Abstract Book

INRA SCIENCE & IMPACT université de BORDEAUX cirad IUFRO

Next-generation transcriptome assembly of an Amazon palm (Euterpe precatoria)

Maria Teresa G. Lopes, Fernanda A. Gaiotto, Ananda V. de Aguiar, Annette Fahrenkrog, Flora Bittencourt, Christopher Dervinis, Bárbara S. F. Müller, Rodrigo F. dos Santos, Regina C. Quisen, Matias Kirst

Euterpe precatoria is a palm species of the Arecaceae family. The species occurs in the northwest and central regions of the Amazon rainforest, and it grows preferably on well-drained and low fertility soils. Euterpe precatoria produces açai fruits that are processed and consumed as smoothie, jelly, juice, candies and ice cream. Additionally, it produces a palm heart that can be consumed. Natural populations are threatened by predatory exploitation and deforestation in the Amazon. With the advent of next-generation sequencing technologies (RNAseq), the genetic diversity of many forest species, as well as their evolutionary processes, can be better understood. Here we describe the generation of a reference transcriptome for E. precatoria using RNAseq, developed to support population and genetic studies. Leaves of one adult individual were collected in the Amazon rainforest (Brazil), and immediately frozen in liquid nitrogen and lyophilized. Total RNA was isolated and cDNA libraries were sequenced with the Illumina NextSeq platform. A total of 95,232,362 raw reads (paired-end reads of 151 bp length) were filtered by quality with Trimmomatic and assembled into 241,205 transcripts with Trinity. The E. precatoria de novo transcriptome assembly contains 201,545 unigenes represented by 86 Mbp, with a median (mean) contig length of 282 bp (359 bp) and a GC content of 44.76%. Unigenes were annotated for their putative functions based on the Arabidopsis thaliana transcriptome database. A total of 12,575 annotated unigenes were categorized into 31 functional groups under Gene Ontology terms. In the biological process category, cellular processes (41.07%) and metabolic processes (37.84%) were the predominant groups. For cellular component category the predominant were cell part (53.64%) and organelle (34.02%). The main distributions in the molecular function category were catalytic activity (38.79%) and binding (36.39%). The E. precatoria reference transcriptome was also analyzed for the identification of simple sequence repeat (SSR) markers. A total of 5,099 SSRs were identified along the transcriptome using 10,4,4,4,4,4 motifs repeats criteria for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides SSRs. Dinucleotide repeats were the most abundant type of repeat, representing 59.15% of the total. Trinucleotide repeats constituted roughly 20.48% of all the SSRs detected. The most common dinucleotide motif was AG/CT-GA/TC, and corresponded to 47.84% of the 3,016 SSRs identified in this category. This transcriptome represents a valuable genomic resource for E. precatoria that will be used for future research on genetic diversity, evolution and breeding for this species.

Poster number : S5.5

Transcriptome analysis of Euterpe edulis and identification of microsatellite markers

Ananda V. de Aguiar, Maria Teresa G. Lopes, Fernanda A. Gaiotto, Flora Bittencourt, Christopher Dervinis, Bárbara S. F. Müller, Rodrigo Furtado dos Santos, Regina C. Quisen, Matias Kirst

Euterpe edulis is a palm species from the Arecaceae family native to the Brazilian Atlantic forest. It is a plant that prefers shady and humid environments, and plays a very important role in forest dynamics. It produces an abundance of fruits that serve as food for many species of wildlife. The palm heart is the main exploited product. The growing demand for that product in Brazil is driving this species to local extinction. The genetic characterization of E. edulis is indispensable to propose management and conservation strategies for the remaining natural population. With the objective of developing genomic resources for this species, total RNA was isolated from leaves of one adult individual of E. edulis, converted into cDNA, and used to prepare sequencing libraries. The Illumina NextSeq platform was used to produce transcriptome sequences. A total of 81,724,584 raw reads (paired-end read of 151 length) were filtered by quality with Trimmomatic and assembled into 288,275 transcripts with Trinity. The E. edulis de novo transcriptome assembly contains 235,419 unigenes represented by 120 Mbp, with a median contig length of 292 bp, mean contig length of 372 bp, and a GC content of 45.21%. Of all unigenes, 8,428 were functionally annotated to one or more Gene Ontology categories based on Arabidopsis thaliana. Two predominant groups in the biological process category were cellular (40.97%) and metabolic processes (37.87%). For the cellular component category the predominant groups were cell structure (52.39%) and organelle (32.57%). The main distributions in the molecular function category were catalytic activity (39.48%) and binding (36.47%). Beyond SNPs, we can also developed microsatellites from transcriptomes. About 12,346 sequences were examined, anda total of

