sup>1</sup>, Ivan Schuster<sup>1</sup>, Welison Andrade Pereira<sup>2</sup>, Evandro Novaes<sup>2</sup>

<sup>1</sup>*LongPing High-Tech, SP-330, Km 296, Cravinhos, Brazil*

<sup>2</sup>*Universidade Federal de Lavras, Departamento de Biologia, Lavras, Brazil*

---

<sup>1</sup>Artigo submetido para a revista *Crop Breeding and Applied Biotechnology*. Elaborado de acordo com as normas do periódico.

## ABSTRACT

The advances in genotyping technologies have transformed the way breeding programs manage their genetic resources. The identification of single nucleotide polymorphisms (SNPs) can improve the understanding of genetic diversity of maize (*Zea mays*) inbred lines and their classification into heterotic groups, which is useful to orientate certain crosses to obtain hybrids with higher performance. Genetic diversity of 293 genotypes was investigated with 5252 SNPs with minor allele frequency >5%. SNPs distribution averaged 525 per chromosome. Polymorphism information content averaged 0.297. The unweighted pair group method with arithmetic mean analysis and principal component analysis based on a genetic distance matrix revealed similar clusters with high cophenetic correlation coefficients (0.953 and 0.863, respectively). The results demonstrated consistency compared to previously established heterotic groups using pedigree and breeding information, and also allowed exploration of genetic variability by revealing subgroups within each group.

**Keywords:** Genetic diversity. Single nucleotide polymorphism. *Zea mays*.



## 1 INTRODUCTION

In maize (*Zea mays*) breeding programs, there has always been the interest in classifying germplasm into heterotic groups, which are defined as sets of related genotypes from same or different populations. Genotypes from the same group show similar combining ability and, when crossed with genotypes from another group, exhibit heterosis. High performance hybrids can be explored in allogamous plants throughout crosses between inbred lines (Souza Júnior 2011).

To classify maize inbred lines into heterotic groups, genetic diversity studies are commonly accomplished (Wu et al. 2016, Leng et al. 2019, Silva et al. 2020). An important requirement for heterosis manifestation (Prasad and Singh 1986). The classification of inbred lines into heterotic groups using genetic diversity quantification orientates crosses between those from contrasting groups, giving rise to hybrids with higher heterosis. Consequently, the number of hybrid crosses performed is reduced and breeding programs efficiency is increased (Reif, Hallauer and Melchinger 2005). Thus, a systematized knowledge of maize genetic resources has become necessary for better evaluations regarding their diversity.

Genetic diversity studies to classify inbred lines into groups can be accomplished by using morpho-agronomic characteristics and combining ability analyses, such as diallels and topcrosses. However, these field trials are onerous, since they require large field areas, manual pollination and detasseling labor. Hence, the number of evaluated hybrids is usually restricted (Fernandes et al. 2015, Wu et al. 2016, Kulka et al. 2018). Likewise, genotypes can be allocated into groups based on their genealogy. Although simple, this method requires detailed pedigree information, which is not always available (Lee, Ash and Good 2007, Adu et al. 2019b, Leng et al. 2019).

The use of molecular markers has become the best way to make predictions about genetic diversity among genotypes (Muhammad et al. 2017, Scherlosky et al. 2018, Adu et al. 2019a, Silva et al. 2020), without the need of innumerable crosses and hybrids evaluation in the field. Some molecular markers are highly polymorphic, and independent of environmental effects and plant physiological stages, which are advantages for selecting more divergent parents that will give rise to populations with high variability and adaption ability (Govindaraj, Vetriventhan and Srinivasan 2015, Nadeem et al. 2018).

Molecular information allows estimating genetic similarity among individuals in terms of identity by descent (IBD) or identity by state (IBS) alleles (Messmer et al. 1993). The IBD

between a pair of individuals is the probability that an allele in a given locus of a genotype and an allele from the same locus of another genotype are copies from a common ancestor (Cox et al. 1985). In contrast, IBS is the genotypic similarity of alleles alike in “state”, that is, indistinguishable by their effects and ancestry. The estimates of genetic similarity based on molecular markers reveal the proportion of IBS alleles, regardless of whether their identity is caused by IBD or IBS alleles (Messmer et al. 1993).

Single nucleotide polymorphisms (SNPs) are the most abundant source of variation in organisms, showing dense coverage across genomes compared to other types of molecular markers. This increases the likelihood of some SNPs being associated with genes of interest, which can contribute to improve accuracy of genetic diversity evaluation in breeding programs. Besides, new generation sequencing technologies have offered high-throughput and reduced costs with automated platforms that are appropriate to breeding programs requirements (Rasheed et al. 2017, Guo et al. 2019).

SNP data can be transformed into dissimilarity estimates, followed by clustering methods that assist their classification into heterotic groups. The interpretation of these results can orientate breeders when choosing the most promising maize lines from different groups to be crossed, reducing the number of crosses required during hybrid trials. Consequently, the efficiency of maize breeding programs can be enhanced.

This study assessed how accurate are the clustering methods of maize inbred lines that use genetic dissimilarity information obtained from SNP data, compared to the heterotic groups previously established with pedigree and breeding information.

## 2 MATERIALS AND METHODS

This study comprised 293 maize inbred lines from the breeding program from LongPing High-Tech, Brazil. These lines belonged to four heterotic groups of tropical and temperate backgrounds (Table 1), previously defined by pedigree and topcross information.

**Table 1.** Heterotic groups of 293 inbred lines based on pedigree and breeding information.

Heterotic groups	Characteristics	Number of lines
G1	Tropical	56
G2	Tropical	79
G3	Temperate	4
G4	Tropical	154

The DNA of each maize line was extracted from leaf samples according to the Fast ID Genomic DNA Extraction Kit protocol (Fast ID NA, Inc., Fairfield, IA). DNA quantity and quality were checked by fluorimetry using Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) before genotyping. The SNP genotyping was performed using the MaizeSNP50 BeadChip according to the Infinium HD Assay Ultra Protocol Guide (Illumina, Inc., San Diego, CA). A total of 6231 SNPs were available for this study.

Markers with minor allele frequency (MAF) of less than 5% and with missing data of more than 5% were filtered out. Genetic information of SNPs was assessed by polymorphism information content (PIC), which provides the discriminatory degree of each marker, depending on the number of alleles per locus and its relative frequencies (Botstein et al. 1980). The parameters above were estimated using PowerMarker version 3.25 (Liu and Muse 2005). The physical distribution of SNPs across the maize chromosome was produced using the web-based software PhenoGram (Center for System Genomics, Pennsylvania State University; <http://visualization.ritchielab.psu.edu/>).

Genetic distances were estimated by the complement of identity by state ( $1 - IBS$ ) with TASSEL 5.0 software (Bradbury et al. 2007). Based on the distance matrix, the maize lines were clustered by the unweighted pair group method with arithmetic mean (UPGMA) and principal component analysis (PCA) using R 3.5.1 (R Core Team 2018), with the *ggplot2* version 3.1.0 (Wickham 2016) and *ape* version 5.3 (Paradis and Schlie 2018) packages, respectively. PCA plots were generated with DataWarrior 5.0.0 (Sander et al. 2015). The cophenetic correlation coefficient was calculated, and Mantel's test was performed to check

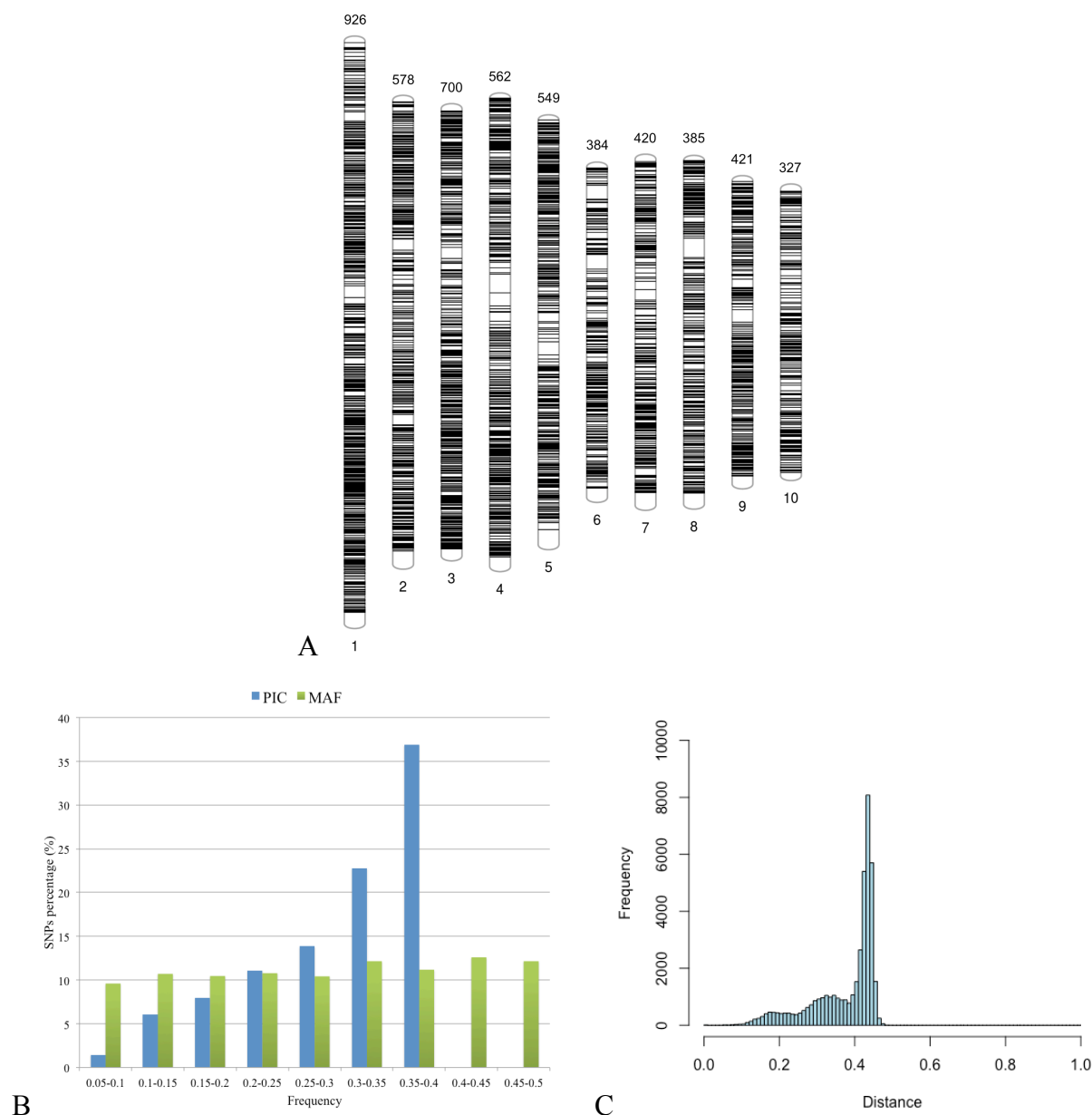
cluster analysis fitness to the genetic distance matrix using *stats* version 3.5.1 (R Core Team 2018) and *ade4* version 1.7-13 (Dray and Dufor 2007) packages.

