

## RESUMO - BIOTECNOLOGIA E GENÔMICA

### **A CHROMOSOME SCALE GENOME ASSEMBLY OF PENSACOLA BAHIAGRASS (PASPALUM NOTATUM CV. PENSACOLA)**

*Joyce Etsuko Arakaki (joyce.e.arakaki@gmail.com)*

*Bianca Baccili Zanotto Vigna (bianca.vigna@embrapa.br)*

*Marco Pessoa Filho (marco.pessoa@embrapa.br)*

*Mauricio De Alvarenga Mudadu (mauricio.mudadu@embrapa.br)*

*Esteban Rios (estebanrios@ufl.edu)*

*Alessandra Pereira Fávero (alessandra.favero@embrapa.br)*

The development of new, cost-effective sequencing technologies, along with user-friendly computational tools, facilitates genomic studies of non-model species. *Paspalum notatum* is an important forage and turfgrass worldwide. Pensacola is a diploid ( $2n=20$ ) cultivar with haploid genome size of 600Mb, estimated by flow cytometry. Pensacola is native to Argentina and has long narrow leaves, early flowering and cold tolerance. It is related to five ('Tipher 1', 'Tipher 2', 'UF-Riata', 'Tifton 9' and 'Tifquick') of the 17 released cultivars and the most adopted cultivar in Southeastern Florida and South of Brazil, indicating its agro-economic importance. Here, we present a chromosome-scale genome assembly of *Paspalum notatum* cv. Pensacola. To obtain the PacBio HiFi sequences, Pensacola seeds were germinated and grown in a greenhouse at the University of Florida, and high molecular weight DNA extraction was performed from leaf tissue. PacBio HiFi sequencing was performed in a Sequel II system at Maryland Genomics, using one SMRT cell 8M in Circular

Consensus Sequencing mode (CCS) with 30h of sequencing time. About 31.6Gb were obtained in ~2.4M reads with an average of 13.4kb. The assemblies and haplotype separation were performed using Hifiasm and HiCANU+Purge\_dups tools. For the assembly, different combinations of softwares and types of sequences were tested: A) Hifiasm + PacBio/HiFi; B) Hifiasm + Pacbio/HiFi and Illumina/Hi-C (NCBI data - SRX13424350); and C) HiCanu + PacBio/HiFi. The sizes of the observed assemblies ranged between 540 and 560Mb, close to the genome size and to a recently published *P. notatum* assembly (cv. Crowver) with 541Mb (GCA\_022530915.1). Assemblies's statistics were: A) 479 contigs obtained with N50 = 54Mb, Mean contig size (Mcontig) = 1,2Mb, Longest contig size (Lcontig) = 78.3Mb and Shortest contig size (Scontig) = 15,7kb; B) 520 contigs, N50 = 47.2Mb, Mcontig = 995kb, Lcontig = 57.3Mb and Scontig = 2,4kb; C) 127 contigs, N50 = 41.6Mb, Mcontig = 3.7Mb, Lcontig = 60.2Mb and Scontig = 15.4kb. Assembled genomes were evaluated for their repetitive content with RepeatModeler and RepeatMasker programs, which masked ~60% of the genome. BUSCO software assessed the assemblies completeness and the results ranged between 97% and 99% of complete BUSCOs. Ragout was performed in Assembly A, using Crowver as reference, to obtain chromosome-scale scaffolds and resulted in 10 scaffolds ranging from 42Mb to 78Mb, whose sizes were similar to the chromosomes from cv. Crowver. Comparative analysis with D-GENIES showed collinearity between scaffolds obtained from Ragout and Crowver assembly, with a few inversions in centromeric regions. Prediction of gene regions will be performed by ab initio analyses and also based on transcriptome data. A high-quality chromosome-scale genome assembly of *P. notatum* cv. Pensacola will boost breeding programs by providing reliable information for marker and gene discovery, genomic selection, among others.