Aspects of Applied Biology ***96****, 2010*
Agriculture: Africa's "engine for growth" - Plant science and biotechnology hold the key

Perspectives on molecular breeding of Africa's main staple food crops - cassava and yam

By MELAKU GEDIL and ALIEU SARTIE

[1]*Central Biotechnology Laboratory, International Institute of Tropical Agriculture, PMB 5320, Oyo Road, Ibadan, Nigeria*
Email: m.gedil@cgiar.org; a.sartie@cgiar.org

Introduction

Root and tuber crops are grown in different parts of the world for different reasons, but in Africa, cassava (*Manihot esculenta*) and yam (*Dioscorea* sps.) are grown as a main staple food for hundreds of millions of people. Widely cultivated in several African countries, cassava and yam provide food security and income generation for millions of small-holding farmers.

With respect to breeding for improved varieties, cassava and yam face similar challenges: 1) The genetics of the crops are poorly understood; 2) Due to lengthy growing season/generation time, the breeding cycle is considerably long; 3) Shy flowering or incompatibility hampers hybridization to develop breeding population; 4) High level of heterozygosity caused by allogamy; 5) Vegetative propagation results in low multiplication ratio of plant propagules; 6) Paucity of genomic resources; 7) Inadequate critical mass of researchers and lack of international collaboration; 8) Absence of research activity by the private sector; 9) High genotype by environment interaction requires breeding for specific agroecology zones; 10) Cultivated by resource-limited small scale farmers with a resultant low productivity or large yield gap; 11) Limited diversity as both cassava and most yam species are not native to Africa; 12) High magnitude of postharvest losses due to bulkiness and perishability.

Enhancing agricultural productivity in Africa is a daunting task entailing a multifaceted approach combining conventional and new technological advances. The genomic boom in the past few decades has opened a new avenue of research for agricultural scientists. Conventional breeding has been augmented by various innovative molecular-marker aided techniques. A wide array of genomic milestones in the past decades has triggered the development and deployment of innovative techniques (e.g. chip based assays, tilling, high throughput genotyping). As the genomic revolution continues to generate large amount of data, a variety of methods of reverse and forward genetics are being developed to aid in efficient and effective plant breeding.

IITA has the global responsibility for yam research and an African responsibility for cassava improvement. In IITA, the development of new genomic tools for molecular breeding and gene discovery is underway for these crops. New candidate single nucleotide polymorphisms (SNP), simple sequence repeat (SSR), and conserved orthologue set (COS) markers have been identified, in silico, from cassava expressed sequence tags (ESTs). Development of ESTs and associated tools in yam is underway. This paper reviews the current status of cassava and yam improvement and gives a perspective on integration of novel genomics technologies with classical breeding schemes for enhanced/accelerated crop productivity in sub-Saharan Africa (SSA).

Cassava

Cassava (*Manihot esculenta*) is the sole member of the Euphorbiaceae family that is grown mainly for food. Dubbed Africa's food insurance, owing to its stable yield under severe abiotic stresses

such as drought and low soil fertility, cassava is not only a famine reserve crop but increasingly becoming an industrial raw material and animal feed (Dixon, 2003). Cassava breeding endeavors in the last few decades have helped elevate the status of cassava from a subsistence (poor man's) crop to a modern food, feed, and industrial crop.

Breeding challenges in cassava

The time required to develop an improved cassava genotype by conventional breeding can be up to 10 years before it gets to the national variety release process, which can take at least six additional years (Okechukwu & Dixon, 2008). Under special circumstances, however, very intensive fast track participatory approaches were developed to release several varieties in 2 years (Dixon *et al*., 2008). The cassava plant generated from stem cutting takes a minimum of 12 months to complete its cycle. The inability of some of the breeding materials to flower and the high rate of abortion subsequent to pollination are additional hurdles to developing breeding populations by hybridization. Being an alogamous species, cassava is characterized by a high level of heterozygosity, which limits the options for genetic and genomic analysis (Ceballos *et al*., 2004). The low multiplication ratio of vegetative plant propagules influences the design of variety trials in terms of plot size and replicated multi-environment evaluations (Ojulong *et al*., 2008).

