

High Throughput Phenotyping and Genomics Assisted Breeding for Low Hydrogen Cyanide in Cassava

Michael Kanaabi

Abstract

Cassava has diverse uses for food, feed and industry. However, the utility of the crop is restricted by the presence of hydrogen cyanide (HCN) in the roots. The HCN not only makes roots bitter to the taste, but can also be poisonous to fatal levels. Limiting societal exposure to high HCN cassava by breeding for low HCN varieties is the most sustainable way of limiting HCN dietary toxicity. However, progress in breeding low HCN cassava has been slow, partly due to the difficulty in HCN phenotyping. This study aimed to contribute to accelerated breeding of low HCN cassava varieties by developing faster, accurate and reproducible HCN prediction methods based on Near Infrared Spectroscopy (NIRS) and molecular markers to inform selection decisions. Multi-location field trials were established and leaf samples picked three months after planting and shipped to a genotyping facility for genotyping with kompetitive allele specific polymerase chain reaction (KASP) markers and for genotyping by sequencing (GBS). Roots were harvested 12 months after planting for HCN quantification and collection of NIRS spectra. Highly significant differences were observed among clones ($p < 0.001$) and locations ($p < 0.001$) with significant clone by environment interactions ranging from ($p < 0.05$) to ($p < 0.001$). Locations Arua and Serere were associated with high HCN phenotypic expression while Namulonge and Tororo were associated with low HCN phenotypic expression. Three KASP markers snpME00404, snpME00405 and snpME00406 showed significant co-segregation of genotypes with HCN phenotype, explaining 27%, 17% and 14% of the HCN phenotypic variation respectively. Quantitative NIRS predictions with modified partial least squares (MPLS) algorithm had low accuracy ($R^2_p = 0.38$). However, binary classification (low or high HCN) with machine learning algorithms achieved up to 99% accuracy with partial least squares discriminant analysis (PLS-DA) using the full spectral range (400 – 2500 nm). The wavelengths 961, 1165, 1403–1505, 1913–1981, and 2491 nm were influential in discrimination of low and high HCN accessions and achieved up to 100% classification accuracy with Logistic Regression (LR) algorithm. The accuracy of genomic prediction models was low to moderate ($r = 0.18 - 0.52$), with Reproducing Kernel Hilbert Spaces (RKHS) and Bayesian Lasso (BL) being the most accurate ($r = 0.52$). All Bayesian models evaluated, with the exception Bayesian Ridge Regression (BRR) were more accurate than the Genomic Best Unbiased Linear Predictions (GBLUP) model ($r = 0.22$). Collectively, these results offer frameworks for accelerated low HCN cassava breeding in Uganda. The NIRS and KASP markers can be used in routine phenotyping and selection in early generation trials while the genomic selection models RKHS or GBLUP can be deployed in cyclic population improvement for low HCN.