### 3 RESULTS AND DISCUSSION

A total of 293 maize inbred lines were genotyped with the Illumina MaizeSNP50 array, and 6231 SNPs were available for this study. To assess the genetic diversity information among the inbred lines, 979 markers were filtered out, in which 1.68% represented the monomorphic SNPs; 11.35% referred to those with minor allele frequency (MAF) of less than 5%; and 3.13% were SNPs with more than 5% of missing data. Therefore, the diversity analyses were carried out with a total of 5252 polymorphic SNPs. The missing rate was low, considering an expected 10% of random loss of beads within the Illumina MaizeSNP50 runs (Xu et al. 2017).

The SNPs were distributed all over the maize genome. In average, there were 525 SNPs in each chromosome, varying from 327 in chromosome 10 to 926 in chromosome 1 (Figure 1A). Genomic coverage was proportional to similar studies (Li et al. 2018, Guo et al. 2019, Leng et al. 2019, Silva et al. 2020). Compared to primary molecular marker systems, the usage of high-density marker genotyping makes genetic diversity analysis more powerful. SNP markers provide higher genomic coverage, increasing the chances they will be in the core genome, associated with important agronomic traits (Xu et al. 2017, Scherlosky et al. 2018).

The magnitude of markers informativeness depends on their degree of polymorphism, which reflects on the genetic diversity among genotypes under study (Chesnokov and Artemyeva 2015). To assess the 5252 SNPs informativeness across the 293 inbred lines, polymorphism information content (PIC) and minor allele frequency (MAF) were estimated. They ranged from 0.092 to 0.375 (mean of 0.297) and 0.051 to 0.5 (mean of 0.284), respectively (Figure 1B). Li et al. (2018) and Liu et al. (2015) found similar PIC and MAF range values when studying genetic properties of Chinese maize germplasm using SNP markers data. Wu et al. (2016) found similar values of PIC analyzing tropical and temperate maize inbred lines from CIMMYT. According to Botstein et al. (1980), PIC values ranging from 0.25 to 0.5 indicate that markers are moderately informative. However, the authors considered multiallelic markers for this classification. Considering the biallelic nature of SNPs, in which the maximum value of PIC is 0.375, we can consider PIC values in the higher quartile highly informative. Based on these criteria, 65.6% of the markers in this study were highly informative (Figure 1B).



**Figure 1.** Physical distribution of the 5252 SNPs across the maize genome (A). Distribution of polymorphism information content (PIC) and minor allele frequency (MAF) for 5252 SNPs in a population of 293 maize inbred lines (B). Genetic distances frequency among all maize inbred line pairs (C).

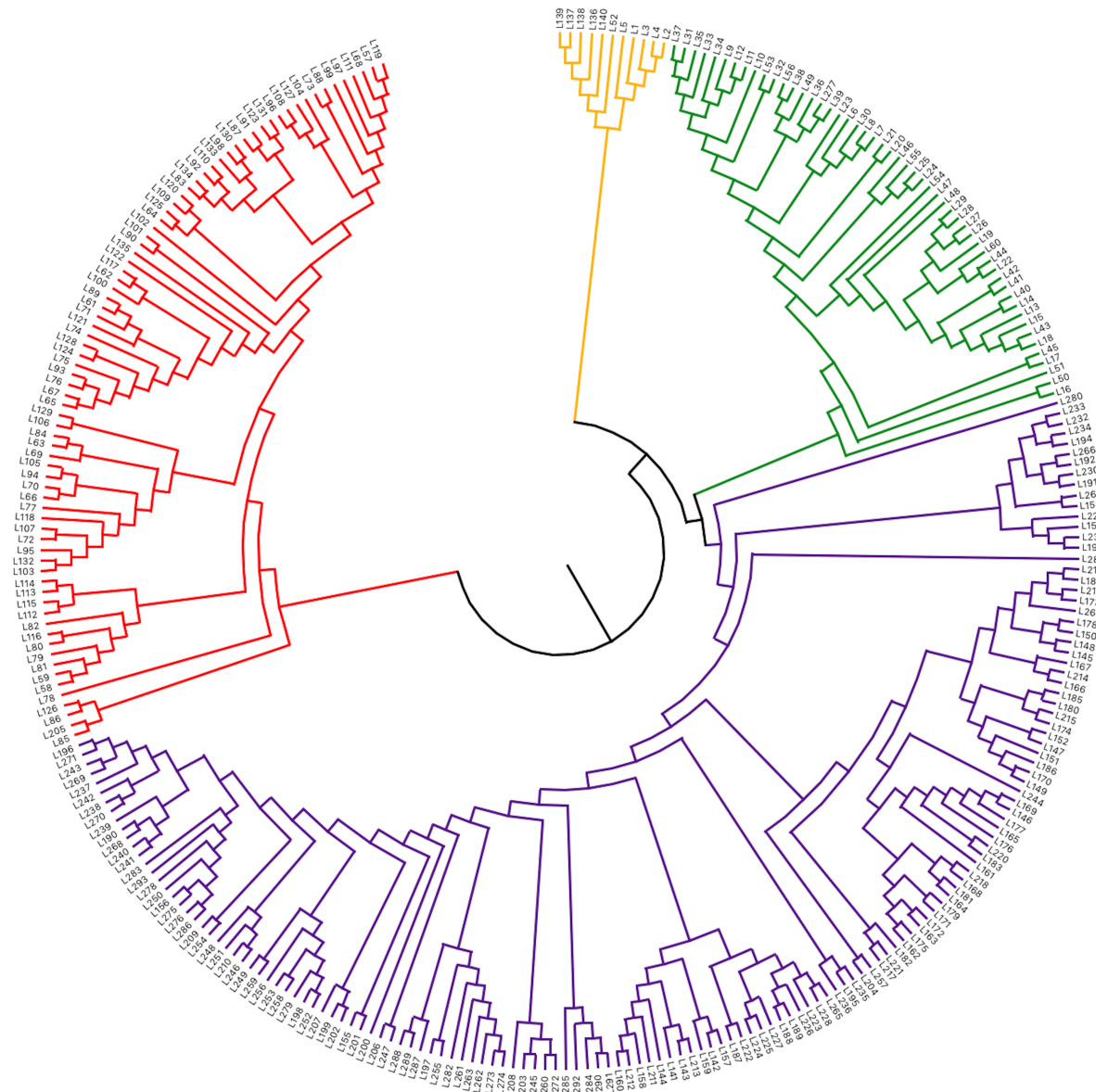
Based on the 5252 polymorphic SNPs, a genetic distance matrix was built among all inbred line pairs, ranging from 0 to 0.491, with an average of 0.375. Figure 1C shows the genetic distance frequencies across all lines. The distance range was higher than those estimated by Ertiro et al. (2017) and Silva et al. (2020), in which distances of tropical maize lines based on IBS genetic similarity ranged from 0.011 to 0.346 (average of 0.313) and 0.003 to 0.253 (average of 0.193), respectively. The smallest distances were obtained between seven line pairs: L61 x L89, L57 x L119, L98 x L130, L124 x L128, L154 x L193, L154 x L231 and

L173 x 216. The largest distance obtained was between lines L136 x L159. These results show that genetic diversity analysis using SNP markers was coherent, once heterotic pattern based on pedigree and breeding information classified line L136 in a temperate group, and line L159 in a tropical group. All the other line pairs with the smallest genetic distances agreed to the tropical groups. Compared to the cited studies, our distances estimates presented a higher average, indicating that there is still genetic variability among lines, even though we evaluated elite maize genotypes originated from breeding programs, which in theory could have led to narrowing of variability (Scherlosky et al. 2018).