Priority traits for cassava improvement

Decades of cassava breeding have generated hundreds of improved clones with disease resistance, drought tolerance, high root yield and quality. Cassava is under constant threat from diseases and pests. The major biotic stresses inflicting devastating damage to cassava production are Cassava Mosaic Disease (CMD) and Cassava Brown Streak Disease (CBSD). However, abiotic stresses such as drought also take a significant toll on cassava yield. Therefore, drought, CMD, and CBSD together constitute the bulk of IITA's and partners' research portfolio. The main objectives of the cassava breeding program are:

- Multiple durable resistance to pests and diseases (pyramiding)
- Drought tolerance
- Post-harvest physiological deterioration/biodeterioration
- Enhanced nutritional quality – to mitigate malnutrition or 'hidden hunger'
- Low cyanogenic potential. Different cassava cultivars contain a varying level of potentially toxic levels of cyanogen called linamarin which is associated with certain health disorders.
- Nutrient responsiveness
- Speciality traits for industrial high value root quality
  - High dry matter yield per unit time and area
  - Increased sugar content of roots

Current status of cassava genomics

Recent cytogenetic studies have confirmed that cassava is diploid with 2n=36 chromosomes (Nassar, 2000; de Carvalho & Guerra, 2002). The genome size of cassava is approximately 800mbp (C-values of 0.83 pg) (Awoleye *et al*., 1994). According to GenBank release 172.0 (includes information available as of 10 June 2009), around 81,000 nucleotide entries could be found in public databases (Table 1). Complete chloroplast genome sequence (Genbank accession number EU117376) has been determined for an African land race, TME3 (Daniell *et al*., 2008). The most promising stride in cassava genomics is the near completion of cassava genome sequencing (Steve Rounsley, personal communication). The importance of cassava genome sequencing was strongly advocated by the Global Cassava Partnership (GCP-21) to pave the way for the application of modern molecular bioscience technologies to cassava improvement (Raven *et al*., 2006). Spearheaded by the Danforth Center, the sequencing work had begun at The US Department of Energy Joint Genome Institute (DOE-JGI) a few years back and taken up by a consortium of private, government and non-profit sectors. The outcome of the sequencing project will certainly usher this vital crop to the genomics era opening up a new avenue of efficient and effective approaches to untap the potential of cassava and its wild relatives.

Table 1. *Entrez records for Manihot esculenta at NCBI*

Sub database	Entries	Sub database	Entries
Nucleotide	515	Popset	17
Nucleotide EST	80,459	3D Domains	42
Nucleotide GSS	1,821	GEO Datasets	6
Protein	466	UniSTS	156
Structure	7	PubMed Central	124
Genome Sequences	1	Gene	131
Genome Projects	1	Taxonomy	4

Molecular markers

As most of the traits targeted by breeding programmes are low heritability and quantitative in nature, the use of molecular marker assisted tools is imperative. Currently, the cassava research community has developed several hundreds of molecular marker of which SSRs account for the largest proportion (Fregene *et al*., 1997; Mba *et al*., 2001; Okogbenin *et al.,* 2006). Recently, following development of ESTs in various labs around the world (Anderson *et al.,* 2004; Lokko *et al.,* 2007; Sakurai *et al*., 2007), SNP markers are becoming markers of choice owing to their suitability to the latest high throughput genotyping platforms such as Illumina (Morag Ferguson, personal communication). The present status of molecular genetic analysis in cassava has been thoroughly reviewed in Lokko *et al*. (2007).

Current projects involving molecular work

1. *Drought tolerance*:

A project funded by the Generation Challenge Program (GCP), aims to develop molecular markers associated with drought tolerance traits via QTL and candidate gene mapping in cassava for efficient and cost-effective breeding approach.

2. *Genetic basis of drought tolerance:*

Microarray based expression study to identify key genetic determinants of drought tolerance in cassava and use the information to develop tools for crop improvement.

3. *Development of molecular markers for pyramiding disease resistance genes*:

The objective of the project was to tag different sources of natural resistance to CMD and pyramid CMD resistant genes in cassava germplasm, develop markers for combining different sources of resistance genes, and to initiate molecular marker-assisted pyramiding of resistance genes into elite parental lines for use in developing durable disease and pest resistant germplasm for several African countries.

