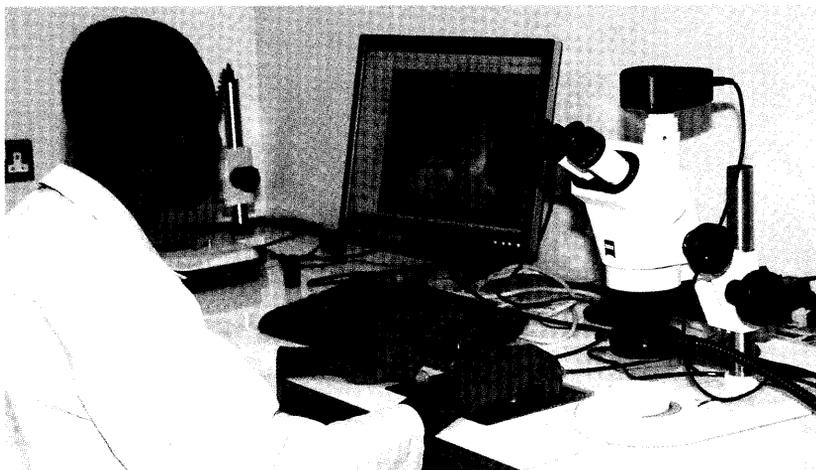


New genomic tools for improving African crops

Melaku Gedil, m.gedil@cgiar.org



Digital imaging and microscopy: a tool for research and training. Photo by O. Adebayo, IITA

The increase in genomic techniques in the past few decades has thrown the doors of research wide open to agricultural scientists. Conventional breeding has been augmented by various innovative molecular marker-aided techniques. The first wave of molecular marker technology introduced biochemical markers (isozymes and allozymes).

These quickly gave way to the first generation DNA-level markers such as Restriction Fragment Length Polymorphism (RFLP, DNA analysis),

Randomly Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), and simple sequence repeat (SSR)—all mouthfuls to the layperson. Those that lend themselves well to automation and multiplexing (use of simultaneous or more than one set of primers in the reaction mix) prevailed because of their cost-effectiveness.

Advances in sequencing technology enhanced the use of DNA sequence-based markers such as SSR and single nucleotide polymorphism (SNP),

allowing the development of automated, high throughput (output) genotyping platforms. In a decade, the cost of genotyping has dramatically declined with various techniques developed that allow flexibility under different circumstances. This emphasized the feasibility of molecular breeding.

New tools

Some of the new genetic engineering tools used at IITA include molecular markers for marker-assisted breeding, resistance gene analogs (RGA), Targeting

Induced Local Lesions In Genomes (Tilling), DNA chips, application of DArT markers, and bioinformatics.

In IITA, the development of new genomic tools for molecular breeding and gene discovery is under way for the mandate crops. For instance, new markers have been identified, *in silico* (online), from cassava Expressed Sequence Tags and hundreds of markers validated using a diverse panel of cultivated cassava varieties. After filtering with various criteria, over a hundred new markers were developed, useful for fingerprinting and other molecular genetic applications.

The rapid accumulation of genome sequence data led to the development of an array of functional genomics tools that are being used to understand the complex pathways involved in host plant–pathogen interaction. The RGA technique has applications in cloning, profiling, and host–pathogen interaction.

The RGA technique was used in IITA to assess DNA sequence variation in several elite cassava clones, resulting in several novel sequences, some of which were found to be similar to previously reported RGAs. This information is expected to facilitate the identification of gene-targeted markers

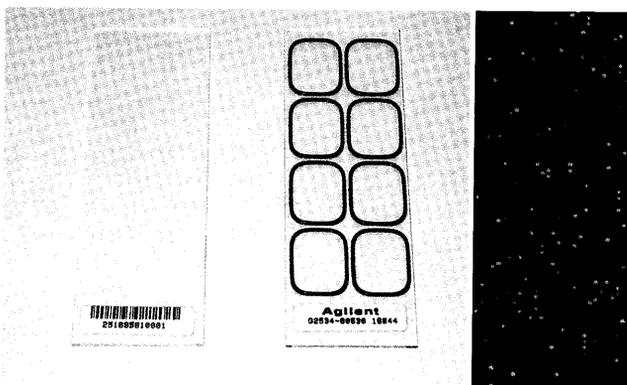
for molecular breeding and gene discovery in cassava.

Another new tool is Tilling, a popular technique of reverse genetics for detecting mutations in a target gene, followed by the assignment of phenotypes to the gene sequence. It rapidly gained popularity because it is suitable for automation and for screening thousands of samples. Besides being a non-GMO approach for broadening the genetic base, it provides tools for developing markers for marker-assisted breeding for traits that are cumbersome and expensive to measure.

Tilling work to discover induced and natural mutation in cassava was geared towards specific traits that are intractable (or not easily managed or manipulated) using conventional methods. Adaptation of the technique to other IITA mandate crops such as

yam, banana, and cowpea entails the selection of target tissue or organ for mutation, and the selection of similar or different target genes. Crops such as maize and soybean have numerous germplasm resources that can be easily adopted and adapted.

