

POSTER PRESENTATION

Open Access

Comparative transcriptome analysis of tree *Eucalyptus* species using RNAseq technology: analysis of genes interfering in wood quality aspects

MM Salazar^{1*}, LC Nascimento¹, ELO Camargo¹, RO Vidal², J Lepikson-Neto¹, DC Goncalves¹, WL Marques¹, PJSL Teixeira¹, GAG Pereira¹

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery Arraial d'Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

The *Eucalyptus* wood is one of the most important raw materials for pulp and paper industry. Brazil is currently the first producer of short-fiber pulp and sixth in total production of cellulose. To maintain the industrial competitiveness, investment in genomic research started in 2002 with the *GENOLYPTUS* project (Brazilian Network of Eucalyptus Genome Research). Recently, a new transcriptome library was generated using Next Generation RNA Sequencing by Illumina's sequencing by synthesis technology.

Different species of *Eucalyptus* are recognized for their superior characteristics in terms of growth, wood quality and resistance to different types of stress (1). Such features are probably driven by the coordinated expression of numerous genes involved in processes of structural and regulatory genes in xylogenesis. Therefore, the main purpose of this study is to identify genes and key metabolic compounds directly involved in wood quality, as well as transcription factors involved. An extensive data mining in the RNAseq database was conducted to identify sequences over expressed in xylem and those that were differentially expressed between species.

Methods

Genolyptus Sanger sequenced ESTs (167,271) and NCBI *Eucalyptus* ESTs (36,981) were assembled using the program CAP3 (2). All unigenes were automatically

annotated using BLAST (3) (e-value cutoff of 1e-5) against protein databases, including: non-redundant (NR) database, uniref (4), pfam (5) and keg (6). Moreover, a functional annotation using the BLAST2GO software was performed (7). The RNA-Seq reads produced from three different xylem libraries (*Eucalyptus globulus*, *E. grandis* and *E. urophylla*) were aligned against the assembled unigenes using the SOAP2 aligner (8) configured to allow up two mismatches, discard sequences with "N"s and return all optimal alignments. In order to perform the differential expression analysis between libraries, a normalization and statistical pipeline were applied using DEG-seq (9) software considering a 99% confidence rate (cut-off of 0.01). From this analysis we obtained xylem genes and transcription factors differentially expressed between the three species.

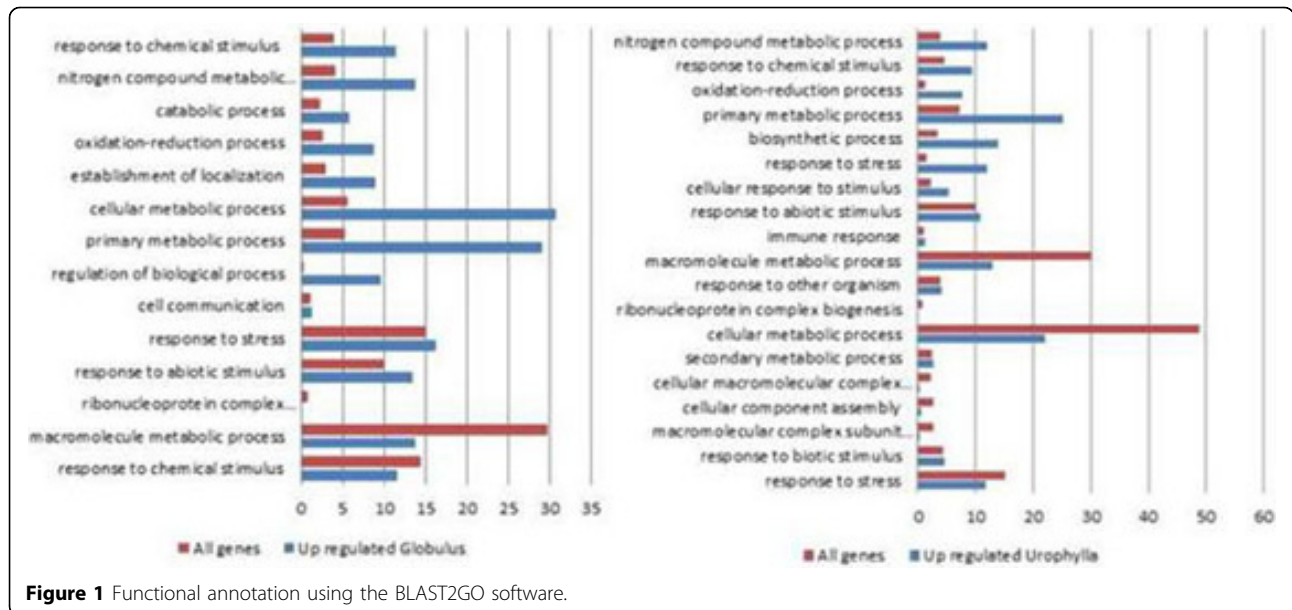
Results and discussion

The assembly produced 53,412 unigenes (18,098 contigs and 35,314 singlets). The xylem libraries produced a large number of RNAseq reads (35bp). About 28 million reads were produced for the *E. globulus* library, 25 million for *E. grandis* and 25 million for *E. urophylla*. About 2% of reads were discarded after filtering. Most part of RNAseq reads mapped into the new EST assembly: 69.27% for *E. globulus*, 71.97% for *E. grandis* and 67.90% for *E. urophylla*. As a result, 33,599 unigenes were aligned to the RNAseq libraries. The functional annotations (Figure 1) show percent of genes related to the most relevant GO categories represented in each of the species pairs studied for Biological Process, level 3.

