SBG

59º Congresso Brasileiro de Genética

Functional annotation of genomic sequences of *Hevea brasiliensis*

Conson, ARO[1]; Souza, LM[1]; Toledo-Silva, G[1]; Cardoso-Silva, CB[1]; Mantello, CC[1]; Silva, CC[1]; Le Guen, V[2]; Garcia, AAF[3]; Souza, AP[1,4].

[1]Molecular Biology Center and Genetic Engineering, UNICAMP, Campinas, SP, Brazil; [2]CIRAD, UMR AGAP, Montpellier, Hérault, France; [3]Department of Genetics, ESALQ - University of São Paulo, Piracicaba, SP, Brazil; [4]Department of Plant Biology – Biology Institute and Molecular Biology Center and Genetic Engineering, UNICAMP, Campinas, SP, Brazil

andreconson@gmail.com

Keywords: Rubber tree, gene annotation, high-throughput sequencing, ontology, Euphorbiaceae

The rubber tree *(Hevea brasiliensis)* is a recent domesticated endemic tree species from Amazonian rain forest. It is an important crop to rubber industries due the latex producing, which has better quality compared to the synthetic rubber. Difficulties for genetic improvement were attributed to its perennial nature and long juvenile period, so genomic resources should provide insights into the chromosome organization and help elucidate the regulatory architecture of loci involved in biological processes, including metabolism, growth, development and immunity mechanisms. In this context, functional annotation were developed from the genomic sequencing of two rubber tree cultivars (GT1 and RRIM 701) aiming the better understanding of the *Hevea* genome. DNA sequencing libraries were previously constructed from the cultivars of rubber tree using the Illumina GAIIx. A stringent filtering resulted in 6,995 contigs, which were screened for homolog sequences through Blastx search (e-value: $\leq 10^{-6}$) against NCBI's non-redundant protein database. Also, a bi-directional best hit blast was performed against Kyoto Encyclopedia of Genes and Genomes (KEGG) database using KEGG Automatic Annotation Server. After the step of clustering and homology search via blast, 2,369 ORFs were found, from which, 490 complete, 748 internal and 1,131 partial. Blastx search returned a total of 2,176 hits for a range of plant species. Among the various plants that have protein sequences in GenBank the top-hit species were attributed to *Ricinus communis* (1,146 hits), followed by *Populus trichocarpa* (350 hits) and *Vitis vinifera* (214 hits).A total of 1,523 ORFs were assigned to Gene Ontology (GO) categories, which were distributed to cellular components (33.53%), biological processes (40.40%) and molecular functions (26.07%). The most representatives GO terms among the biological processes were metabolic processes (11.05%) and cellular processes (9.94%). Both *Hevea brasiliensis* and *Ricinus communis* are classified in the Euphorbiaceae family, which may explain for the high similarities that were observed. Representative sequences were mapped to important pathways, such as glycolysis, carbon fixation, starch and sucrose metabolism, amino acid biosynthesis, plant hormone signal transduction, biosynthesis of secondary metabolites and plant-pathogen interactions, using the KEGG database. The possibility of utilize genome sequence information form a potential resource to study the biology, evolution, and future genomic-assisted crop improvements of the rubber tree. Financial support: FAPESP and CNPq