BAC combined with overgos gene-rich strategies for sugarcane genome sequencing

SIDENY LIMA NUNES [1], ROBERTA ALVARES CAMPOS [1], ABDALA ALMEIDA [1], MILTON NISHIYAMA [1], CHANGSOO KIM [3], NATHALIA DE SETTA [2], CAROLINA LEMBKE [1], GLAUCIA MENDES SOUZA [1]

1. Depto. de Bioquímica, Instituto de Bioquímica, Universidade de São Paulo, Brazil.
2. Depto. de Botânica, Instituto de Biociências, Universidade de São Paulo, Brazil.
3. Center for Applied Genetic Technologies, University of Georgia, USA.

Sugarcane is a major feedstock used for biofuel production worldwide. Sequencing of sugarcane genome is a key step for the development of the of molecular biology tools that will provide genes and promoters to the use in the generation of transgenic varieties and molecular markers for breeding. Hybridization using overgo probes is an established approach for screening arrayed bacterial artificial chromosome (BAC) libraries. We have improved the use of overgos by increasing the yield of positive clones without use radioisotopes, instead overgo gene-rich regions for construct BAC sugarcane libraries. Also, a total of 6,021 overgos hybridized to sugarcane BACs were extracted from the BACMan database with corresponding nucleotide sequences. The overgos were blasted against sorghum, rice, and selected sugarcane EST clusters. Since an overgo is a short sequence (40 nucleotides). Aiming to increase accuracy and generate quality data for assembly in one-contig sugarcane BAC, with overgo approach we combine a technique called pooling BAC, for the use of the same space physical gasket, with specially designed adapters for sequence connection and identification called Mid-tag (Multiplex Identifiers). The Mid-tag had unique sequence for identification and fishing each BAC sequence on gasket. We obtained a first race in 454 Genome Sequencer FLX System (Roche), with 18 sugarcane BACs sequenced. Resulting in 1,169,584 sequences considered good quality (Genome Analyzer 454, Roche), 495 bp average size, resulting 377 Mbp of high quality. All generated sequences were assembled with PHRAP Program. Assemblies generated one-contig size of average 110 kb. The BACs had 150 kb size, but contains vector-related sequences (E. coli $\leq$ 1% and vector pBeloBAC11 $\leq$ 10%), trimmed. Genes analysis in AUGUSTUS Predictor Pipeline conclude that approximately 30% of BACs selected using overgos contained target genes sequences, others 20% of overgos BACs had genes from the same structural family or orthologous genes.

Word Keys: sugarcane, BAC, overgo
Supported by: FAPESP, CNPq and CAPES