

Genetic Analysis of Anti -Nutritional Factors in Soybean

Noberto Jose Palange

Abstract

Soybean (*Glycine max*) is an important crop for human food and animal feed because of its high protein and oil content. However, anti-nutritional factors (ANFs) can reduce nutrient utilization and limit the nutritional value of soybean-based foods and feeds. This study investigated the genetic variability, high-throughput phenotyping, genetic associations, and genomic prediction of phytate and TTI in soybean germplasm. A set of 308 soybean genotypes was evaluated for phytate and TTI using standard spectrophotometric methods. Significant phenotypic variation was observed for both traits ($p < 0.001$). Phytate concentrations ranged from 0.02 to 6.92 mg g⁻¹, whereas TTI concentrations ranged from 0.02 to 1.53 mg g⁻¹.

In the second study, near-infrared spectroscopy (NIRS) combined with machine learning (NIRS-ML) was evaluated for rapid prediction of phytate and TTI. A randomly selected subset of 190 genotypes previously characterized using wet-chemistry methods was scanned over the 400–2500 nm spectral range. Partial least squares and Random Forest (RF) regression models were evaluated for both traits. RF outperformed PLS under the evaluated modelling conditions, achieving R² values of 0.97 for phytate and 0.96 for TTI.

In the third study, the genetic architecture of phytate and TTI was investigated using genome-wide association analysis of 11,805 quality-controlled SNP markers. Multiple complementary GWAS approaches identified marker–trait associations across several soybean chromosomes for both traits. Candidate genes located within genomic regions surrounding significant markers were subsequently identified through gene annotation.

In the fourth study, genomic prediction of phytate and TTI was evaluated using G-BLUP, RR-BLUP, Bayesian methods, LASSO, Elastic Net, and RF under different training population sizes and marker densities. RF showed the most consistent performance across the two traits, although some linear and Bayesian models performed better for phytate. Increasing training population size or marker density did not consistently improve prediction performance.

The findings provide baseline information for the further validation of NIRS-ML models, marker–trait associations, and genomic prediction strategies in independent and genetically diverse soybean populations.