Journal of Biotechnology and Biomaterials

ExtendedAbstract Open Access

Long Non-Coding RNome of Soursop (Annonamuricata) Fruit Exocarp: First draft

Domínguez-Rosas, Edmundo; Tafolla-Arellano, Julio-César; Hernández-Oñate, Miguel-Angel; Fernández-Valverde, Selene-Lizbeth and Tiznado-Hernández, Martín-Ernesto

Coordinación de Tecnología de Alimentos de Origen Vegetal. Centro de Investigación en Alimentación y Desarrollo, A.C. Carretera Gustavo Enrique Astiazarán Rosas, México· Email: tiznado@ciad.mx

Abstract

Soursopfruit (*Annonamuricata)* had received considerable interest due to its nutritional, economical, and medicinal properties, yet its commercialization has been limited due to its short post-harvest life.Nonetheless, the increase in the post-harvest shelf lifewill allow fruit exportation to international markets.Numerous studies have been carried out to elucidate the phenomena that control the duration of post-harvest of fruits. Data generated in later years, strongly indicates that the cuticle controls the post-harvest life of the fruits. The fruit cuticle is the outermost layer of the fruit, composed of high molecular weight compounds such as very long chain fatty acids, waxy alcohols and esters, long chain alkanes, ketones as well as terpenes, phenolic acids, among others.Studies with the goal to understand the molecular mechanism of cuticle biosynthesis had shown that it is regulated by a complex gene network. However, the potential role of non coding RNA in the biosynthesis and regulation of the cuticle biosynthesis have not been studied yet, to our knowledge.Several studies had shown the regulatory role in different phenomena of non-coding RNA such as the long noncoding RNAs (lncRNAs). These genetic elements are characterized by having sequences greater than 200 nucleotides (nt) and a null capacity to be translated into proteins. In plants, lncRNAs had been shown to regulate various phenomena such as fruit ripening, flowering, anthocyanin synthesis, and plant response to salt stress. Based on the above, the objective of the present work was to study the possible participation of lncRNAs in the regulation of the molecular mechanism of soursop cuticle biosynthesis.The fresh fruits of soursop used for the RNA-seq profiling, were obtained from an orchard located at Compostela, Nayarit, Mexico. Three fruits were collected at physiological mature (PM), ripe (R) and over-ripe (OR) stage of development. The exocarp tissue of each fruit was excised and stored at -80 ºC to isolate total RNA. Three strand-specific RNA-Seq libraries for each developmental stage, were constructed. These were sequenced in paired-end mode with a read length of 150 bp using an IlluminaNextSeq 500 system obtaining around 80 million pair-end reads per library. Subsequently, the reads were filtered by quality with the software Trimmomatic version 0.36 and were used to assemble the transcripts *de novo* using the software Trinity. With the aim to eliminate exogenous RNAs transcripts, the resulting assembled transcripts were contrasted with the GenBank Nucleotide (nt) database (http://www. ncbi.nlm.nih.gov/genbank/) using the Basic Local Alignment Search Tool (BLAST). Transcript having hits to sequences from viruses, bacteria, and archaea were discarded. After that, rRNA, low-complexity, and polyA/T sequences were removed with the tool SeqClean. With the aim to remove redundancies, the remaining transcripts were clustered using CD-HITS with a minimum of 90% identity. Further, transcripts showing an expression of less than 0.01 transcripts per million and shorter than 200 bp were also removed.The annotation of the transcripts was carried out by querying against the databases of NCBI's non-redundant protein database (nr), SwissProt, RefSeq, KEGG, GO, TF plant with blastx using an E-value of $1e^{-20}$. Further, a functional domain search was made using the InterPro database. Transcripts with functional annotation were considered as coding genes, while the transcripts without significant homology to known proteins were retained and used for identification of putative long non-coding RNAs. With this analysis, out of the 188,614 soursop library transcripts, it was found that 82,181 does not have any hit, and that could be considered as non-coding sequences. Thereafter, the transcripts identified were evaluated for their coding capacity with the software coding potential calculator (CPC) and coding potential calculator 2. (CPC2), and transcripts showing a coding potential score greater than zero were discarded. With this analysis, we filtered out 3,788 transcripts showing coding capacity, remaining 78,393. Subsequently, transcripts with an open reading frame (ORF) longer than 75 aminoacids were also eliminated. Out of the 78,393 remaining transcripts, 10,182 were found to have an ORF of 75 aminoacids and were removed. The remaining transcripts were filtered with kallisto to retain the most abundant isoform. With this analysis, from 68,211 transcripts, we finally get 65,142 transcripts with a large potential to be lncRNAs.With the goal to obtain the final putative lncRNAs, the 65,142 remaining transcripts will be analyzed with the software infernal to query these transcripts against Rfam to find tRNAs, rRNAs, and other non-coding RNAs. Additionally, we are also going to use the software's signalP and tmhmm to detect signal peptides, and transmembrane helices, respectively. Later, with the aim to find the lncRNAs playing a regulatory role in the molecular mechanism of cuticle biosynthesis, we will analyze the differential expression of lncRNAs and mRNAs encoding proteins playing a role in cuticle biosynthesis, to build co-expression networks between the samples at different developmental stages. Finally, the lncRNAs differential expression calculated *in silico*, will be validated by qRT-PCR analysis with the $2^{-\Delta\Delta CT}$ method to calculate the gene expression.