4. *Targeting Induced Local Lesions In Genomes (Tilling)*:

Of the several techniques of reverse genetics such as RNA interference suppression and transposon tagging, Tilling rapidly gained popularity due to its amenity to automation. High throughput mutation discovery is vital for screening thousands of samples. Besides being a non-GMO approach for broadening the genetic base, it provides tools for development of markers for marker-assisted breeding for traits that are cumbersome and expensive to measure. Preliminary Tilling work to discover induced and natural mutation in cassava was initiated in IITA with a focus on traits that are intractable using conventional methods. Whereas initial work has been done in collaboration with IAEA, external funds and strong collaboration with advanced labs is in view.

5. *The Use of Comparative Genomics*:

Aims to tap new sources of disease resistance genes in cassava and its wild relatives by the technique of resistance gene analogs (RGA). The rapid accumulation of genome sequence data sparked the development of an array of functional genomics tools that are being employed to understand the complex pathways involved in host plant-pathogen interactions. Structural analysis of plant R proteins revealed that most R genes encode homologous proteins featuring nucleotide binding site (NBS) and leucine-rich repeats (LRR). As a result, two major subfamilies of the NBS-LRR type resistance genes were described based on the type of the amino-terminal domains. The NBS-LRR proteins carrying Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) motifs are known as TIR-NBS-LRR and CC-NBS-LRR, respectively. The existence of such conserved domain led to comparative genomics approach to discover candidate disease resistance genes in many plant species. The technique involves three major steps, 1) amplification of resistance gene analog (RGA) region, 2) cloning and sequencing, 3) sequence analysis and similarity search. Such homology-based identification of RGAs has been successfully utilized as an alternative method of resistance gene tagging and genome-wide survey for a complete set of genes featuring the NBS-LRR domain. A study was carried out using several pairs of degenerate primers matching the conserved domains of R genes to amplify putative RGAs in cassava (*Manihot esculenta*), *M. glaziovii*, *M. eprunosa* and castor bean (*Ricinus communis*). In the first phase, PCR products derived from two different clones of cassava were separated by cloning and individually subjected to nucleotide sequence analysis. Each of the 15 recovered sequences was used to perform a BLAST search against the Genbank database as well as against the library of cassava DNA sequences. Two of these sequences were similar to previously reported NBS-LRR genes while the rest showed varying degrees of similarity to cassava ESTs and genomic sequences. Additional RGA regions have been identified in several elite cassava genotypes as well as other Manihot species (Unpublished data, Table 2). Assessment of DNA sequence variation in the amplified region is underway. Further amplification and sequence characterization of other conserved regions, using a different set of primer pairs, is also in progress. Genetic information derived from this experiment is expected to facilitate the identification of gene-targeted markers for molecular breeding. Furthermore, the findings of this study can be complemented with advanced bioinformatics analysis geared towards gene discovery.

Table 2. *Identification of RGAs in Manihot species and castor bean*

Species	Clones identified	Clones sequenced	Unique sequences
M.esculenta			
*TME 279(R)		32	17
*TME 56(S)	42		in progress
*TME 6(R)	40		in progress
*TME 52(S)	116		in progress
M .epuinosa A	71		in progress
M .epuinosa B		53	20
M .tripartita	67		in progress
M .brachyandra	64		in progress
M .elebu	74		in progress
Castor bean	39		in progress

6. *High-density oligonucleotide arrays (DNA chips):*

A genome-wide DNA microarray for cassava with ~ 14,000 probes has been developed. The microarray has been utilized for transcriptome analysis of cassava. Candidate genes that are differentially expressed after virus infection have been identified (unpublished data). In addition to gene discovery, DNA chips provide a reverse genetics tool for identification of gene-targeted markers for molecular breeding. A comprehensive cassava DNA microarray has been developed and tested (Ingelbrecht, personal communiction). In addition, following the development of cassava ESTs, work is in progress to utilize the SNP markers for linkage analysis and molecular characterization by using advanced genotyping platforms such as Illumina (Ingelbrecht et al. 2008).

7. *Application of Diversity Array Technology (DArT) markers in cassava genotyping:*

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% (unpublished data). DArT markers provide reliable and high throughput molecular information for the management of biodiversity in germplasm collections and enables rapid genome profiles for QTL mapping.

8. *Marker development for pro-vitamin A carotenoid (pVAC) in cassava:*

This project is at the very initial stage. Plan is in place to perform in silico research in cassava genome sequence to identify and characterize pVAC-related genes with the support from the Harvest Plus project.

