

Modeling allele-specific expression in complex polyploids

Agnelo Furtado, Antonio Augusto Franco Garcia, Robert James Henry, Gabriel Rodrigues Alves Margarido, Fernando Henrique Correr

Escola Superior de Agricultura "Luiz de Queiroz" / Universidade de São Paulo (ESALQ/USP)

Abstract

Allele-specific expression (ASE) represents the difference in the magnitude of expression between haplotypes of the same gene. Assessing ASE relies on identifying polymorphisms from genotypic data, whose allele expression levels are measured by high-throughput methods. Allelic imbalance occurs if the ratio of expression between two alleles shows deviations from their expected equivalent expression. However, this is not straightforward for polyploids, especially autopolyploids, as knowledge about the dosage of each allele is required for accurate estimation of ASE. This is the case for the genomically complex *Saccharum* species, characterized by high levels of ploidy and aneuploidy. Two species in this genus were the basis for developing sugarcane cultivars, which are interspecific hybrids. We propose a model to test for allelic imbalance in *Saccharum* that can be easily expanded to other polyploids. Our study is the first approach to assess ASE in a complex polyploid system using estimated allelic dosages. First, we identified SNPs and estimated allele dosages in a panel of *Saccharum* and other closely-related accessions. Then, we quantified the expression of each allele using sequenced libraries from leaves of six genotypes. To test for ASE in the i -th SNP of the k -th genotype, our null hypothesis was that the proportion of the reference allele from RNA counts (π_{ik}) was equal to ratio of the dosage of this allele (P_{ik}) in the genome. Our model followed a Beta-Binomial distribution in which the a priori distribution of π_{ik} was modeled by a Beta distribution using as parameters the dosage of each allele. To obtain the a posteriori distribution we used the Bayesian Markov chain Monte Carlo procedure, calling an ASE SNP if P_{ik} was outside the high density interval of π_{ik} . We found that genes showing ASE were common in *Saccharum*, with highest frequencies in sugarcane hybrids. Genes with ASE were related to a broad range of processes, mostly associated to the general metabolism, organelles, responses to stress and responses to stimuli. Although many processes were specifically associated to particular genotypes, we found that conserved Liliopsida orthologs were significantly enriched with genes showing ASE. However, there was no significant enrichment among orthologs of *Saccharinae* or *Saccharum*. We then hypothesize that monocot core genes show ASE to preserve essential functions. These results provide evidence of the ASE importance in the evolution of *Saccharum*, justifying the maintenance of higher expression levels of some beneficial alleles.

Funding:

Link to Video: