

# Impact of non-synonymous mutations in adaptive diversification and domestication of soybean

Kanhu Charan Moharana<sup>1</sup>, Thiago Motta Venancio<sup>1</sup>

<sup>1</sup>*Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, Rio de Janeiro, Brazil*

Identification of adaptive mutations is of great interest in understanding and improving desirable phenotypes in plant genetic improvement programs. However, most DNA mutations are generally neutral or slightly deleterious. Recently genome re-sequencing and identification of many trait-associated single nucleotide polymorphisms (SNPs) in wild (*Glycine soja*) and modern (*Glycine max*) soybean cultivars have been reported. However, the adaptive potential of such SNPs remain poorly studied. In the present work we used the publicly available interspecies-specific SNP data to predict mutations with potentially beneficial impact on protein structures. In a bottom-up approach we used *G. max* reference genome (Glyma1.0) to study the impact of SNPs located within quantitative trait loci. We identified 1841 unique non-synonymous SNPs along 1212 *Glycine max* genes. The effect of such substitutions depends on the extent to which it affects the protein structure. Hence a protein structure based approach was used to estimate the free energy difference ( $\Delta\Delta G = \Delta G_{G,max} - \Delta G_{G,soja}$ ) caused by each amino acid substitution. Using homology modeling we created theoretical protein structures encoded by *G. max* genes. Only 247 encoded proteins (20% of 1212) showed significant hit in protein-BLAST search against the PDB database. We used the *FoldX* software to simulate the impact of the non-synonymous mutations in the candidate proteins. We mutated amino acid residues with that present on wild-soybean and measured the change in free energy. While 58 mutations were observed to have destabilization ( $\Delta\Delta G \geq 1\text{kcal/mol}$ ) effect, 47 mutations were observed further enhancing the protein structure stability ( $\Delta\Delta G \leq -1\text{kcal/mol}$ ) in modern-soybean. Annotations based on *Arabidopsis* orthology indicated that these mutations are involved in various biotic and abiotic stress resistances. For example a starch synthase gene has undergone mutations that made it more stable in cultivated soybeans. Several genes involved in disease resistance also displayed differential stability between wild and cultivated soybeans. Although preliminary, our results highlight the importance of deeper computational analyses of GWAS and next-generation sequencing data in identifying the genetic basis of agronomically relevant phenotypes.

**Funding:** The authors acknowledge Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil and the following Brazilian funding agencies for supporting their research: Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).