Integrated transcriptomic and metabolomic analyzes applied to cane-energy: a new variety of cane with high biomass productivity

Jovanderson Jackson Barbosa da Silva, Luís Guilherme Furlan de Abreu, Nicholas Vinicius Silva, Antônio Pedro de Castello Branco da Rocha Camargo, Camila P. Cunha, Gonçalo Amarante Guimarães Pereira, Marcelo Falsarella Carazzolle

Universidade Estadual de Campinas

Abstract

Energy-cane (CE) is a commercial hybrid originated from the crossing of the same parents as sugarcane, Saccharum officinarum and S. spontaneum, but has very distinct characteristics, such as a high fiber content (13.5% higher), low sucrose content (4.1% lower) and higher productivity (85% higher) compared to commercial sugarcane hybrids (CA). These characteristics make energy-cane a promising plant for the production of first and second generation bioethanol and bioelectricity. Although sugarcane is widely studied from a genomic and molecular biology point of view, very little knowledge has been generated about energy-cane, in particular, in order to expand the development of these new hybrids through the use of genomics and systems biology. In this context, this study has generated transcriptomics (RNA-Seq) and metabolomics data from various tissues of energy-cane and sugarcane during the night cycle (18 hours and 24 hours) and day cycle (6h and 12h). For transcriptomic data, several bioinformatics pipelines were developed to evaluate the best approach for assembling the energy-cane transcriptome using de novo approach or guided by the reference genome of S. spontaneum (one of the parental species). An analysis of the metabolites profile in leaf and culm tissues during the night and daytime variation was also performed. De novo transcriptome assembly was performed by Trinity software and guided assembly was performed by the combination of Hisat2 and StringTie software. In both cases, the transcripts obtained were quantified by the Kallisto software. Separation into coding and non-coding transcripts was performed by the combination of Transdecoder and RNAsamba software. The metabolic profile analysis was performed on a Q-TOF Ultima-API mass spectrometer, with ESI ionization source, coupled with a UPLC Acquity and processed by the Global Natural Products Social Molecular Networking (GNPS) platform. Although reference-genome-guided assembly generated far more transcripts (25, 065) compared to de novo assembly (14, 825), some transcripts were obtained exclusively by de novo assembly representing candidate transcripts specific to energy-cane. From the analysis in GNPS, 246 metabolites in CE and CA were recovered in 6h, of these only 26 were identified, at 12h, 202 compounds, 18 were identified, at 18h, 230 compounds, of these 22 were identified, at 00h, 220 metabolites, being 19 identified, at 6h (24 hours after the first collection), 129 compounds, 14 identified. The integrated analysis of these omics approaches is a fundamental step to improve our understanding of the molecular biology of energy-cane compared to sugarcane, generating new insights for the development of more productive varieties.