**Table 2.** Dendrogram clusters information.

<b>Dendrogram clusters</b>	<b>Number of lines</b>	<b>Corresponding group of the majority of the lines</b>	<b>Number of lines from other groups</b>
C1 (yellow)	11	G3	6 from G1; 1 from G4
C2 (green)	52	G1	1 from G2; 1 from G4
C3 (purple)	151	G4	-
C4 (red)	79	G2	1 from G4

The UPGMA clustering analysis based on the genetic distance matrix resolved 4 distinct clusters, as exhibited in the dendrogram (Figure 2). The significant cophenetic correlation coefficient ( $r = 0.953$ ;  $P < 0.0001$ ; 10,000 permutations) indicated that the clustering analysis showed a very good fit to the genetic distances matrix on which it was based, according to Mohammadi and Prasanna (2003). Beckett et al. (2017) agrees that an accurate dendrogram is important to help breeders classifying their germplasm. Fernandes et al. (2015) and Nikolic et al. (2015) also found positive cophenetic values when inferring genetic diversity using microsatellites ( $r = 0.59$  and  $r = 0.80$ , respectively), but not as high as in this study using SNP data.



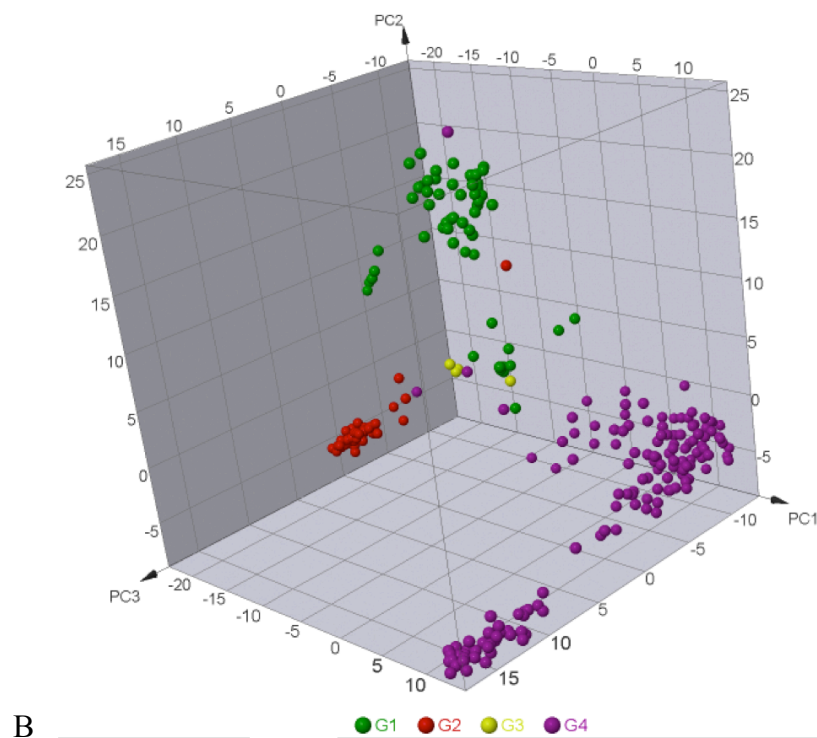
**Figure 2.** Dendrogram displaying 4 clusters: C1 (yellow), C2 (green), C3 (purple), C4 (red).

Originally, the population of this study was organized into four heterotic groups (G1, G2, G3 and G4) based on their genealogy and breeding information (Table 1). The dendrogram also revealed four distinct clusters (C1, C2, C3 and C4), in which the majority of the inbred lines were grouped as expected. However, not all the clusters were entirely coincident with the known heterotic groups (Table 2). For instance, seven inbred lines usually identified as tropical (6 from G1 and 1 from G4) were allocated into the C1 cluster. The C1 cluster contains all temperate maize lines from G3, and based on this cluster analysis, these seven inbred lines are likely to have a temperate genetic background. In summary, from the 293 maize inbred lines analyzed, 10 were classified into other groups (3.4%). This shows the importance of associating molecular and conventional breeding for a more accurate genetic



diversity analyses (Wu et al. 2016). The inconsistencies found in the classification of genotypes using marker data could be related to errors in pedigree information or genetic drift during the process of inbred line creation (Nikolic et al. 2015), as well as mixtures between heterotic groups.

Principal component analysis was performed with the 5252 SNPs and the first three principal components were plotted to investigate cluster patterns, resulting in the discrimination of four clusters (Figure 3). The cophenetic correlation value ( $r = 0.863$ ;  $P < 0.0001$ ; 10,000 permutations) between the PCA distances matrix and the genetic distances matrix suggested that this clustering method was also reliable, once lines were clustered in a manner that is for the most part consistent with their known pedigree and breeding information. The first three principal components (PCs) explained variation proportions of 30.5%, 14.6%, and 7.7%, respectively.



**Figure 3.** Principal component analysis (PCA) based on genetic distance using the first three PCs. Colored dots refer to heterotic groups previously established.

The results demonstrated that both clustering methods provide consistent clusters with ease of visualization, which can help breeders orientate maize lines for topcrosses. Although the dendrogram provides a better visual separation of inner subgroups within each group, the 3D plot generated using the DataWarrior interface makes the interaction between the software

and the user much easier. The PCA method also allowed to infer two big groups (G4 versus the remainder groups), which could benefit breeders while planning hybrid crosses, since they would have less possible hybrid combinations to be evaluated in the field. At the same time, fewer testers would be necessary during topcrosses trials. Therefore, using less heterotic groups could lead to time and cost savings, since smaller planting area and less manual labor would be required.

Although most of the genotypes under this study were classified into the four heterotic groups previously established by the breeding program, it was possible to visualize subgroups within each of the groups, revealing the abundant genetic variability existing in maize germplasm. This was brought to attention thanks to the SNP markers genomic high-density and their intrinsic nature of not being influenced by genotype-by-environment interactions.

The results also support other genetic diversity studies using SNPs as an important tool for clustering inbred lines into genetically related groups and orientating hybrid crosses between individuals of different groups to obtain higher performance (Richard et al. 2016, Dari et al. 2018, Silva et al. 2020). Besides, SNPs are useful to quickly determine the genetic relationship of new inbred lines by genotyping them in early stages of the plant and including them in a new cluster analysis (Beckett et al. 2017).

The main goal of this study was to show that the grouping of maize lines using high-density molecular markers data is a precise methodology, without the need of executing vast combining ability experiments in the field for this classification. However, it is important to say that the information provided by genetic distances do not ensure good hybrids, since their parents need to show genetic complementarity between them (HABLAK, 2018). Classifying maize lines using SNPs showed high accuracy to the classification previously done with pedigree and breeding information.

Besides, the genotyping data generated in this study can be used in further models of hybrid prediction, allowing the identification of hybrids with cost-effective labor, thus improving breeding efficiency.

#### **4 CONCLUSION**

Clustering methods using genetic diversity estimates based on SNP markers are a reliable source for classifying maize inbred lines into heterotic groups, since they have shown strong consistency compared to the previous identification using pedigree and breeding information. The PCA grouping method based on the highly dense markers also allowed the definition of two bigger groups, which could benefit breeders while planning hybrid crosses to be evaluated in the field. Besides, the use this type of molecular marker provides detailed information about each group, allowing exploration of genetic variability within them.

**5 ACKNOWLEDGEMENTS**

This study was financially supported by the company LongPing High-Tech, Brazil.

## 6 REFERENCES

- Adu GB, Awuku FJ, Amegbor IK, Haruna A, Manigben KA and Aboyadana PA (2019a) Genetic characterization and population structure of maize populations using SSR markers. **Annals of Agricultural Sciences** **64**: 47–54.
- Adu GB, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ and Gedil M (2019b) Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. **PLoS ONE** **14**: 1–12.
- Beckett TJ, Morales AJ, Koehler KL and Rocheford TR (2017) Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. **PLoS ONE** **12**: 1–23.
- Botstein D, White RL, Skolnick, M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics** **32**: 314–331.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics** **23**: 2633–2635.
- Chesnokolov Yu. V. and Artemyeva AM (2015) Evaluation of the measure of polymorphism information of genetic diversity. **Agricultural Biology** **50**: 571–578.
- Cox TS, Lookhart GL, Walker DE, Harrell LG, Albers LD and Rodgers DM (1985) Genetic Relationships Among Hard Red Winter Wheat Cultivars as Evaluated by Pedigree Analysis and Gliadin Polyacrylamide Gel Electrophoretic Patterns. **Crop Science** **25**: 1058–1063.
- Dari S, MacRobert J, Minnaar-Ontong A and Labuschagne MT (2018) SNP-based genetic diversity among few-branched-1 (Fbr1) maize lines and its relationship with heterosis, combining ability and grain yield of testcross hybrids. **Maydica** **63**: 1–14.
- Dray S and Dufour A (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. **Journal of Statistical Software** **22**: 1–20.
- Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, Wegary D, Azmach G, Ogugo V, Keno T, Abebe B, Chibsa T and Menkir A (2017) Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. **BMC Genomics** **18**: 1–11.
- Fernandes EH, Schuster I, Scapim CA, Vieira ESN and Coan MMD (2015) Genetic diversity in elite inbred lines of maize and its association with heterosis. **Genetics and Molecular Research** **14**: 6509–6517.
- Govindaraj M, Vetriventhan M and Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. **Genetics Research International** **2015**: 1–14.

