Characterization of repetitive DNA sequences in organelle genomes from *Hymenaeae stigonocarpa* (Mart.) using high throughput sequencing data

Cruz, Lays Karolina Soares¹,³; Antunes, Adriana Maria¹,³; Melo, Arthur Tavares de Oliveira²; Coelho, Alexandre Siqueira Guedes²; Novaes, Evandro²; Soares, Thannya Nascimento³; Collevatti, Rosane Garcia³; Telles, Mariana Pires de Campos³.

¹ Bolsista CAPES de doutorado no Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Ciências Biológicas, UFG, Goiânia, GO.
² Laboratório de Genética e Genômica de Plantas, Escola de Agronomia, UFG, Goiânia, GO.
³ Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, UFG, Goiânia, GO.

Autor para correspondência: Lays K. S. Cruz e-mail: layskarolina02@gmail.com

*Hymenaea stigonocarpa* Mart. (Fabaceae), popularly known as “Jatobá-do-Cerrado”, is an important medicinal tree species commonly found in the Brazilian Cerrado biome. Its bark is commonly used to relieve gastric pains, ulcers, diarrheas and inflammations, whereas the edible fruit pulp is used for home-made recipes such as flour, jam and jellies. *H. stigonocarpa* is a common Cerrado biome tree with uncharacterized genome. This study aimed to identify and characterize microsatellites and repetitive DNA elements in genomic sequences from mitochondria and chloroplast genome. DNA was extracted from leaves of five individual of *H. stigonocarpa* sampled from germplasm collection located in EA/UFG. The DNA library was prepared according Nextera Illumina protocol (Illumina, Inc.). Paired-end libraries were sequenced using the Illumina MiSeq platform. Read qualities were evaluated using FastQC and Trimmomatic softwares. Phred Q < 30 was taken as the minimum quality threshold. Reads smaller than 50 bp were discarded. High quality reads were separated into nuclear and organelles DNA sequence using the bowtie2 aligner against chloroplast and mitochondrial databases set up from NCBI data. Draft de novo assemblies of each organellar reads were obtained using Velvet. The organelles reads were assembled in 96 and 180 scaffolds with N50 of 2094 and 1389 bp for chloroplast and mitochondria respectively. These drafts assemblies were forward processed used in IMEx and RepeatMasker web servers for repetitive elements search. Were found 2 retroelements in mitochondrial assembly (L1/CIN4 – LINEs family and Gypsy/DIRS1 – LTR elements family), and other 2 retroelements in chloroplast assembly (L1/CIN4 – LINEs family and Gypsy/DIRS1 – LTR elements family) representing 1.57% and 3.20% of total nucleotide sequences used in assemblies, respectively. According to the literature these transposons are the most common to find in plant genomes and we found no difference when we search for these elements using different plant genomes as a reference in RepeatMasker. Were also found 82 different microsatellites in chloroplast genome, all of them have hexanucleotide motifs size, with only two exception (AATA and AT) that occurs once time in each. Most of elements occurs only once, while three motifs occurs twice (TATAGT, ATTTTC and AAAAGA), We also found 120 different microsatellites in mitochondrial assembly, all of them have hexanucleotide motifs and most of all elements occurs once, while six motifs occurs twice (TATCTT, ATTTTC, AGGAAA, AAGTGA, AACCTC and AAGGAA). All of these hexanucleotide microsatellites probably belongs to coding regions because of the motif size, which makes sense considering they are organelles genomes. This study shows a very interesting approach that could help us to understand better an uncharacterized plant genome.

Keywords: IMEx, Jatobá-do-Cerrado; Microsatellites; RepeatMasker; Transposons.