Cassava improvement in the era of "agrigenomics": the road to next-generation breeding

Ismail Yusuf Rabbi (i.rabbi@cgiar.org), Melaku Gedil, Morag Ferguson, and Peter Kulakow

In the last 45 years, IITA has played a pivotal role in the genetic improvement of cassava for resource-poor farmers in sub-Saharan Africa (SSA). More than 400 varieties have been developed that are not only high yielding but also resistant to diseases and pests. Many of these improved varieties have been extensively deployed in SSA and have helped to avert humanitarian crises caused by the viral disease pandemics that devastated local landraces in East and Central Africa.

The cassava breeding program in Ibadan has a collection of more than 750 elite cassava clones representing current and historical materials accumulated over the last 45 years. These materials, referred to as the genetic gain collection (GGC), are accompanied by extensive field evaluation (phenotypic) data. In addition, the active breeding collection contains over 1000 African landraces and more than 400 new advanced breeding clones that are also accompanied by phenotypic data, including observations of disease and pest resistance, plant architecture, flowering ability, and performance in storage root yield.

Pro-vitamin A 'yellow root' cassava developed by the IITA cassava breeding program. Photo by IITA.

I. Rabbi, Postdoctoral Fellow (Molecular Genetics); M. Gedil, Head, Bioscience Center, IITA, Ibadan, Nigeria; M. Ferguson, Molecular Geneticist, IITA, Nairobi, Kenya; and P. Kulakow, Cassava Breeder, IITA, Ibadan, Nigeria.

The most recent success of the conventional cassava breeding program culminated in the release of three vitamin A cassava varieties by the Government of Nigeria. These varieties (IITA TMS I011368, IITA TMS I011371, and IITA TMS I011412) were first cloned from seedlings in Ibadan in 2001 and have been subjected to extensive field testing throughout Nigeria. While almost all cassava in Nigeria are currently white fleshed, vitamin A cassava produces yellow-fleshed roots with nutritionally significant concentrations of carotenoids that produce vitamin A in the human body when consumed as yellow *gari* or *fufu*. In cooperation with HarvestPlus, IITA and partners will distribute vitamin A cassava planting materials to more than 25,000 farmers in 2013. New yellow-fleshed genotypes in the pipeline promise continued improvement in pro-vitamin A content, yield, and dry matter in the coming years.

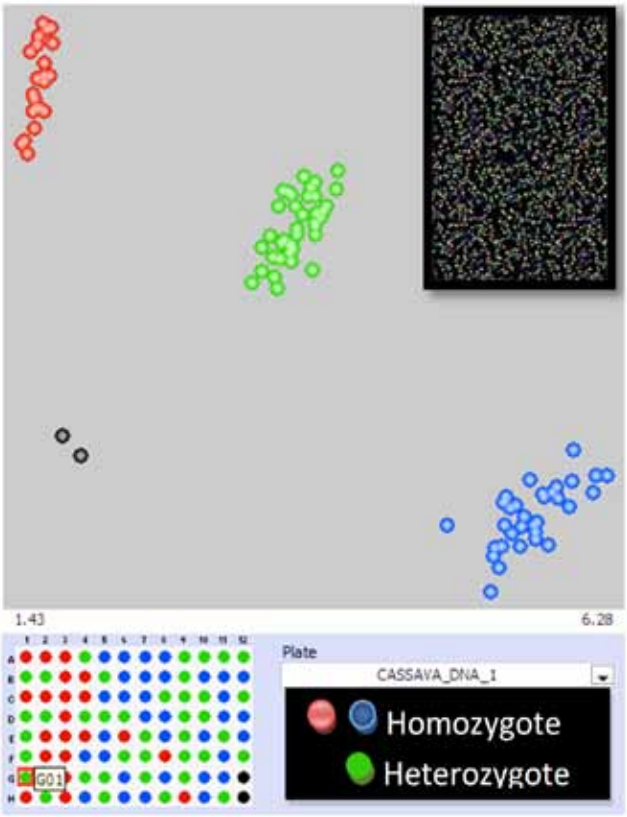
As the vitamin A cassava illustrates, the genetic improvement of cassava has mostly been achieved through conventional breeding methods based on phenotypic selection. The only known direct application of molecular markers in cassava breeding is selection for resistance to cassava mosaic disease and cassava green mite. Recent advances and a reduction in the cost of the next-generation sequencing technologies now promise to usher in a new era for cassava breeding that will combine the success of conventional hybridization, selection, and multilocational yield trials with the latest advances in genomic resources.

Preparation of cassava DNA for genotyping by sequencing. Photo by IITA.

Setting the stage for "next-generation cassava breeding"

Cognizant of the potential of marker technologies to improve the efficiency and effectiveness of cassava breeding, IITA, in collaboration with partners, embarked on the development and deployment of molecular markers[1]. With the recent accumulation of genomic resources in cassava research, including the first full cassava genome sequence[2], our emphasis at IITA has shifted towards the application of these resources in molecular breeding[3].