* Correspondence: marcelamsalazar@gmail.com

¹Laboratório de Genômica e Expressão - LGE- UNICAMP, Brazil

Full list of author information is available at the end of the article



In the *E. globulus* X *E. grandis* comparison, most genes are in the macromolecule metabolic process category that includes genes for pectin, cellulose and hemicellulose metabolism and also transcription factors involved in such pathways. Over 10% of these genes are over-expressed in *E. globulus*. Over 30% of the genes are over-expressed in *E. globulus* in the category metabolic cellular process. In the *E. urophylla* X *E. grandis* comparison, the metabolic cellular process category is representative of the total number of contigs, however, the number of genes over-expressed in *E. urophylla* is much lower. This may be an indicative that genes that participate in such pathways can contribute to the differential wood qualities found in *E. globulus*.

The new assembly, RNAseq libraries and Gbrowse are available at www.lge.ibi.unicamp.br/eucalyptus. *E. globulus* and *E. urophylla* libraries were compared against *E.*

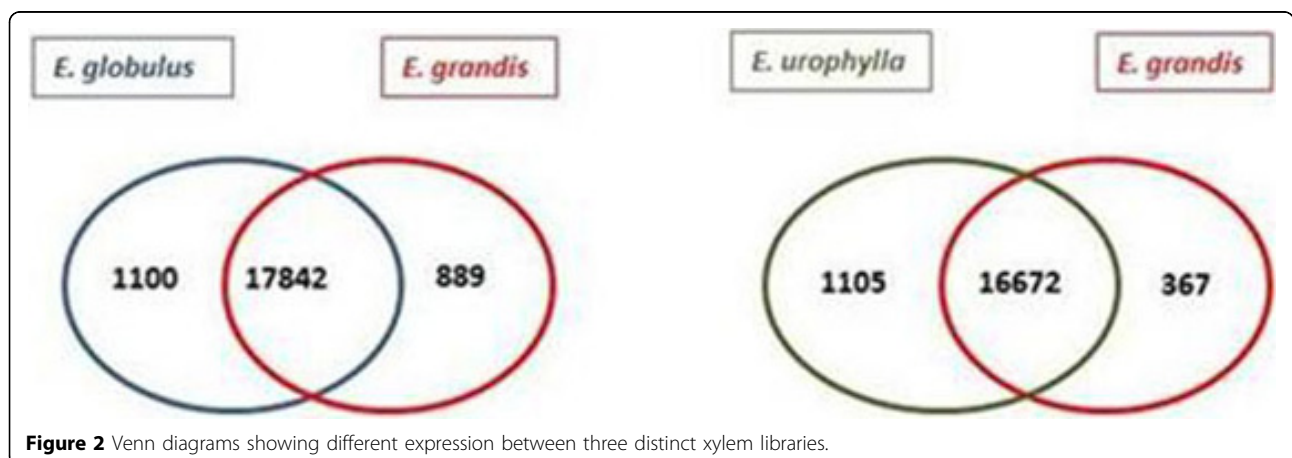
grandis library in order to access differentially expressed genes (considering 99% of confidence rate - cut-off of 0.01). As a result, 19,828 genes were differentially expressed in the *E.gl* X *E. gr* comparison (51.43%) and 18,142 (49.27%) in *E.ur* X *E. gr*. Also in these groups there were genes not expressed in one of the species, as can be seen in Venn diagram below (Figure 2).

These results may contribute to the understanding of wood formation processes and possibly help guide its improvement. The increase in wood quality and productivity has significant economic impacts especially in the pulp and paper industry.

Author details

¹Laboratório de Genômica e Expressão - LGE- UNICAMP, Brazil. ²Laboratório Nacional de Biotecnologias - CNPEM/ABTLuS, Brazil.

Published: 13 September 2011



References

1. Grattapaglia D, Kirst M: **Eucalyptus applied genomics: from gene sequences to breeding tools.** *New Phytologist* 2008, **179**:911-929.
2. Huang X, Madan A: **CAP3: A DNA Sequence Assembly Program.** *Genome Research* 1999, **9**:868-877.
3. Alschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool.** *Journal of Molecular Biology* 1990, **215**:403-410.
4. Suzek BE, Huang H, McGarvey P, Mazumber R, Wu CH: **Uniref: comprehensive and non-redundant UniProt reference clusters.** *Bioinformatics* 2007, **23**(10):1282-1288.
5. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Griffith-Jones S, Howe KL, Marshall M, Sonnhammer ELL: **The Pfam Protein Families Database.** *Nucleic Acids Research* 2002, **30**(1):276-280.
6. Kanehisa M, Goto S: **Kegg: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Research* 2000, **28**(1):27-30.
7. Conesa A, Gotz S, Garcia-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
8. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, Kristiansen K, Wang J: **SOAP2: an improved ultrafast tool for short read alignment.** *Bioinformatics* 2009, **25**(15):1966-1967.
9. Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X, Peng L: **Expression profiling and integrative analysis of the CESA/CSL superfamily in rice.** *BMC Plant Genomic* 2010, **10**(282):1-16.

doi:10.1186/1753-6561-5-S7-P175

Cite this article as: Salazar et al.: Comparative transcriptome analysis of tree *Eucalyptus* species using RNAseq technology: analysis of genes interfering in wood quality aspects. *BMC Proceedings* 2011 5(Suppl 7):P175.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

