

Next generation transcriptome assembly for *Euterpe oleracea*

Maria Teresa G. Lopes, Ph.D., Univerdade Federal do Amazonas; Ananda V. de Aguiar*, Ph.D., Embrapa Florestas, Fernanda A. Gaiotto, Ph.D., Universidade Estadual de Santa Cruz; Annette Fahrenkrog, Ph.D., University of Florida; Flora Bittencourt, Ph.D., Universidade Estadual de Santa Cruz; Christopher Dervinis, Master, University of Florida; Bárbara S. F. Müller, Master, Universidade Fderal de Brasilia; Rodrigo F. dos Santos, Master, University of Florida; Regina C. Quisen, Ph.D., Embrapa Florestas; Matias Kirst, Ph.D., University of Florida

Euterpe oleracea, Arecaceae family, is used especially for fruit. Its natural habitat is Amazon rainforest. Here we used high throughput RNA sequencing (RNAseq) to obtain the reference transcriptome for this species, establishing a critical genomic resource necessary for future genetics studies. Leaves of one adult individual of *E. oleracea* were collected in the Amazon rainforest (Brazil). They were immediately frozen in liquid nitrogen and lyophilized. Total RNA was isolated and converted into cDNA prior to sequencing using an Illumina NextSeq platform. A total of 102,576,656 raw reads (pair-end read of 151 length) were filtered by quality using Trimmomatic and assembled into 193,487 transcripts with Trinity. The *E. oleracea* novo transcriptome assembly contains 153,805 unigene databases represented by 117 Mbp, with a median (mean) contig length of 334 bp (608 bp) and a GC content of 44.9%. The databases were annotated for their putative functions based on Arabidopsis thaliana transcriptome database resource. Total of 14,207 annotated unigene databases were categorized into 30 functional groups under Gene Ontology terms. In the biological process category, cellular processes (39.09%) and metabolic processes (35.92%) were the predominant groups. For the cellular component category, the predominant were cell parts (51.98%) and organelles (32.6%). The main distributions in the molecular function category were catalytic activity (36.33%) and binding (35.48%). This is a critical resource to be used for the development of new molecular tools for conservation genetics and evolutionary studies.

Take Away Notes:

- The audience can get a greater knowledge of the technology that helps to understand the genetic makeup of a species which is very important to human and animal foods.
- The study presents objectively modern and efficient techniques to reach its goals.
- Taking under consideration the conference theme, many of the participants will have similar objectives in their respective studies.
- The audience can use the results of this work for many purposes including: to comparing the transcriptome and genome other palm species around the world; to obtaining some marks as SNPs; further in depth studying about conservation, evolution and breeding palm species. This will help to obtain SNPs that will be used a variety of studies on genetic themes. RNA sequencing was the strategy to access the functional component, especially since it is a species without a reference genome. In the future, this set of SNPs can be used to design one chip for the palm species. This kind of information about this native's species quite possibly has an even greater significance to Brazilian researchers.

Biography

PhD in Genetics and Plant Breeding at Universidade Federal de Goiás and post-doctorate in forestry genomics at the University of Florida (2016). I am a Senior Researcher in Forestry and Genetic Resources at Brazilian Agricultural Research Corporation (EMBRAPA) and Professor of Forest Breeding and Genetic Conservation of Forest Species in graduate course at Universidade Estadual de São Paulo (UNESP). The area of expertise is focused on conservation, breeding and genomics with emphasis on forest species. The author also serves as a consultant to scientific funding agencies in Brazil and acts as a technical consultant in the pine tree breeding area for forestry companies in Brazil.