Resumos do 59º Congresso Brasileiro de Genética • 16 a 19 de setembro de 2013
Hotel Monte Real Resort • Águas de Lindóia • SP • Brasil
www.sbg.org.br - ISBN 978-85-89109-06-2

Genome sequencing of a *Hevea brasiliensis* for single nucleotide polymorphism discovery

Souza, LM[1]; Toledo-Silva, G[1]; Cardoso-Silva, CB[1]; Conson, AR[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

Keywords: Rubber tree, SNP, molecular markers

The rubber tree (Hevea spp.), is the primary plant used in natural rubber production. Historically, the breeding of rubber trees has been based on techniques involving statistics and quantitative genetic approaches to determine the best genotypes to be used as new cultivars. The discovery of molecular genetic markers has provided new possibilities for characterizing genotypes for the purpose of identifying cultivars, analyzing genetic diversity, establishing relationships between agricultural traits and genetic factors (QTLs), and identifying genes of interest. The application of next generation sequencing technology has brought a new opportunity for high throughput single nucleotide polymorphism (SNP) discovery. Knowledge about SNPs markers is extremely important in the development of genotyping assays, allowing improvements in plant breeding through marker-assisted selection. In this project, we carried out genomic sequencing of two rubber tree cultivars. The DNA libraries were constructed for two cultivars of rubber tree (two from GT1 and two from RRIM701) and sequenced in Illumina plataform. The resulting short reads (72 bp) were submitted to quality filtering and then were *de novo* assembled using the CLC software. Next, Burrows-Wheeler Aligner (BWA) aligned short reads back to assembled contigs, and Freebayes was used to perform variant calling and snp detection. Genotype likelihoods were computed and variable positions in the aligned reads were compared to the reference contigs. Using the varFilter command of VCFutils script, SNPs were filtered only for positions with a minimal mapping quality (-Q) and coverage (-d) of 30 and 10 respectively. Unique and shared SNPs among the two cultivars were extracted with the VCFtools software. SNPs located in contigs containing open reading frames (ORFs) ≥200 bp were also extracted using transdecoder script. A total of 10,993,648 reads were obtained. Only 10,071 contigs were retained after assembly and removal of singletons and repetitive regions. The contig length median was 3078 bp (N50), and GC content was 35.4%. After the step of clustering and homology search against *H. brasiliensis* draft genome via blastx, the remaining 6,995 contigs were used as reference for mapping and SNP calling. In total, Freebayes detected a total of 59,116 (39,455 transitions and 19,812 transversions, Ts/Tv=1,99) different heterozygous SNP position in sequences using the stringent parameters. Of these SNPs, 41,621 (70.4%) were found in contigs containing predicted open reading frames (ORFs). Regarding different genotypes, 19,118 putative SNPs were found in GT1 and 25,360 in RRIM701. These cultivars shared 14,568 SNPs positions. The results show that it is possible to combine next generation sequencing data combined with high-density SNP detection methods to discover large numbers of putative SNPs in *Hevea brasiliensis*, providing a framework for further population genomic studies to identify the molecular basis underlying phenotypic variation of relevant traits in a non-model species.
Financial Support: FAPESP