Knowledge of the nucleotide sequences of the target genes is a prerequisite for Tilling. The major IITA mandate crops—cassava, yam, and banana—have very limited genomic resources. To date, nucleotide sequence information for a very few, largely chloroplast, genes could be found in Entrez Gene (<http://www.ncbi.nlm.nih.gov/>). Investigations in the past decade resulted in the cloning and characterization of expressed cassava genes involved in starch, cyanogen glucosides, and carotenoid biosynthesis. We have used about six full gene sequences (both



A genome-wide 14K DNA chip for cassava (left) and a scan showing 14,000 different genes (right). Photo by IITA.

nuclear and chloroplast genes) and three promoter sequences of cassava for designing primers. However, even in the absence of a nucleotide sequence for the gene of interest, comparative genomics has been successfully used to identify candidate genes. The completion of the genome sequence of poplar and, more recently, of castor bean, is expected to provide useful genetic tools for identifying candidate genes in cassava. Besides, the ongoing cassava genome sequencing (<http://www.jgi.doe.gov/sequencing/why/CSP2007/cassava.html>) is anticipated to be completed soon, opening a new avenue of research in functional postgenomic studies such as Tilling.

DNA chips have also become popular tools for gene discovery and

also for diagnostics. They also provide a reverse genetics tool for identifying gene-targeted markers for molecular breeding. A genome-wide DNA microarray for cassava with ~14,000 probes has been developed at IITA. This is the most comprehensive DNA chip for cassava available to date. This microarray has been used for transcriptome analysis of cassava. Candidate genes that are differentially expressed after virus infection have also been identified.

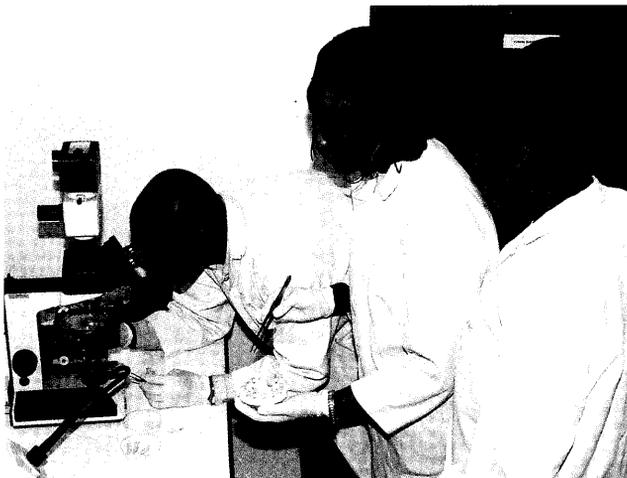
A cassava DArT chip with 735 polymorphic markers was used to fingerprint a diverse cassava population comprising genotypes from Africa, Latin America, Asia, and breeder lines maintained at IITA. Overall reproducibility of the

marker set was very high and average call rate was 97%. DArT markers provide reliable and high throughput molecular information for managing biodiversity in germplasm collections and make rapid genome profiles possible for quantitative trait loci (QTL) mapping.

Bioinformatics

Advances in bioscience technologies such as sequencing, synthesis, imaging, and various other nanoscale assays, have dramatically increased the volume of biological data, which in turn, started the concurrent growth of bioinformatics tools. Bioinformatics is broadly defined as the application of computer technology to the storage, retrieval, and analysis of large amounts of biological information.

The major areas of high-end bioinformatics include the development of databases and algorithms for analyzing and annotating various types of microarray platforms, high-density oligonucleotide chips, variety of mass spectrometry, and diverse platforms of new-generation sequencing data. However, the majority of life science scientists and investigators tend to turn to the Internet to seek end-user web tools and resources (software packages). Countless institutions in the West provide a myriad biological data



Photomicroscopy of transformed material, Biotech Lab, IITA. Photo by IITA

resources and services, including expert-curated databases of nucleic acid and protein sequences, data and text mining tools, genome and transcriptome analysis; protein and other macromolecular structure analysis; networks, pathways, and systems biology; evolution and systems biology tools.

The major tools in the public domain are, however, the development of peer-reviewed, up-to-date, web-accessible databases and web tools (analysis software packages). These resources typically provide an advanced query interface.

The explosive growth of web sites has necessitated that users distinguish between inaccurate personal web sites and reliable resources maintained by a consortium of investigators and/or a legitimate institution. The journal *Nucleic Acid Research* began to publish annually a collection of molecular biology databases and bioinformatics links directory. The most recent updates of molecular biology databases (<http://www.oxfordjournals.org/nar/database/c/>) feature over 1000 databases, over 300 of which are on plants, whereas the latest Bioinformatics Links Directory published by the same journal (http://bioinformatics.ca/links_directory/) lists over 1200 links.

Another outstanding issue in the use of online bioinformatics tools is that, as the number of such web resources grows astronomically, even learning how to use the interface is becoming cumbersome, prompting the need for one-stop gateway type of tools for integrated querying (e.g., BioMart, OBRC from the University of Pittsburgh; Bioclipse).

One of the advances in bioinformatics is the availability of programming and scripting languages (Perl, Bioperl, Python, and Java) for automating complex but routine steps, such as search, retrieval, and parsing (resolving into and examining component parts) search results. While varieties of commercial integrated analysis packages are available, the cost of initial installation and maintenance becomes prohibitive. Developing our capacity for such routine end-user applications is vital to the support of our molecular biology work.

African researchers working on well-studied crops such as rice, wheat, maize, and soybean will have the best genomic resources at their fingertips, provided that they have Internet connection. To take advantage of publicly accessible web resources,

including the variety of databases, online software, publications, and multimedia learning materials, African scientists and students need institutional support and considerable internal and external funding. As in other fields of science, bioinformatics lags in SSA due, partly, to poor or nonexistent Internet connection. Fast and broad Internet connection is the key to successful online research.

Research in molecular biology is slowly gaining ground in Africa. Any molecular biology research needs to be augmented by a bioinformatics database and online tools.

There is no shortage of available tools for agricultural research or agricultural information and database management. The challenge is in finding the best ones or combinations that suit institutional needs, resources, or preferences. IITA will continue to use suitable and affordable conventional and new genomic tools to undertake research on its mandate crops.