Yam

Yam (*Dioscorea* spp.), a multi-species, polyploid and vegetatively propagated crop, is an economically important staple food crop for more than 300 million people in low income food-deficit countries of the tropics. In addition to their food values, most *Dioscorea* species contain saponins, sapogenins such as diosgenin (Sartour *et al*., 2007), and other alkaloids, which have been exploited for making poisons (Neuwinger, 1996) and pharmaceutical products (Chu & Figueiredo-Ribeiro, 1991). Out of the more than 600 species, 10 are generally cultivated as food: *D. alata*, *D. rotundata*, *D. cayenensis*, *D. bulbifera*, *D. esculenta*, *D. opposita-japonica*, *D. nummularia*, *D. pentaphylla*, *D. transversa* and *D. trifida* (Lebot, 2009). In West Africa, the most commonly cultivated species are *D. alata* and *D. rotundata*. In this sub-region, about 48 million tons of yams are produced annually on 4 million hectares of land. The five major yam producing countries (Benin, Cote d'Ivoire, Ghana, Nigeria and Togo) account for 93% of world production. The average *per capita* consumption of yam in these countries ranges from 204 kcal per day in Nigeria to 395 in Benin. Hence yam is the key to food security (as mainstay for at least 60 million people) and income generation.

In spite of its multiple uses, the average yield of yam in West Africa is far below the potential yield of the crop. In Nigeria, for instance, Olayide (1972) estimated the average yield of yam to be about 14% of the potential yield. This low yield is as a result of factors including the deteriorating soil structure and fertility. Over the past decades a trend of increasing production has been recorded, the bulk of which is due principally to increased area of cultivation rather than yield per ha. The production of this crop is constrained by various biotic and abiotic stresses. In the backdrop of climatic change and poverty, genetic improvement could make a considerable contribution to minimizing the yield gap in yam.

Breeding challenges in yam

Yam genetic improvement is faced with several constraints including the long growth cycle (about 8 months or more, which does not allow the generation of the crop for more than once in a year); dioecy; poor to nonflowering plants; polyploidy; vegetative propagation; heterozygous genetic background; and poor knowledge about the genetics of the crop (Mignouna *et al*., 2007). In addition, although yams are monocots, they are only very distantly related to grasses; for example, banana and wheat are more closely related to each other than either is to yam. Thus there is no convenient model system for yam genomics. In recent years, some progress has been made in germplasm characterization and the development of molecular markers for genome analysis.

Priority traits for yam improvement

Important traits for yam improvement include: high tuber yield per unit of area, time and labour; resistance to diseases (e.g. anthracnose, viruses, tuber rots) and pests (e.g. nematodes); tuber characteristics that facilitate harvesting and are valued by consumers (e.g. size, shape, flesh colour, dry matter content, culinary quality, storability, dormancy); tolerance to abiotic stresses (e.g. mid-season and terminal drought; low soil nutrients); and suitability to prevalent cropping systems (e.g. plant architecture, vigour, and maturity period).

Current status of yam genomics

The knowledge on the genomics of yam is currently very scarce despite its nutritional and medicinal values. Presently, very few nucleic acid and protein sequences for *Dioscorea* species are available in public databases. For instance, a nucleotide sequence search in the recent release of the GenBank database (GenBank release 172) returned only 1015 nucleotide sequences and 1555 protein sequences for the entire genus of *Dioscorea*, and the bulk of these were partial sequences of house-keeping genes derived from organelle genomes (chloroplast and mitochondria) (Table 3). Furthermore, large numbers of entries of the available genomic data in *Dioscorea* were obtained from non-cultivated species while the most important cultivated species such as *D. rotundata* and *D. cayenensis* had significantly lower numbers of entries.

Table 3. *Entrez records of* Dioscorea *species at NCBI*

Sub database	Entries	Sub database	Entries
Nucleotide	1015	Popset	50
Nucleotide EST	39	3D Domains	0
Nucleotide GSS	0	GEO Datasets	0
Protein	1555	UniSTS	4
Structure	0	PubMed	
Central	135		
Genome Sequences	1	Gene	141
Genome Projects	1	Taxonomy	1

Genome size

The database of plant genome size developed and maintained by Bennett & Leitch (2005) contains DNA C-values for 11 yam species, all from the genus *Dioscorea*. It is notable that the Mbp values span a wide range of values from as low as 466 (1C=0.48 pg) in *D. togoensis* to 2352 (1C=2.4 pg) in *D.villosa*, with the C-value of *D. alata* and *D. rotundata* lying between 500–600 Mbp. The difference in genome size is partly attributable to the ploidy level even though values are not available for some species for this to be concluded with certainty (Table 4). Most recently, Arnau *et al*. (2009), have investigated the ploidy of *D. alata* and concluded that most of the 110 accessions have 2n=40 and that the basic number of *D. alata* is x=20. Despite the apparent variation in the estimated values, in general, the genome size of yam species is relatively low.

