

## **GENOME-WIDE SNP ALLELOTYPING OF THE BRAZILIAN MAIZE CORE COLLECTION WITH THE EMBRAPA MULTISPECIES 65K INFINIUM CHIP**

**Bianca de Sousa Alcântara**<sup>1,2</sup>; **Flávia Franca Teixeira**<sup>3</sup>; **Guilherme Ferreira Simiqueli**<sup>1</sup>; **Dario Grattapaglia**<sup>1,2</sup>

<sup>1</sup>. . Embrapa Recursos Genéticos e Biotecnologia; <sup>2</sup>. . Universidade de Brasília; <sup>3</sup>. . Embrapa Milho e Sorgo

### **Abstract:**

The core collection of the Brazilian Maize germplasm collection (BMCC) with 300 accessions was established in 1999 to represent the genetic diversity of more than the 4000 accessions organized into four races: indigenous, old commercial, recent commercial and exotic commercial. This classification, based on morphology and geography could be enhanced or validated by genomic data to better clarify its genetic structure and diversity. An accession of outcrossed maize corresponds to a population with wide diversity among individuals. However, the analysis of several individuals per accession makes the process impracticable. To enable fast and precise genetic analysis of the core collection we adopted an allelotyping approach of pools of DNA, to estimate allele frequencies for a set of 3526 SNPs genotyped on the EMBRAPA 65K Multispecies Infinium chip. We initially evaluated different methods of preparing DNA pools. For each of four accessions, DNA samples were extracted from each of 16 seedlings. These DNA samples were quantified by Nanodrop absorbance and Q-bit fluorescence. Three pool assembly treatments with four technical replicates were compared: (1) Pools of equimolar nanodrop measured DNA; (2) Pools of equimolar Q-Bit measured DNA; (3) Direct DNA extraction of bulked leaf samples weighed to the nearest 0.1 gram. The root-mean-square deviation (RMSD) was used to quantify the difference between the allele frequencies estimated from the pooled sample and the true values calculated from genotype data of the same 16 individuals. Allele frequency estimates for the pools were obtained by the GenomeStudio B-Allele Frequency (BAF), a measure used in SNPchip-based analysis of CNVs (copy number variants) in human cancer. BAF is a normalized measure of the allelic intensity ratio of the two SNP alleles (A and B). BAF near 1 or 0 indicates the complete absence of one of the alleles (AA or BB), and intermediate BAF values can be precisely estimated. Surprisingly, the treatment providing the lowest RMSD was the bulked leaf extraction method with an average RMSD across the accessions and replicates of 5.7% against an RMSD of 6.6% for nanodrop and 8.4% for Q-bit. DNA pools from bulked leaf samples were genotyped for 287 BMCC accessions, and two up to four biological replicates for 64 of them to evaluate reproducibility. Reproducibility among technical replicates was assessed by Wright's fixation index ( $F_{st}$ ) across all 3526 SNPs. The average  $F_{st}$  between replicates across the 64 accessions was 0.0047 indicating that accurate estimates can be obtained with minimal sampling error. A k-means clustering based on  $F_{st}$  among all 287 accessions ( $F_{st}$  from 0.0017 to 0.57) detected six genetic groups in the BMCC also visualized in a dendrogram. Matching between these groups and the originally intended subclassifications of the BMCC, including landraces, introductions and improved material, as well as tropical or temperate and the four grain types (pop, flint, floury, dent) is currently underway. In a follow-up study we will use allelotyping to study 200 indigenous Brazilian maize varieties to understand their divergence, structure and potential genomic signatures of domestication.

**Palavras-chave:** Germplasm; Allelotyping; SNPs; *Zea mays*; Genetic diversity.

### **Support / Acknowledgment**

Fundação de Apoio à Pesquisa do Distrito Federal - FAPDF.