- Guo Z, Wang H, Tao J, Ren Y, Xu C, Wu K, Zou C, Zhang J and Xu Y (2019) Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. **Molecular Breeding** **39**: 1–12.
- Hablak S (2019) The Concept of Allelic and Nonallelic Mechanism of Heterosis. **Biochemistry and Molecular Biology** **3**: 77-82.
- Kulka V, Aparecida Silva T, Contreras-Soto RI, Maldonado C, Mora F and Scapim CA (2018) Diallel analysis and genetic differentiation of tropical and temperate maize inbred lines. **Crop Breeding and Applied Biotechnology** **18**: 31–38.
- Lee EA, Ash MJ and Good B (2007) Re-examining the relationship between degree of relatedness, genetic effects, and heterosis in Maize. **Crop Science** **47**: 629–635.
- Leng Y, Lv C, Li L, Xiang Y, Xia C, Wei R, Rong T and Lan H (2019) Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). **Molecular Breeding** **39**: 1–19.
- Li T, Qu J, Wang Y, Chang L, He K, Guo D, Zhang X, Xu S and Xue J (2018) Genetic characterization of inbred lines from Shaan A and B groups for identifying loci associated with maize grain yield. **BMC Genetics** **19**: 1–12.
- Liu C, Hao Z, Zhang D, Xie C, Li M, Zhang X, Yong H, Zhang S, Weng J and Li X (2015) Genetic properties of 240 maize inbred lines and identity-by-descent segments revealed by high-density SNP markers. **Molecular Breeding** **35**: 1-12.
- Liu K and Muse SV. (2005) PowerMaker: An integrated analysis environment for genetic maker analysis. **Bioinformatics** **21**: 2128–2129.
- Messmer MM, Melchinger AE, Herrmann RG and Boppenmaier J (1993) Relationships among early european maize inbreds: II. Comparison of pedigree and RFLP data. **Crop Science** **33**: 944-950.
- Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants - Salient statistical tools. **Crop Science** **43**: 1235–1248.
- Muhammad RW, Qayyum A, Ahmad MQ, Hamza A, Yousaf M, Ahmad B, Younas M, Malik W, Liaqat S and Noor E (2017) Characterization of maize genotypes for genetic diversity on the basis of inter simple sequence repeats. **Genetics and Molecular Research** **16**: 1–9.
- Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G and Baloch FS (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. **Biotechnology and Biotechnological Equipment** **32**: 261–285.
- Nikolić A, Ignjatović-Micić D, Kovačević D, Čamdžija Z, Filipović M and Drinić SM (2015) Genetic diversity of maize inbred lines as inferred from SSR markers. **Genetika** **47**: 489–498.

Paradis E and Schliep K (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. **Bioinformatics** **35**: 526-528.

Prasad SK and Singh TP (1986) Heterosis in relation to genetic divergence in maize (*Zea mays* L.). **Euphytica** **35**: 919–924.

R CORE TEAM (2013) **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Available at <<http://www.R-project.org/>>. Accessed on February 29, 2020.

Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK and He Z (2017) Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. **Molecular Plant** **10**: 1047–1064.

Reif JC, Hailauer AR and Melchinger AE (2005) Heterosis and heterotic patterns in maize. **Maydica** **50**: 215–223.

Richard C, Osiru DS, Mwala MS and Lubberstedt T (2016) Genetic diversity and heterotic grouping of the core set of southern African and temperate maize (*Zea mays* L) Inbred lines using SNP markers. **Maydica** **61**: 1–9.

Sander T, Freyss J, von Korff M and Rufener C (2015) DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. **Journal of Chemical Information and Modeling** **55**: 460–473.

Scherlosky A, Sérgio Marchioro V, De Assis Franco F, Braccini AL and Schuster I (2018) Genetic variability of Brazilian wheat germplasm obtained by high-density SNP genotyping. **Crop Breeding and Applied Biotechnology** **18**: 399–408.

Silva KJ, Guimarães CT, Guilhen JHS, Guimarães PE de O, Parentoni SN, Trindade R dos S, de Oliveira AA, Bernardino K da C, Pinto M de O, Dias KO das G, Bernardes C de O, Dias LA dos S, Guimarães LJM and Pastina MM (2020) High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. **Crop Science** **2020**; 1-9.

Souza Júnior CL (2011) Cultivar development of allogamous crops. **Crop Breeding and Applied Biotechnology** **11**: 8–15.

Wickham H (2016) **ggplot2: Elegant Graphics for Data Analysis**. Springer, New York. 276p.

Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X and Babu R (2016) Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. **Theoretical and Applied Genetics** **129**: 753–765.


Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P and Zou C (2017) Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. **Molecular Breeding** **37**: 1-12.

## **ANEXOS**

### **ANEXO A – PRODUTO TÉCNICO**

**Tutorial de análise molecular de diversidade genética para classificação de linhagens de milho**



		<b>TUTORIAL</b>	
ANÁLISE MOLECULAR DE DIVERSIDADE GENÉTICA PARA CLASSIFICAÇÃO DE LINHAGENS DE MILHO			
<b>Elaborador</b>	<b>Revisor</b>	<b>Aprovador</b>	<b>Data de aprovação</b>
<i>Lander Oliveira</i>	<i>Ivan Schuster</i>	<i>Ivan Schuster</i>	<i>22/06/2020</i>

## INTRODUÇÃO

A análise de diversidade genética vem se tornando uma prática comum no contexto de ensino, pesquisa e empresarial. No caso da LongPing High-Tech não é diferente, dado a sua missão de prover sementes de alta tecnologia, com agilidade e excelência, e contribuindo para alimentar a humanidade. A importância desta análise para a classificação de genótipos é uma realidade e há muita expectativa em relação aos ganhos efetivos que esta metodologia pode proporcionar. Para isso, é essencial que haja um procedimento bem delineado que assegure o mínimo de desvios durante sua execução. A equipe do Laboratório de Pesquisa e Desenvolvimento vem trabalhando neste processo, que tem se mostrado adequado para os interesses da empresa. Portanto, este produto técnico, atrelado ao trabalho de conclusão de curso de Mestrado Profissional, é elaborado para padronização das atividades e a sua descrição detalhada garante a qualidade do resultado final.

A partir da próxima seção, o tutorial será apresentado de forma objetiva, com a finalidade central de fornecer um caminho seguro e eficaz para que os colaboradores da equipe possam executar, chegando com precisão ao resultado final em suas análises de rotina. Assim, espera-se que este documento siga propiciando boas diretrizes para a obtenção de genótipos superiores.

## TERMOS E DEFINIÇÕES

<b>Termo</b>	<b>Definição</b>
MAF	Frequência do alelo menos frequente.
PCA	Análise de componentes principais.
PIC	Conteúdo de informação de polimorfismo. Índice calculado para cada marcador molecular, que expressa a informatividade do marcador.
SNP	Polimorfismo de nucleotídeo único.

UPGMA	Método de agrupamento hierárquico baseado nas médias das distâncias.
-------	--