One recent achievement is the identification and validation of nearly 1500 single nucleotide polymorphism (SNP) markers through an international collaboration led by IITA's geneticist, Morag Ferguson[4]. These SNPs have been converted to a highly parallel hybridization-based genotyping system that has been shared with the international cassava research community through partnership with the Generation Challenge Program (GCP).



An example of an SNP genotyping data plotted with KBioscience's SNPviewer software. Inset: *raw SNP genotyping data from Illumina's GoldenGate®assay.*

In addition, the first SNP-based genetic linkage map of cassava has been developed by IITA in collaboration with Heneriko Kulembeka of the Agricultural Research Institute (ARI), Ukiriguru, Tanzania. A linkage map is analogous to landmarks (SNP markers in this case) placed along chromosomes that guide researchers to genes or genomic regions controlling traits of interest. Such a linkage map is an indispensable tool for marker-assisted selection (MAS). SNP and SSR markers have also been applied to uncover quantitative trait loci (QTL) associated with resistance to cassava brown streak disease (CBSD)—which is ravaging cassava

production in Eastern and Southern Africa—in a collaboration between IITA, CIAT, and ARI-Tanzania.

Another dramatic development in cassava genomics is the recently completed sequencing of the cassava genome through the partnership of the US Department of Energy's Joint Genome Institute and 454 Life Sciences[2].

Genotyping-by-sequencing

The progress in next-generation technologies has drastically reduced the costs of DNA sequencing so that genotyping-by-sequencing (GBS) is now feasible for species such as cassava, ushering in a new era of agricultural genomics[5]. This will revolutionize the application of genomic tools for cassava improvement. GBS involves the cutting of genomic DNA into short pieces at specific locations using a restriction enzyme. The ends of these pieces are sequenced using techniques that allow sequencing of many samples at the same time. The beauty of this method is the use of adaptors containing barcodes (unique tags) that are enzymatically joined to the digested DNA fragments, enabling simultaneous sequencing or multiplexing of up to 384 samples in one sequencing reaction. This economy of scale greatly reduces the cost of processing each individual DNA to less than $10/sample. Approximately 200,000 markers can be identified and mapped in a very short time. With this powerful tool, breeders may conduct genomics-based research that was inconceivable a couple of years ago. Some of the exciting new research applications include polymorphism discovery, high-density genotyping for QTL detection and fine mapping, genome-wide association studies, genomic selection, improving reference genome assembly, and kinship estimation.

High-density QTL mapping and fine mapping

In the past, a limitation for QTL mapping was the number of markers on a genetic linkage map. With new SNP-based technologies this is no longer a limitation. This allows for fine mapping of QTLs so long as a sufficient number of individuals in the mapping population can be developed. IITA, in collaboration with national partners [ARI-Tanzania and National Crops Resources Research Institute (NaCRRI), Uganda], is using SNPs to discover QTLs associated with sources of tolerance for CBSD.

The next frontier for cassava genomics

Using the genotyping by sequencing approach, scientists from IITA and Cornell University, USA, are currently genotyping more than 2000 accessions of cassava, including released varieties, advanced breeding lines, and landraces from Africa. This is a pilot study of genomic selection funded by the Bill & Melinda Gates Foundation to explore the potential for using the IITA breeding collection, including genetic gain, local germplasm, and current advanced breeding lines, as the base population to begin genomic selection for West Africa. The IITA breeding collection has been extensively characterized in many locations and over many years.

