



Phylogenomic analysis of transcription factor families and the evolution of nitrogen fixation in legumes

Kanhu Charan Moharana, Thiago Motta Venancio

Transcription factors (TF) are essential for proper plant growth and development. Legumes, particularly soybean, are remarkable for their ability to form symbiotic associations with nitrogen-fixing bacteria. Here we report the evolutionary analysis of major TF families in legumes and their potential association with the emergence of nitrogen fixation in this group. We gave particular attention to the two rounds of whole-genome duplications (WGD; ~58 *mya* and ~13 *mya*) in the soybean lineage. In total, we analyzed the genomes of 15 legume species and 5 other plant species. Known TF DNA binding domains were used to systematically identify TFs. We estimated changes in gene family sizes in various ancestor nodes and found a strong correlation with WGD events. Later nodes with significantly rapid changes ($P < 0.05$) in gene family sizes were integrated with gene expression data from various species. We identified a total of 35,271 TFs from 58 families. The percentage of TF ranged from 3-8% of the gene complements. Bursts of TF expansions were detected at ages that correspond to that of known WGD events. We observed the maximum number of expansions in the common ancestor of wild and cultivated soybeans, which can be involved in important soybean traits. As compared to wild soybean, 38 cultivated soybean TF families were significantly expanded. Several genes from these families had seed and pod specific expression and were found within seed specific quantitative trait loci (QTL) like weight, bean size, oils. Such genes might have retained in multiple copies in cultivated soybean, during domestication process, while preferentially lost in the wild counterpart. Similarly, the last common ancestor of legumes showed a significant expansion of TFs. By analyzing the families expanded at the legume-specific WGD, we found that only 14 of the 58 families significantly expanded in cultivated soybean, including ERF and bHLH genes that are preferentially expressed in root tissues, which are possibly related with nitrogen fixation. Our findings will help elucidate aspects pertaining to soybean domestication and soybean-symbiont associations.

Keywords: transcription factors, phylogenomics, legume evolution.

Institution of support: CAPES, FAPERJ, UENF