**ETAPAS**

Etapa	Ação																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
<p><b>1</b></p> <p>Preparo dos dados</p>	<ul style="list-style-type: none"> <li>A Figura 1 mostra os dados brutos de genotipagem.</li> </ul> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> </tr> <tr> <th>1</th> <th>###Chr</th> <th>1</th> <th>1</th> <th>2</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>8</th> </tr> <tr> <th>2</th> <th>Pos</th> <th>1234</th> <th>2345</th> <th>3456</th> <th>4567</th> <th>5678</th> <th>6789</th> <th>7890</th> <th>8901</th> <th>9012</th> </tr> <tr> <th>3</th> <th>MrkID</th> <th>Marker 1</th> <th>Marker 2</th> <th>Marker 3</th> <th>Marker 4</th> <th>Marker 5</th> <th>Marker 6</th> <th>Marker 7</th> <th>Marker 8</th> <th>Marker 9</th> </tr> </thead> <tbody> <tr><td>4</td><td>Linhagem 1</td><td>T/T</td><td>C/C</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>C/C</td><td>T/T</td><td>C/C</td></tr> <tr><td>5</td><td>Linhagem 2</td><td>T/T</td><td>C/C</td><td>C/C</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>C/C</td><td>C/C</td></tr> <tr><td>6</td><td>Linhagem 3</td><td>T/T</td><td>C/C</td><td>C/C</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>C/C</td><td>T/T</td></tr> <tr><td>7</td><td>Linhagem 4</td><td>T/T</td><td>C/C</td><td>C/C</td><td>G/G</td><td>A/A</td><td>G/G</td><td>A/A</td><td>C/C</td><td>T/T</td></tr> <tr><td>8</td><td>Linhagem 5</td><td>T/T</td><td>./</td><td>C/C</td><td>G/G</td><td>A/A</td><td>G/G</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> <tr><td>9</td><td>Linhagem 6</td><td>T/T</td><td>C/C</td><td>A/A</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> <tr><td>10</td><td>Linhagem 7</td><td>./</td><td>C/C</td><td>C/C</td><td>G/G</td><td>A/A</td><td>G/G</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> <tr><td>11</td><td>Linhagem 8</td><td>./</td><td>C/C</td><td>A/A</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> <tr><td>12</td><td>Linhagem 9</td><td>T/T</td><td>G/G</td><td>A/A</td><td>A/A</td><td>C/C</td><td>T/T</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> <tr><td>13</td><td>Linhagem 10</td><td>T/T</td><td>G/G</td><td>C/C</td><td>G/G</td><td>A/A</td><td>G/G</td><td>A/A</td><td>T/T</td><td>T/T</td></tr> </tbody> </table> <p style="text-align: center;"><b>Figura 1.</b> Dados genotípicos brutos.</p> <ul style="list-style-type: none"> <li>A Figura 2 mostra os dados da planilha de Excel anexa a este tutorial.</li> </ul> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> <th>K</th> <th>L</th> <th>M</th> <th>N</th> <th>O</th> <th>P</th> <th>Q</th> <th>R</th> <th>S</th> <th>T</th> <th>U</th> <th>V</th> <th>W</th> <th>X</th> <th>Y</th> </tr> <tr> <th>1</th> <th>Dados Perdidos</th> <th>Dados Válidos</th> <th>% Dados Perdidos</th> <th>A/A</th> <th>G/G</th> <th>C/C</th> <th>T/T</th> <th>A/C</th> <th>A/T</th> <th>A/G</th> <th>G/A</th> <th>G/C</th> <th>G/T</th> <th>C/A</th> <th>C/G</th> <th>C/T</th> <th>T/A</th> <th>T/G</th> <th>T/C</th> <th>Freq. A</th> <th>Freq. G</th> <th>Freq. C</th> <th>Freq. T</th> <th>MAF</th> <th>PIC</th> </tr> </thead> <tbody> <tr><td>2</td><td>0</td><td>0</td><td>#DIV/0!</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr> <tr><td>3</td><td>0</td><td>0</td><td>#DIV/0!</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr> <tr><td>4</td><td>0</td><td>0</td><td>#DIV/0!</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr> <tr><td>5</td><td>0</td><td>0</td><td>#DIV/0!</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr> </tbody> </table> <p style="text-align: center;"><b>Figura 2.</b> Colunas da planilha anexa.</p> <ul style="list-style-type: none"> <li>Copie as células com dados genotípicos e cole-as transpostas (Figura 3A) na coluna “AA” da planilha de Excel anexa (Figura 3B).</li> </ul> <div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p><b>A</b></p> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th>AA</th> <th>AB</th> <th>AC</th> <th>AD</th> <th>AE</th> <th>AF</th> <th>AG</th> <th>AH</th> <th>AI</th> <th>AJ</th> <th>AK</th> <th>AL</th> <th>AM</th> </tr> <tr> <th>###Chr</th> <th>Pos</th> <th>MrkID</th> <th>Linhagem 1</th> <th>Linhagem 2</th> <th>Linhagem 3</th> <th>Linhagem 4</th> <th>Linhagem 5</th> <th>Linhagem 6</th> <th>Linhagem 7</th> <th>Linhagem 8</th> <th>Linhagem 9</th> <th>Linhagem 10</th> </tr> </thead> <tbody> <tr><td>1</td><td>1234</td><td>Marker 1</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>./</td><td>./</td><td>T/T</td><td>T/T</td></tr> <tr><td>1</td><td>2345</td><td>Marker 2</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td><td>./</td><td>C/C</td><td>C/C</td><td>C/C</td><td>G/G</td><td>G/G</td></tr> <tr><td>2</td><td>3456</td><td>Marker 3</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td><td>A/A</td><td>C/C</td><td>A/A</td><td>A/A</td><td>C/C</td></tr> <tr><td>2</td><td>4567</td><td>Marker 4</td><td>A/A</td><td>A/A</td><td>A/A</td><td>G/G</td><td>G/G</td><td>A/A</td><td>G/G</td><td>A/A</td><td>A/A</td><td>G/G</td></tr> <tr><td>3</td><td>5678</td><td>Marker 5</td><td>C/C</td><td>C/C</td><td>C/C</td><td>A/A</td><td>A/A</td><td>C/C</td><td>A/A</td><td>C/C</td><td>C/C</td><td>A/A</td></tr> <tr><td>4</td><td>6789</td><td>Marker 6</td><td>T/T</td><td>T/T</td><td>T/T</td><td>G/G</td><td>G/G</td><td>T/T</td><td>G/G</td><td>T/T</td><td>T/T</td><td>G/G</td></tr> <tr><td>5</td><td>7890</td><td>Marker 7</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td></tr> <tr><td>6</td><td>8901</td><td>Marker 8</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td></tr> <tr><td>8</td><td>9012</td><td>Marker 9</td><td>T/T</td><td>C/C</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td></tr> </tbody> </table> </div> <div style="width: 48%;"> <p><b>B</b></p> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> <th>K</th> <th>L</th> <th>M</th> <th>N</th> <th>O</th> <th>P</th> <th>Q</th> <th>R</th> <th>S</th> <th>T</th> <th>U</th> <th>V</th> <th>W</th> <th>X</th> <th>Y</th> <th>AA</th> <th>AB</th> <th>AC</th> <th>AD</th> <th>AE</th> <th>AF</th> </tr> <tr> <th>1</th> <th>Dados Perdidos</th> <th>Dados Válidos</th> <th>% Dados Perdidos</th> <th>A/A</th> <th>G/G</th> <th>C/C</th> <th>T/T</th> <th>A/C</th> <th>A/T</th> <th>A/G</th> <th>G/A</th> <th>G/C</th> <th>G/T</th> <th>C/A</th> <th>C/G</th> <th>C/T</th> <th>T/A</th> <th>T/G</th> <th>T/C</th> <th>Freq. A</th> <th>Freq. G</th> <th>Freq. C</th> <th>Freq. T</th> <th>MAF</th> <th>PIC</th> <th>###Chr</th> <th>Pos</th> <th>MrkID</th> <th>Linhagem 1</th> <th>Linhagem 2</th> <th>Linhagem 3</th> </tr> </thead> <tbody> <tr><td>2</td><td>0</td><td>8</td><td>20.0%</td><td>0</td><td>0</td><td>8</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td><td>100%</td><td>0.00</td><td>1</td><td>1234</td><td>Marker 1</td><td>T/T</td><td>T/T</td><td>T/T</td></tr> <tr><td>3</td><td>1</td><td>9</td><td>10.0%</td><td>0</td><td>2</td><td>7</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>22%</td><td>78%</td><td>0%</td><td>22%</td><td>0.35</td><td>1</td><td>2345</td><td>Marker 2</td><td>C/C</td><td>C/C</td><td>C/C</td></tr> <tr><td>4</td><td>0</td><td>10</td><td>0.0%</td><td>3</td><td>0</td><td>7</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>70%</td><td>0%</td><td>30%</td><td>0.42</td><td>2</td><td>3456</td><td>Marker 3</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td></tr> <tr><td>5</td><td>0</td><td>10</td><td>0.0%</td><td>6</td><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>60%</td><td>40%</td><td>0%</td><td>40%</td><td>0.48</td><td>2</td><td>4567</td><td>Marker 4</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td></tr> <tr><td>6</td><td>0</td><td>10</td><td>0.0%</td><td>4</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>60%</td><td>0%</td><td>40%</td><td>0.48</td><td>3</td><td>5678</td><td>Marker 5</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td></tr> <tr><td>7</td><td>0</td><td>10</td><td>0.0%</td><td>0</td><td>4</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>40%</td><td>0%</td><td>60%</td><td>0.48</td><td>4</td><td>6789</td><td>Marker 6</td><td>T/T</td><td>T/T</td><td>T/T</td><td>T/T</td></tr> <tr><td>8</td><td>0</td><td>10</td><td>0.0%</td><td>10</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>100%</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td><td>0.00</td><td>5</td><td>7890</td><td>Marker 7</td><td>A/A</td><td>A/A</td><td>A/A</td><td>A/A</td></tr> <tr><td>9</td><td>0</td><td>10</td><td>0.0%</td><td>0</td><td>0</td><td>4</td><td>6</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>40%</td><td>60%</td><td>40%</td><td>0.48</td><td>6</td><td>8901</td><td>Marker 8</td><td>C/C</td><td>C/C</td><td>C/C</td><td>C/C</td></tr> <tr><td>10</td><td>0</td><td>10</td><td>0.0%</td><td>0</td><td>0</td><td>1</td><td>9</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0%</td><td>0%</td><td>10%</td><td>90%</td><td>0.18</td><td>8</td><td>9012</td><td>Marker 9</td><td>T/T</td><td>C/C</td><td>C/C</td><td>T/T</td></tr> </tbody> </table> </div> </div> <p style="text-align: center;"><b>Figura 3.