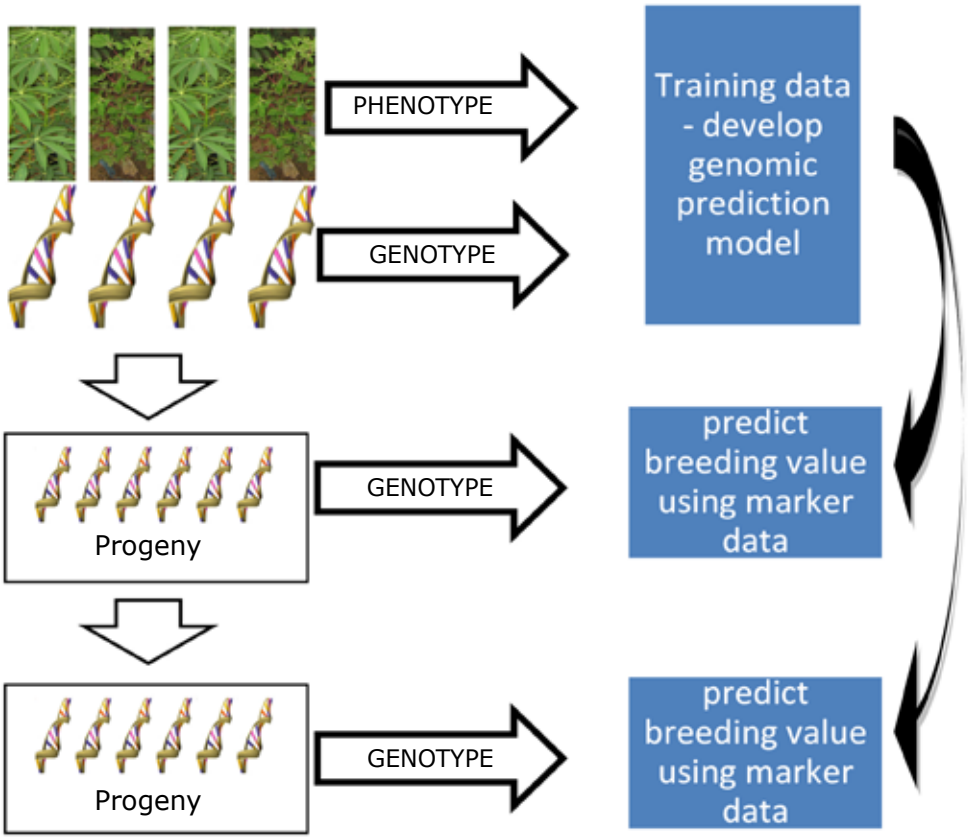
The convergence of high-density SNP data and extensive phenotypic data in

IITA's cassava collection sets the stage for the implementation of genome-wide association studies (GWAS) and genomic selection (GS) in breeding. The aim of GWAS is to pinpoint the genetic polymorphisms underlying agriculturally important traits. In GWAS, the whole genome is scanned for significant marker-trait associations, using a sample of individuals from the germplasm collections, such as a breeder's collection. This approach of "allele mining" overcomes the limitations of traditional gene mapping by (a) providing higher resolution, (b) uncovering more genetic variants from broad germplasm, and most importantly, (c) creating the possibility of exploiting historical phenotypic data for future advances in breeding cassava.

GS is a breeding strategy that seeks to predict phenotypes from high-density genotypic data alone, using a statistical model based on both phenotypic and genotypic information from a "training population". For cassava, phenotyping is the slowest and most expensive phase of the crop's breeding cycle because of the crop's low multiplication ratio of between 5 and 10 cuttings/plant. Thus, it takes several cycles of propagation (up to 6 years) to carry out a proper multilocational field trial evaluation. The implementation of GS at the seedling stage should: (a) dramatically reduce the length of the breeding cycle, (b) increase the number/unit time of crosses and selections, and (c) increase the number of seedlings that could be accurately evaluated. The reduced breeding cycle means that the "engine of evolution," i.e., recombination and selection, can proceed at a rate that is three times as fast as phenotypic-based selection, while saving resources.

Preparation of gari, the most popular food product from cassava, made from yellow root cassava. Photo by IITA.

In conclusion, cassava breeding in IITA is being redefined, thanks to the increasing availability and deployment of genomic resources. Combining these resources with IITA's long-standing conventional breeding pipeline means that the best days of cassava improvement lie ahead. These efforts will ultimately satisfy the increasing need for more healthy and nutritious food produced in environmentally sustainable ways.



A schema of genomic selection (GS) processes, starting from phenotyping and genotyping of the training population and selection of parental candidates via genomic estimated breeding value (GEBV)–based selection. Note that selection model improvement can be performed iteratively as new phenotype and marker data accumulate.

References

[1]Lokko et al. 2007. Cassava. In: Kole et al (ed). Genome mapping and molecular breeding in plants, Vol. 3. Pulses, Sugar and Tuber Crops. Springer-Verlag Berlin Heidelberg.

[2]Prochnik S., P.R. Marri,B. Desany, P.D. Rabinowicz, et al. 2011. The cassava genome: current progress, future directions. Tropical Plant Biology. doi:10.1007/s12042-011-9088-z.

[3]Ferguson M., I.Y. Rabbi, D-J. Kim, M. Gedil, L.A.B. Lopez-Lavalle, and E. Okogbenin. 2011a. Molecular markers and their application to cassava breeding: past, present and future. Tropical Plant Biology. DOI 10.1007/s12042-011-9087-0.

[4]Ferguson M.E., S.J. Hearne, T.J. Close, S. Wanamaker, W.A. Moskal, C.D. Town, J. de Young, P.R. Marri, I.Y. Rabbi, and E.P. de Villiers. 2011b. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theoretical and Applied Genetics. DOI: 10.1007/s00122-011-1739-9.

[5]Elshire R., J. Glaubitz, Q. Sun, J. Poland, and K. Kawamoto. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.