

## Analysis of the repetitive DNA fraction in species of the caesalpinia group (Leguminosae) reveal high abundance of chromovirus retrotransposons

Van-Lume, B<sup>1,2</sup>; Báez, M<sup>1</sup>; Huettel, B.<sup>2</sup>; Pedrosa-Harand, A.<sup>1</sup>;  
Souza, G<sup>1,2</sup>

### Abstract/Resumo

The Caesalpinia group includes ~205 species and 24 genera with pantropical distribution, of which only three occur in northeastern Brazil: *Cenostigma*, *Libidibia* and *Paubrasilia*. These genera, placed in three different clades, are characterized by having a numerically stable karyotype  $2n = 24$  (16M + 8A) and are the only of the Caesalpinia group presenting large heterochromatic chromomycin A3 (CMA<sup>+</sup>) bands at proximal chromosome regions. In order to characterize the repetitive fraction of these Northeastern Caesalpinia genera, genomic DNA was extracted from *C. microphyllum*, *L. ferrea* and *P. echinata* and sequenced by Illumina technology (2 x 250 bp, paired-end read mode). Repetitive DNA was analyzed using the Galaxy/RepeatExplorer-Elixir platform, which groups sequences based on similarity, generating clusters for different repetitive DNA families. A genome coverage of  $0.1 \times$  (1C = 616.14, 891.88 and 917.02 Mbp for *P. echinata*, *C. microphyllum*, and *L. ferrea*, respectively) was analyzed for each species, and only clusters with relative abundance of at least 0.01% of the genome were considered. The 298 clusters generated for *P. echinata*, the 260 clusters for *C. microphyllum* and the 272 clusters for *L. ferrea* constitute approximately 60%, 38% and 42% of each genome, respectively. The analyzes revealed only three high confidence satellite DNA families for *C. microphyllum* and two for *L. ferrea*, with low genome abundance each. All species presented a greater proportion of Ty3-Gypsy elements than Ty1-Copia, with the Chromovirus lineage being the most abundant. Phylogenetic analysis based on the Chromovirus reverse transcriptase domain suggests homology of the most abundant elements identified in each species. This suggests that these retroelements may be strong candidates to occur at pericentromeric CMA<sup>+</sup> heterochromatic bands. Comparative chromosomal mapping of these repetitive elements is necessary to confirm the presence of Chromovirus at proximal regions and will help understanding how amplification/desamplification dynamics drives heterochromatic diversity in Caesalpinia.

Keyword/Palavras-chave: Heterochromatin; Repetitive DNA; Caesalpinia; NGS

1 Laboratory of Plant Cytogenetics and Evolution, Federal University of Pernambuco - [brenavan-lume@hotmail.com](mailto:brenavan-lume@hotmail.com)

2 Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding, Cologne, Germany