

ABSTRACT

Soybean constitutes about 40% protein, 20% oil; a highly nutritious legume that has enormous potential to improve dietary quality for people throughout the world. To understand the genetic diversity and population structure of tropical soybean germplasm, 89 genotypes from diverse sources were analyzed using 7,962 SNP markers. The results showed low genetic diversity among the studied germplasm and the polymorphism information content (PIC) was 0.27. The phylogenetic tree and Principal Component Analysis (PCA) both showed that the 89 soybean genotypes were grouped into three major clusters; while population structure grouped the genotypes into two major subpopulations. On the other hand, the average Roger genetic distance within the study population was 0.34.

The variability of different nutritional traits of 52 tropical soybean core collections from diverse sources, showed that traits varied significantly depending on genotypes and country of origin.

Total protein content ranged from 35.07% to 50.4% and genotypes Sline 5.18, BSPS 48A-8 and BPS 48A-27-1 had the highest protein content of 50.40%, 48.88% and 48.08% respectively. On the other hand, NIIXGC 17.3 and Nam II had the lowest protein content of 30.07% and 35.57% respectively. Total oil content ranged from 14.94% to 23.48% where genotypes G32B, Roan and AGS 338 significantly had the highest oil content among the others while Signal, Maksoy 5N and Sline 16.2 had the lowest. The relative percentage of major fatty acids ranged from 10.58% to 21.18% for palmitic acid (16:0), 4.93% to 16.76% for stearic acid (18:0), 22.69% to 39.95% for oleic acid (18:1), and 30.60% to 51.72% for linoleic acid (18:2). Total oil content varied significantly between origins with genotypes from Seed Co having the highest mean of 20.13% while those from AVRDC had the lowest mean of 18.32%. Genotypes from Uganda had the highest percentages of oleic acid; followed by genotypes from Japan and AVRDC.

The GWAS based on 92 soybean genotypes revealed two significant associations ($-\log[P\text{-value}] > 5$) with oil content for two SNPs, rs2291820 and rs22918919 on chromosome 7 and 10 respectively. A significant association ($-\log[P\text{-value}] > 2.5$) with protein content was detected for 3 SNPs, rs 22918920, rs 22918919 and rs 1494480 located on chromosomes 7, 10 and 20

respectively. The study found that both rs2291820 and rs22918919 located on chromosomes 7 and 10 were associated with both oil and protein content.

Genotype NII X GC 20.3 had the highest mean protein content of 43.0%, and BSPS 48A-9-2 and BSPS 48A-28 were superior for mean grain yield ($1,207 \text{ kg ha}^{-1}$). Bulindi was the most discriminating and representative test environment for soybean yield. A weak and negative correlation ($r = -0.1^{**}$, $df = 29$) was detected between protein content (%) and yield (Kg Ha^{-1}).

The current study identified soybean genotypes that can be used to improve the nutritional traits of soybean in Uganda and across the East African region. The study also reported a low diversity in the studied germplasm pool that can lead to genetic erosion of the existing germplasm pool. Therefore addressing these challenges and developing soybean varieties with the desirable traits, requires diversification of the genetic background of the current germplasm pool by incorporating new genetic backgrounds from other countries. Furthermore, the study identified SNP markers that are associated with both total protein and oil content that will hasten the process for the development of soybean varieties with improved nutritional traits in Uganda and across the East African region. The highest yielding and stable genotypes BPS 48A-9-2, BPS 48A-31 and Nam II \times GC 44.2 are recommended for further evaluation under farmers' production conditions for selection and release as new soybean varieties in Uganda.