</b> (A) Dados genotípicos transpostos; (B) Dados colados na coluna “AA” da planilha de fórmulas.</p>		A	B	C	D	E	F	G	H	I	J	1	###Chr	1	1	2	2	3	4	5	6	8	2	Pos	1234	2345	3456	4567	5678	6789	7890	8901	9012	3	MrkID	Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7	Marker 8	Marker 9	4	Linhagem 1	T/T	C/C	A/A	C/C	T/T	A/A	C/C	T/T	C/C	5	Linhagem 2	T/T	C/C	C/C	A/A	C/C	T/T	A/A	C/C	C/C	6	Linhagem 3	T/T	C/C	C/C	A/A	C/C	T/T	A/A	C/C	T/T	7	Linhagem 4	T/T	C/C	C/C	G/G	A/A	G/G	A/A	C/C	T/T	8	Linhagem 5	T/T	./	C/C	G/G	A/A	G/G	A/A	T/T	T/T	9	Linhagem 6	T/T	C/C	A/A	A/A	C/C	T/T	A/A	T/T	T/T	10	Linhagem 7	./	C/C	C/C	G/G	A/A	G/G	A/A	T/T	T/T	11	Linhagem 8	./	C/C	A/A	A/A	C/C	T/T	A/A	T/T	T/T	12	Linhagem 9	T/T	G/G	A/A	A/A	C/C	T/T	A/A	T/T	T/T	13	Linhagem 10	T/T	G/G	C/C	G/G	A/A	G/G	A/A	T/T	T/T		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	1	Dados Perdidos	Dados Válidos	% Dados Perdidos	A/A	G/G	C/C	T/T	A/C	A/T	A/G	G/A	G/C	G/T	C/A	C/G	C/T	T/A	T/G	T/C	Freq. A	Freq. G	Freq. C	Freq. T	MAF	PIC	2	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	3	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	4	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	5	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	###Chr	Pos	MrkID	Linhagem 1	Linhagem 2	Linhagem 3	Linhagem 4	Linhagem 5	Linhagem 6	Linhagem 7	Linhagem 8	Linhagem 9	Linhagem 10	1	1234	Marker 1	T/T	T/T	T/T	T/T	T/T	T/T	./	./	T/T	T/T	1	2345	Marker 2	C/C	C/C	C/C	C/C	./	C/C	C/C	C/C	G/G	G/G	2	3456	Marker 3	C/C	C/C	C/C	C/C	C/C	A/A	C/C	A/A	A/A	C/C	2	4567	Marker 4	A/A	A/A	A/A	G/G	G/G	A/A	G/G	A/A	A/A	G/G	3	5678	Marker 5	C/C	C/C	C/C	A/A	A/A	C/C	A/A	C/C	C/C	A/A	4	6789	Marker 6	T/T	T/T	T/T	G/G	G/G	T/T	G/G	T/T	T/T	G/G	5	7890	Marker 7	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	6	8901	Marker 8	C/C	C/C	C/C	C/C	T/T	T/T	T/T	T/T	T/T	T/T	8	9012	Marker 9	T/T	C/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AA	AB	AC	AD	AE	AF	1	Dados Perdidos	Dados Válidos	% Dados Perdidos	A/A	G/G	C/C	T/T	A/C	A/T	A/G	G/A	G/C	G/T	C/A	C/G	C/T	T/A	T/G	T/C	Freq. A	Freq. G	Freq. C	Freq. T	MAF	PIC	###Chr	Pos	MrkID	Linhagem 1	Linhagem 2	Linhagem 3	2	0	8	20.0%	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	0%	0%	100%	100%	0.00	1	1234	Marker 1	T/T	T/T	T/T	3	1	9	10.0%	0	2	7	0	0	0	0	0	0	0	0	0	0	0	0	0%	22%	78%	0%	22%	0.35	1	2345	Marker 2	C/C	C/C	C/C	4	0	10	0.0%	3	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0%	70%	0%	30%	0.42	2	3456	Marker 3	C/C	C/C	C/C	C/C	5	0	10	0.0%	6	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	60%	40%	0%	40%	0.48	2	4567	Marker 4	A/A	A/A	A/A	A/A	6	0	10	0.0%	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0%	60%	0%	40%	0.48	3	5678	Marker 5	C/C	C/C	C/C	C/C	7	0	10	0.0%	0	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0%	40%	0%	60%	0.48	4	6789	Marker 6	T/T	T/T	T/T	T/T	8	0	10	0.0%	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	0%	0%	100%	0.00	5	7890	Marker 7	A/A	A/A	A/A	A/A	9	0	10	0.0%	0	0	4	6	0	0	0	0	0	0	0	0	0	0	0	0%	40%	60%	40%	0.48	6	8901	Marker 8	C/C	C/C	C/C	C/C	10	0	10	0.0%	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0%	0%	10%	90%	0.18	8	9012	Marker 9	T/T	C/C	C/C	T/T
	A	B	C	D	E	F	G	H	I	J																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
1	###Chr	1	1	2	2	3	4	5	6	8																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
2	Pos	1234	2345	3456	4567	5678	6789	7890	8901	9012																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
3	MrkID	Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7	Marker 8	Marker 9																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
4	Linhagem 1	T/T	C/C	A/A	C/C	T/T	A/A	C/C	T/T	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
5	Linhagem 2	T/T	C/C	C/C	A/A	C/C	T/T	A/A	C/C	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
6	Linhagem 3	T/T	C/C	C/C	A/A	C/C	T/T	A/A	C/C	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
7	Linhagem 4	T/T	C/C	C/C	G/G	A/A	G/G	A/A	C/C	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
8	Linhagem 5	T/T	./	C/C	G/G	A/A	G/G	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
9	Linhagem 6	T/T	C/C	A/A	A/A	C/C	T/T	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
10	Linhagem 7	./	C/C	C/C	G/G	A/A	G/G	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
11	Linhagem 8	./	C/C	A/A	A/A	C/C	T/T	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
12	Linhagem 9	T/T	G/G	A/A	A/A	C/C	T/T	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
13	Linhagem 10	T/T	G/G	C/C	G/G	A/A	G/G	A/A	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
1	Dados Perdidos	Dados Válidos	% Dados Perdidos	A/A	G/G	C/C	T/T	A/C	A/T	A/G	G/A	G/C	G/T	C/A	C/G	C/T	T/A	T/G	T/C	Freq. A	Freq. G	Freq. C	Freq. T	MAF	PIC																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
2	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
3	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
4	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
5	0	0	#DIV/0!	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
###Chr	Pos	MrkID	Linhagem 1	Linhagem 2	Linhagem 3	Linhagem 4	Linhagem 5	Linhagem 6	Linhagem 7	Linhagem 8	Linhagem 9	Linhagem 10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
1	1234	Marker 1	T/T	T/T	T/T	T/T	T/T	T/T	./	./	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
1	2345	Marker 2	C/C	C/C	C/C	C/C	./	C/C	C/C	C/C	G/G	G/G																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
2	3456	Marker 3	C/C	C/C	C/C	C/C	C/C	A/A	C/C	A/A	A/A	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
2	4567	Marker 4	A/A	A/A	A/A	G/G	G/G	A/A	G/G	A/A	A/A	G/G																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
3	5678	Marker 5	C/C	C/C	C/C	A/A	A/A	C/C	A/A	C/C	C/C	A/A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
4	6789	Marker 6	T/T	T/T	T/T	G/G	G/G	T/T	G/G	T/T	T/T	G/G																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
5	7890	Marker 7	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
6	8901	Marker 8	C/C	C/C	C/C	C/C	T/T	T/T	T/T	T/T	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
8	9012	Marker 9	T/T	C/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	AA	AB	AC	AD	AE	AF																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
1	Dados Perdidos	Dados Válidos	% Dados Perdidos	A/A	G/G	C/C	T/T	A/C	A/T	A/G	G/A	G/C	G/T	C/A	C/G	C/T	T/A	T/G	T/C	Freq. A	Freq. G	Freq. C	Freq. T	MAF	PIC	###Chr	Pos	MrkID	Linhagem 1	Linhagem 2	Linhagem 3																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
2	0	8	20.0%	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	0%	0%	100%	100%	0.00	1	1234	Marker 1	T/T	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
3	1	9	10.0%	0	2	7	0	0	0	0	0	0	0	0	0	0	0	0	0%	22%	78%	0%	22%	0.35	1	2345	Marker 2	C/C	C/C	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
4	0	10	0.0%	3	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0%	70%	0%	30%	0.42	2	3456	Marker 3	C/C	C/C	C/C	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
5	0	10	0.0%	6	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	60%	40%	0%	40%	0.48	2	4567	Marker 4	A/A	A/A	A/A	A/A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
6	0	10	0.0%	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0%	60%	0%	40%	0.48	3	5678	Marker 5	C/C	C/C	C/C	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
7	0	10	0.0%	0	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0%	40%	0%	60%	0.48	4	6789	Marker 6	T/T	T/T	T/T	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
8	0	10	0.0%	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	0%	0%	100%	0.00	5	7890	Marker 7	A/A	A/A	A/A	A/A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
9	0	10	0.0%	0	0	4	6	0	0	0	0	0	0	0	0	0	0	0	0%	40%	60%	40%	0.48	6	8901	Marker 8	C/C	C/C	C/C	C/C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
10	0	10	0.0%	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0%	0%	10%	90%	0.18	8	9012	Marker 9	T/T	C/C	C/C	T/T																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
<p><b>2</b></p>	<ul style="list-style-type: none"> <li>Utilizando a opção filtrar, exclua os marcadores que apresentem:</li> </ul>																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			

