

RESUMO - MELHORAMENTO DE ESPÉCIES PERENES

A DRAFT GENOME ASSEMBLY OF THE BARU TREE (DIPTERYX ALATA VOGEL) AS A RESOURCE FOR DOMESTICATION, BREEDING, AND CONSERVATION

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

Thannya Nascimento Soares (autor1@sigaeventos.com.br)

Mariana Pires De Campos Telles2 (autor2@sigaeventos.com.br)

Alexandre Siqueira Guedes Coelho (alexandre_coelho@ufg.br)

Lázaro José Chaves (autor3@sigaeventos.com.br)

The baru tree (*Dipteryx alata*, Fabaceae) is native to tropical savannas in Latin America and provides protein-rich seeds highly regarded for their nutritional benefits. Baru seed production is mostly based on extractivism, which, in addition to habitat loss, led to its inclusion in the IUCN Red List of Threatened Species. Genomic resources and propagation methods will support a program for baru domestication and breeding, so that commercial production systems are established and impacts of extractivism on natural populations are reduced. We present a draft genome assembly of the *D. alata*, based on third-generation PacBio CCS HiFi reads. High molecular weight DNA was extracted from young leaves of a single diploid individual. Sequencing was performed on a Sequel II system at Maryland Genomics, using one SMRT cell 8M on CCS HiFi mode, with a 30-hour runtime. Approximately 16 Gb were obtained in 1,4 million CCS reads, with an average size of 11kbp. The partially phased draft assembly was obtained with Hifiasm in HiFi-only mode. The draft contains 142 contigs with a

total size of 702 Mbp, close to the 800 Mbp previously estimated for the species by flow cytometry. Contig N50 was 27.3 Mbp, and 22 contigs longer than 10 Mbp (15.5% of the total) added up to 85% of the assembly (592.6 Mbp). Assessment of the assembled gene space using BUSCO (embryophyta_odb10) showed 99.4% complete orthologs, with only 0.4% fragmented and 0.2% missing BUSCOs. RepeatModeler and RepeatMasker were used to identify repetitive regions, and 57.9% of the assembly were masked as repeats. Ongoing activities include RNAseq for prediction of gene models and resequencing of germplasm accessions for SNPs discovery. A high-quality assembly for the baru tree will provide a foundation for variant discovery and the development of genotyping platforms that will aid its accelerated breeding, domestication and understanding of its genetic diversity in nature. Accelerated breeding assisted by genomic tools and availability of elite clones will help establish sustainable commercial production systems for smallholder farmers, reducing pressure on natural populations and impacts on biodiversity.