

Unearthing Agave secrets: transcriptome analysis of three species suitable for bioenergy production in semiarid regions

Marina Pupke Marone, Fabio Trigo Raya, Lucas Miguel de Carvalho, Maiki Soares de Paula, Sarita Rabelo, Luciano Freschi, Odilon Reny Ribeiro Ferreira da Silva, Piotr Andrzej Mieczkowski, Gonçalo Amarante Guimarães Pereira, Marcelo Falsarella Carazzolle

USP

Abstract

Agaves are plants that present high productivity in dry areas because of their efficient drought resistance mechanisms. Some Agave species are used to produce alcoholic beverages and others to produce fibers, although studies have suggested that some species could be used as feedstock to produce bioenergy in marginal lands. It is the case of *Agave sisalana*, which is used in the production of sisal fibers; Brazil is the main sisal fiber producer and exporter in the world. This process utilizes only 4% of the leaves; currently, the rest of the material is discarded, but could be used for the production of second generation (2G) ethanol because of its high content of cellulose and hemicellulose, among other compounds. In addition to that, Agaves present low lignin content, a compound that affects negatively the hydrolysis process (recalcitrance) necessary to the 2G production. There is no genome available for any Agave species nor many published studies about the genetic features of the species cultivated in Brazil. In this context, transcriptomic analysis is an appropriate strategy to explore the genetics of these plants without a reference genome. We have sequenced the leaf, stem and root transcriptomes of three of the most fiber-producing genotypes (*A. fourcroydes*, *A. sisalana* and 11648 hybrid) collected from a germplasm bank located on a region with low rainfall. We assembled the 3 transcriptomes separately, finding 26, 779, 30, 962 and 28, 320 genes for *A. fourcroydes*, *A. sisalana* and 11648 hybrid, respectively. The subsequent analyses were used to shed light on the cell wall complexity and abiotic stress mechanisms utilized by these three species. Altogether with the differential expression and orthologous analysis, we could observe that although all three species have many features in common, their differences come from the expression level. For example, while *A. fourcroydes* stem presents higher lignification, it possesses a higher expression of the COMT gene, part of the S lignin synthesis pathway, a type of lignin that is less recalcitrant; *A. sisalana* and 11648 hybrid, on the other hand, have a lower expression of COMT, presenting higher recalcitrance. This suggests that some species have lower S lignin content than others and could be more suitable for 2G production. By analyzing the most highly expressed genes and tissue-specific ones, we have found that the majority of them were related to high temperature and drought resistance. All have higher expressions of heat shock proteins, ubiquitins and LEA, a gene related to hydric stress. Again, the three species present similar strategies, yet some expression level differences exist. Also, a comparative genomics analysis between these Agave species and high and low biomass production species showed us that Agaves have many expanded heat shock protein families, confirming that this is an important mechanism to dealing with high temperatures. Besides, there are four exclusive families related

to cell wall synthesis between Agaves and Saccharum spontaneum, a high yield biomass crop. In conclusion, Agaves show several processes with secondary functions responding to stress and all these mechanisms make them extremely adapted to dry areas and still great biomass producers.

Funding: FAPESP, CNPq, CAPES