

Understanding sorghum race level diversity and development of sorghum genomic resources by using deep learning-based variant calling approach

Abhishek Rathore^{1,2} (abhishek.rathore@cgiar.org), Pradeep Ruperao³, Nepolean Thirunavukkarasu⁴, Roma Rani Das⁵, Baloua Nebie⁶, Harish Gandhi⁷

¹ CIMMYT, Medak, Telangana, India ; ² Excellence in Breeding, International Maize and Wheat Improvement Center (CIMMYT), Hyderabad, Telangana, India ; ³ ICRISAT, Hyderabad, Telangana, India ; ⁴ ICAR-IIMR, Hyderabad, Telangana, India ; ⁵ International Maize and Wheat Improvement Center (CIMMYT), Hyderabad, Telangana, India ; ⁶ International Maize and Wheat Improvement Center (CIMMYT), Dakar, Senegal ; ⁷ International Maize and Wheat Improvement Center (CIMMYT), Nairobi, India

Understanding sorghum race level diversity and development of sorghum genomic resources by using deep learning-based variant calling approach

Evolutionary divergence in sorghum race populations led to significant genetic and morphological variation. This study presents a k-mer-based sorghum race sequence comparison to discover the conserved k-mers of several accessions from sorghum. We further identified race-specific genetic signatures and the gene variability in 10,321 genes (PAVs). A deep learning-based variant calling methodology was used in a set of genotypic data derived from a diverse panel of 272 sorghum accessions with an objective to understand sorghum race structure, its diversity and domestication process. This analysis resulted in 1.7 million very high-quality genome-wide SNPs.

We further identified selective positive and negative signature regions through a deep genome-wide scan with different statistical methods including iHS and XP-EHH. This work identified a total of 2,370 genes associated with selection signatures which also included 179 selective sweep regions. These genes were randomly distributed over ten chromosomes.

Co-localization of these regions undergoing selective pressure with previously reported QTLs and genes revealed that the signatures of selection could be related to the domestication of important agronomic traits such as biomass and plant height.

This study and identified k-mer signatures will be extremely useful in future plant breeding programs aiming identifying the sorghum race and for trait and SNP markers.