Filtragem	<ul style="list-style-type: none"> <li>○ Dados perdidos &gt;5%;</li> <li>○ MAF &lt;5%;</li> <li>○ MAF = 100% (marcadores monomórficos).</li> </ul>
<b>3</b> Interpretação de PIC	<ul style="list-style-type: none"> <li>• Para estimar o conteúdo de informação polimórfica (PIC) dos marcadores selecionados, utilizou-se a fórmula abaixo:                     <math display="block">PIC = 1 - \sum_{j=1}^n p_j^2</math> </li> <li>• Para interpretar os valores de PIC obtidos, deve-se considerar a natureza bi-alélica dos SNPs, isto é, o valor máximo de PIC para estes marcadores é 0,5. Portanto, os marcadores com valores de PIC acima do terceiro quartil são considerados altamente informativos.</li> </ul>

<b>4</b> Preparo do arquivo <i>hapmap</i>	<ul style="list-style-type: none"> <li>• A matriz de distâncias é gerada pelo software TASSEL 5.0.</li> <li>• O arquivo de entrada para o software deve ser do tipo <i>hapmap</i> (.hmp).</li> <li>• Para produzir este arquivo (Figura 4), basta:             <ul style="list-style-type: none"> <li>○ Alterar os genótipos para o formato de 2 letras (por exemplo, “A/A” deve ser “AA”);</li> <li>○ Alterar dados perdidos no formato “./.” para “NN”;</li> <li>○ Alterar os cabeçalhos da colunas “MrkID”, “###Chrm” e “Pos” para “rs#”, “Chrom” e “pos”, respectivamente.</li> <li>○ Adicionar as demais colunas e organizá-las na mesma ordem mostrada na Figura 4.</li> <li>○ Remover qualquer espaço presente nos nomes dos marcadores e das linhagens.</li> </ul> </li> </ul>
--	--

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	rs#	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panel	QCode	Linhagem 1	Linhagem 2	Linhagem 3	Li
2	Marker 3	C/A	2	3456	+	NA	NA	NA	NA	NA	NA	CC	CC	CC	
3	Marker 4	A/G	2	4567	+	NA	NA	NA	NA	NA	NA	AA	AA	AA	
4	Marker 5	A/C	3	5678	+	NA	NA	NA	NA	NA	NA	CC	CC	CC	
5	Marker 6	G/T	4	6789	+	NA	NA	NA	NA	NA	NA	TT	TT	TT	
6	Marker 8	C/T	6	8901	+	NA	NA	NA	NA	NA	NA	CC	CC	CC	
7	Marker 9	C/T	8	9012	+	NA	NA	NA	NA	NA	NA	TT	CC	TT	

**Figura 4.** Modelo do arquivo *hapmap* para o TASSEL.

- A coluna “alleles” refere-se aos alelos que cada marcador possui.

**Tabela 1.** Fórmulas para gerar a coluna “alleles” do arquivo *hapmap*.

A	C	G	T	Concatenar	alleles
=IF(COUNTIF(\$Q2:\$Z2,"AA"),"A", "")	=IF(COUNTIF(\$Q2:\$Z2,"CC"),"C", "")	=IF(COUNTIF(\$Q2:\$Z2,"CC"),"C", "")	=IF(COUNTIF(\$Q2:\$Z2,"TT"),"T", "")	=CONCATENATE(B2,C2,D2,E2)	=CONCATENATE(LEFT(F2),"/",RIGHT(F2))

- Para criar a coluna “alleles”, insira 6 colunas temporárias em seu arquivo (Figura 5) e adicione as fórmulas da Tabela 3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	rs#	A	C	G	T	Concatenar	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panel	QCode	Linhagem 1	Linhagem 2
2	Marker 3	A	C			AC	A/C	2	3456	+	NA	NA	NA	NA	NA	NA	CC	CC
3	Marker 4	A		G		AG	A/G	2	4567	+	NA	NA	NA	NA	NA	NA	AA	AA
4	Marker 5	A	C			AC	A/C	3	5678	+	NA	NA	NA	NA	NA	NA	CC	CC
5	Marker 6			G	T	GT	G/T	4	6789	+	NA	NA	NA	NA	NA	NA	TT	TT
6	Marker 8		C		T	CT	C/T	6	8901	+	NA	NA	NA	NA	NA	NA	CC	CC
7	Marker 9		C		T	CT	C/T	8	9012	+	NA	NA	NA	NA	NA	NA	TT	CC

**Figura 5.** Produção da coluna “alleles” para o arquivo *hapmap* (em cinza).

- Por fim, apague as colunas desnecessárias a fim de manter o arquivo no formato *hapmap*.
- Para correta leitura do arquivo pelo TASSEL, o arquivo precisa estar organizado em ordem crescente de cromossomo (“chrom”) e posição (“pos”).
- Salve o arquivo no formato *.txt* (Tab Delimited Text), adicionando *.hmp* no nome do seu arquivo:
  - Exemplo: *Arquivo\_hapmap.hmp.txt*

## 5

### Matriz de distâncias genéticas

- Abra o programa TASSEL 5.0.
- Clique em File > Open e abra seu arquivo hapmap (Figura 6).

	0: 3456	1: 4567	2: 5678	4: 8901	5: 9012
Linhagem1	C	A	C	T	C
Linhagem2	C	A	C	T	C
Linhagem3	C	A	C	T	C
Linhagem4	C	G	A	G	C
Linhagem5	C	G	A	G	T
Linhagem6	A	A	C	T	T
Linhagem7	C	G	A	G	T
Linhagem8	A	A	C	T	T
Linhagem9	A	A	C	T	T
Linhagem10	C	G	A	G	T

**Figura 6.** Dados importados para o TASSEL a partir do arquivo *hapmap*.

- Clique em Analysis > Relatedness > Distance Matrix (Figura 7).

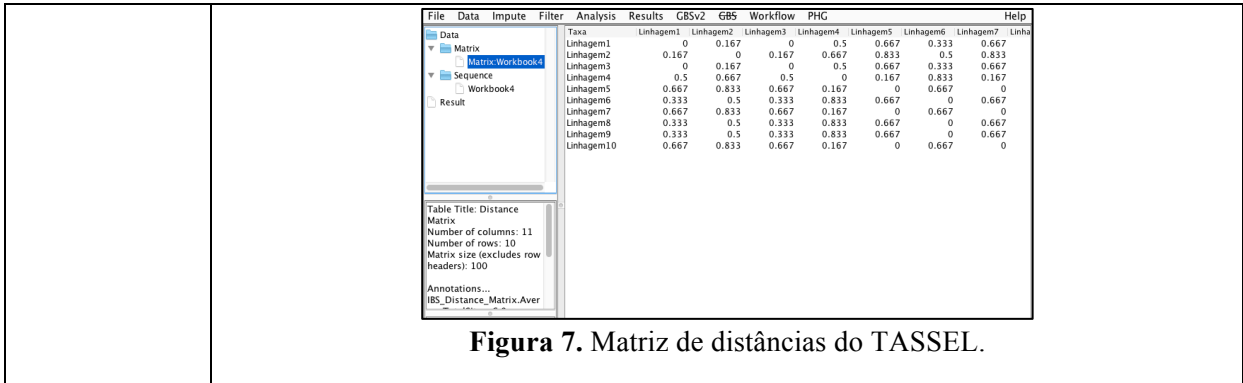


Figura 7. Matriz de distâncias do TASSEL.

**6**  
Análise de componentes principais (PCA)

- Para realizar a análise dos componentes principais (PCA) utilizando o TASSEL, clique em Analysis > Diversity > PCA.
- Na janela que se abre, preencha o campo “number of componentes” com o número 3 (Figura 8).

Figura 8. Número de componentes da análise de PCA.

- Exporte a tabela da Figura 9 clicando em Results > Table.

Taxa	PC1	PC2	PC3
Linagem1	0.625	-0.656	-0.305
Linagem2	0.714	-0.909	0.632
Linagem3	0.625	-0.656	-0.305
Linagem4	-1.007	-0.544	-0.255
Linagem5	-1.126	0.226	0.089
Linagem6	0.807	0.696	-0.011
Linagem7	-1.126	0.226	0.089
Linagem8	0.807	0.696	-0.011
Linagem9	0.807	0.696	-0.011
Linagem10	-1.126	0.226	0.089

Figura 9. Análise de PCA do TASSEL.

- Abra o programa DataWarrior.
- Clique em File > Open para importar a tabela com os dados de PCA. Automaticamente, os gráficos 2D e 3D são gerados (Figura 10).
- Para selecionar um grupo de linhagens, utilize o botão esquerdo do mouse. Para girar o gráfico 3D, utilize o botão direito do mouse.

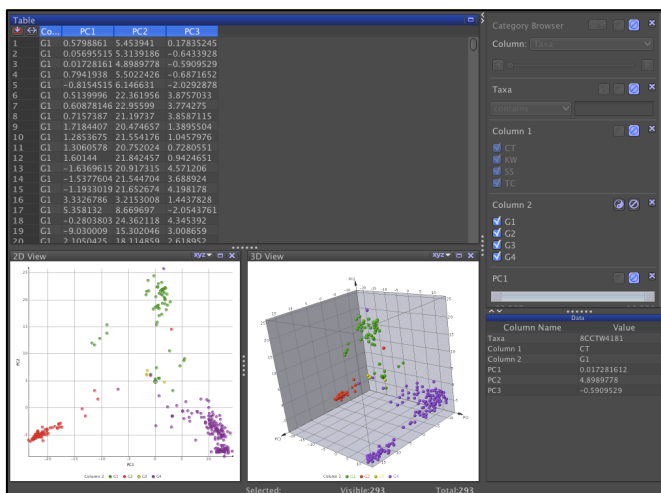


Figura 10. Gráficos de PCA no DataWarrior.