2,724 microsatellites were identified based on 10,6,5,5,4,4 motif repeats criteria, according to the number of sequences that contain mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides. The number of sequences containing more than one microsatellite was 365. Trinucleotide and dinucleotide were more abundant, respectively representing 31.41% and 14.01% of the total SSRs found. The trinucleotide motif AAG/CTT was most represented in this category (23.87%). Potential marker sites identified will be selected, validated, and used to develop EST-SSR primers for this species. Here we present a broad survey of the E. edulis transcriptome, as well as an extensive search for molecular markers that may be applied to analyze the genetic diversity of E. edulis natural populations and germplasm bank.

Poster number : S5.6

Draft Genome and Gene Annotation of Lentinula edodes

Donghwan Shim1, Kwan-Soo Woo1*, Kyung-Hwan Jang1, Sang Urk Han1 and Hojin Ryu2
1 Department of Forest Genetic Resources, National Institute of Forest Science, Suwon 16631, Republic of Korea
2 Department of Biology, Chungbuk National University, Cheongju 28644, Republic of Korea

Lentinula edodes, called the shiitake mushroom, is one of the most important and popular cultivated edible mushroom with its various utilities as foods and medicinal fields. Here, we represent the 46.1 Mb genome of L. edodes, encoding 13,028 predicted gene models. The genome assembly consists of 31 scaffolds and with 0.45% N's (N50, 3.66 Mb; N90, 0.81 Mb, the longest scaffold, 5.85 Mb). Gene annotation provides key information for various signaling pathways and secondary metabolites. This genomic information would contribute to establish the basis for developing molecular genetic markers for MAS/MAB and increasing our understanding of the genomic structure and functions.

Poster number : S5.7

Assessing the utility of genomic selection in Eucalyptus breeding

Ms. Tan - Department of Ecology and Environmental Science, Umeå University, Sweden

Genomic Selection (GS) is a breeding methodology that can increase genetic gains by accelerating the breeding cycle or by improving the selection accuracy. We used 44,250 SNPs scored in a Eucalyptus population comprised of 86 E.grandis, 82 E.urophylla, and their 949 F1 hybrids to develop GS models for four phenotypic traits, basic density, pulp yield, and volume in trees scored at three and six years of age. The model prediction accuracies were optimized by considering the effects of three aspects, (1) statistical algorithms, (2) selection of training set, and (3) SNP subsets. Four statistical algorithms were applied including ridge regression-best linear unbiased prediction (rrBLUP), GBLUP, Bayesian LASSO, and reproducing kernel Hilbert spaces (RKHS). The GS models using rrBLUP, GBLUP and Bayesian LASSO performed similarly well in predicting all the traits. The GS models were found to outperform the conventional pedigree method in predicting traits. We evaluated the impact of four different training sets considering genetic relationship between training and test sets and genetic variance between them. The GS models using the training set that was comprised of all parents and partial F1 progenies gave the best performance in all traits. The average predictive ability across training sets for the rrBLUP algorithm was 0.44 for pulp yield, 0.5 for basic density, 0.18 and 0.32 for volume at 3 and 6 years old, respectively. Furthermore, analyses using subsets of SNPs suggested that using 20k markers is sufficient for GS in our hybrid (E.grandis x E.urophylla) breeding material. Our results suggest that GS model could be applied as a valuable tool for improving Eucalyptus breeding efficiency.

Poster number : S5.8