7  
Preparo do  
arquivo *mega*  
para análise  
de  
agrupamento

- Exporte a matriz de distâncias genéticas geradas no TASSEL clicando em Results > Table.
- Prepare o arquivo de entrada para o programa MEGAX:
  - Converta a matriz de distâncias em um arquivo Excel que contenha somente a porção inferior da matriz. Para isso, exclua a primeira coluna e as duas primeiras linhas da matriz (Figura 11).

	A	B	C	D	E	F	G	H	I	J	K
1	Taxa	Linhagem1	Linhagem2	Linhagem3	Linhagem4	Linhagem5	Linhagem6	Linhagem7	Linhagem8	Linhagem9	Linhagem10
2	Linhagem1	0	0.167	0.000	0.500	0.667	0.333	0.667	0.333	0.333	0.667
3	Linhagem2	0.167	0	0.167	0.667	0.833	0.500	0.833	0.500	0.500	0.833
4	Linhagem3	0.000	0.167	0	0.500	0.667	0.333	0.667	0.333	0.333	0.667
5	Linhagem4	0.500	0.667	0.500	0	0.167	0.833	0.167	0.833	0.833	0.167
6	Linhagem5	0.667	0.833	0.667	0.167	0	0.667	0.000	0.667	0.667	0.000
7	Linhagem6	0.333	0.500	0.333	0.833	0.667	0	0.667	0.000	0.000	0.667
8	Linhagem7	0.667	0.833	0.667	0.167	0.000	0.667	0	0.667	0.667	0.000
9	Linhagem8	0.333	0.500	0.333	0.833	0.667	0.000	0.667	0	0.000	0.667
10	Linhagem9	0.333	0.500	0.333	0.833	0.667	0.000	0.667	0.000	0	0.667
11	Linhagem10	0.667	0.833	0.667	0.167	0.000	0.667	0.000	0.667	0.667	0

Figura 11. Linhas e coluna a serem excluídas.

- Em seguida, utilize a MacroDiagonal fornecida com este tutorial. Abra a macro, clique em Developer > Macros > Run. A matriz final deve parecer com a da Figura 12.

	A	B	C	D	E	F	G	H	I
1	0.167								
2	0.000	0.167							
3	0.500	0.667	0.500						
4	0.667	0.833	0.667	0.167					
5	0.333	0.500	0.333	0.833	0.667				
6	0.667	0.833	0.667	0.167	0.000	0.667			
7	0.333	0.500	0.333	0.833	0.667	0.000	0.667		
8	0.333	0.500	0.333	0.833	0.667	0.000	0.667	0.000	
9	0.667	0.833	0.667	0.167	0.000	0.667	0.000	0.667	0.667

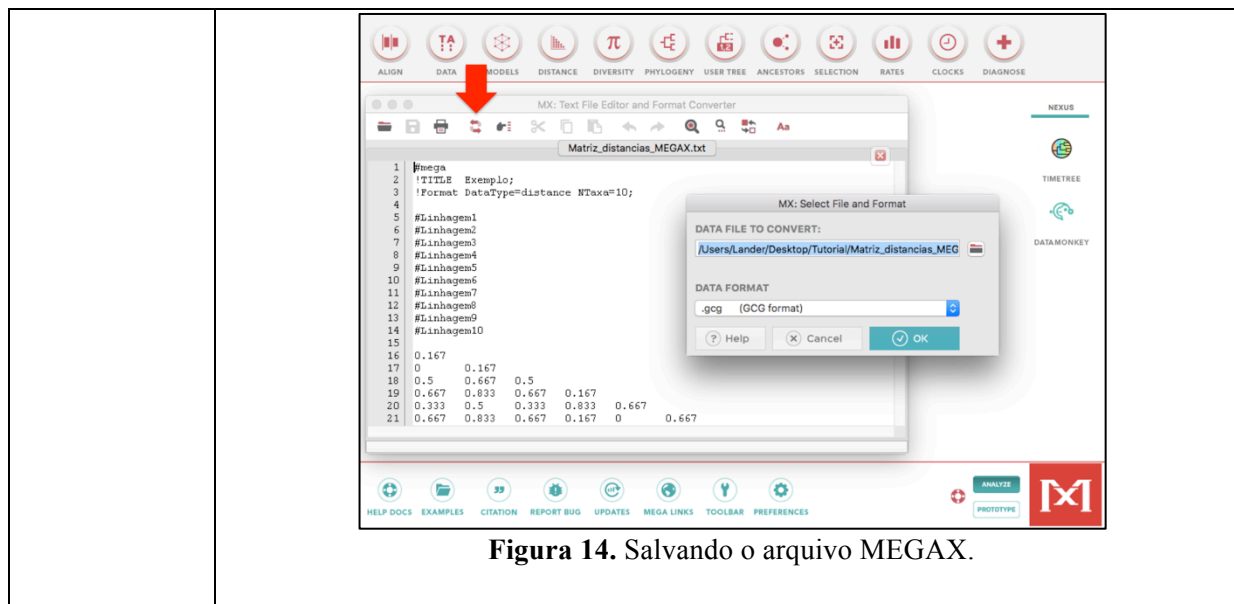
**Figura 12.** Matriz de distâncias inferior.

- Neste arquivo, insira as informações mostradas na Figura 13 acima da matriz inferior.
  - Na célula A3, modifique o número 10 pelo número de linhagens que estão sendo analisadas.
  - Altere o nome do projeto, trocando “Exemplo” pelo nome desejado.

1	#mega								
2	!TITLE Exemplo;								
3	!Format DataType=distance NTaxa=10;								
4									
5	#Linhagem1								
6	#Linhagem2								
7	#Linhagem3								
8	#Linhagem4								
9	#Linhagem5								
10	#Linhagem6								
11	#Linhagem7								
12	#Linhagem8								
13	#Linhagem9								
14	#Linhagem10								
15									
16	0.167								
17	0	0.167							
18	0.5	0.667	0.5						
19	0.667	0.833	0.667	0.167					
20	0.333	0.5	0.333	0.833	0.667				
21	0.667	0.833	0.667	0.167	0	0.667			
22	0.333	0.5	0.333	0.833	0.667	0	0.667		
23	0.333	0.5	0.333	0.833	0.667	0	0.667	0	
24	0.667	0.833	0.667	0.167	0	0.667	0	0.667	0.667

**Figura 13.** Arquivo de dados MEGAX.

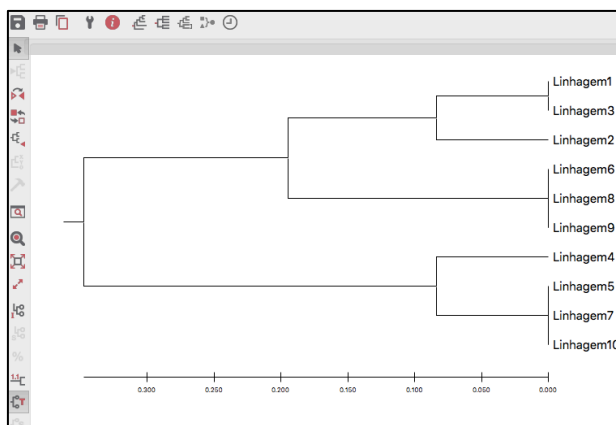
- Salve o arquivo no formato *.txt* (Tab Delimited Text).
- Abra o software MEGAX.
- Clique em Data > Open a File/Session e abra o arquivo em *.txt*.
- Clique nas setas (apontadas na Figura 14). Em seguida, escolha o formato *.gcg* e clique em OK.
- Salve o arquivo com extensão *.meg*.



**Figura 14.** Salvando o arquivo MEGAX.

**8**  
Análise de agrupamento (UPGMA)

- No programa MEGAX, clique em Phylogeny > Construct/Test UPGMA Tree e abra o arquivo com extensão *.meg*.
- Na janela que se abre, selecione as opções “Pairwise Distance” e “Lower Left Matrix” e clique OK nas mensagens que aparecem.
- O dendrograma é então criado. Ele pode ser formatado com as diversas opções da janela (Figura 15).



**Figura 15.** Dendrograma gerado pelo programa MEGAX.

## CONSIDERAÇÕES FINAIS

A partir de 2020, este tutorial passa a ser implantado como um guia de rotina para os colaboradores do Laboratório de Pesquisa e Desenvolvimento da LongPing High-Tech, na realização da análise de diversidade genética. O documento descreve um procedimento objetivo, que otimiza o tempo despendido para as análises e assegura uma execução



padronizada, bem como os resultados obtidos.

Nos próximos anos, o tutorial poderá ser utilizado para treinamento de novos membros da equipe. Adicionalmente, o documento permitirá rastreabilidade ao processo, uma vez que a análise poderá ser realizada pelo aprendiz e treinador, e os dados finais comparados, com a finalidade de detectar possíveis equívocos no procedimento. Além disso, vale ressaltar que o documento será revisado anualmente, como já ocorre com todos os demais procedimentos operacionais da empresa, ajustando os pontos de melhoria observados durante sua execução rotineira no ano anterior.