Molecular markers

Molecular markers (restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), random amplification of polymorphic DNA (RAPD) and SSR) are increasingly used to examine the genetic diversity of cultivated and wild yam species (Mignouna *et al*., 2005; Malapa *et al*., 2005; Egesi *et al*., 2006; Scarcelli *et al*., 2006; Tamiru *et al*., 2007; Tostain *et al*., 2007). SSR markers have been developed and used to characterise the genetic diversity of *D. alata* germplasm collections (Abraham & Arnau, 2007). These markers are now routinely used to select genetically-distant parents to maximise heterozygosity and heterosis in the progenies. The characterisation of the IITA's Core Collection with SSR is also in progress.

Table 4. *Estimated genome sizes of* Dioscorea *species listed in Plant DNA C-values Database release 4.0, October 2005**

Genus	Species	Chrom #	Ploidy level	1C (Mbp)	1C (pg)	Original Reference
Dioscorea	*togoensis*	40	4	466	0.48	Hamon *et al*., 1992 (Ref. 245**)
Dioscorea	*alata*	40	4	564	0.58	Arumuganathan & Earle, 1991 (Ref. 227)
Dioscorea	*abyssinica*	40	4	613	0.63	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*mangenotiana*	40	4	613	0.63	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*praehensilis*	40	4	613	0.63	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*cayenensis-rotundata*	40	4	613	0.63	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*sylvatica*	NA***	NA	833	0.85	Bharathan *et al*., 1994 (Ref. 272)
Dioscorea	*cayenensis-rotundata*	60	6	858	0.88	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*cayenensis-rotundata*	80	8	1274	1.30	Hamon *et al*., 1992 (Ref. 245)
Dioscorea	*villosa*	NA	NA	2352	2.40	Bharathan *et al*., 1994 (Ref. 272)
Dioscorea	*elephantipes*	NA	NA	6615	6.75	Zonneveld *et al*., 2005 (Ref. 466)

* http://www.kew.org/cvalues/homepage.html).
** Numbers refer to references listed at;
http://data.kew.org/cvalues/updates.html#REFERENCES of the C-values database, where C-values refers to the DNA amount in the unreplicated gametic nucleus of an organism, irrespective of the ploidy level of the taxon (Bennett & Leitch, 2005).
*** data not available.

The development of genetic maps allows the use of marker-assisted selection (MAS). Genetic linkage maps based on AFLP markers have been constructed for *Dioscorea tokoro*, a wild yam (Terauchi & Kahl, 1999) and for the cultivated species, *D. rotundata* (Mignouna *et al*., 2002*b*) and *D. alata* (Mignouna *et al*., 2002*c*). The *D. rotundata* map is based on 341 markers segregated in an intraspecific F1 cross. Separate maternal and paternal linkage maps were constructed, comprising 12 and 13 linkage groups, respectively. Several quantitative trait loci (QTL) with an effect on resistance to Yam mosaic virus (YMV) were identified, three on the maternal linkage map and one on the paternal linkage map, showing that both parents contributed to the phenotypic resistance of the progeny. The *D. alata* map is based on 469 markers segregated in an intraspecific F1 cross. These markers were mapped on 20 linkage groups. One QTL located on linkage group 2 was found to be associated with anthracnose resistance, explaining 10% of the total phenotypic variance. The genome coverage of the *D. rotundata* and *D. alata* maps is 56% and 65%, respectively. The saturation of these maps with co-dominant markers such as microsatellites or SSRs and ESTs is necessary for full utilisation of their potential and greater ease of application in yam breeding. Recently, a USAID Linkage Project coordinated by the IITA and Virginia State University (VSU) (USA) generated ≥ 80,000 ESTs. In addition, a joint CIRAD-INRA project is developing new SSR for genetic linkage mapping. To date, only about 60 SSR markers developed from seven different yam species, *D. tokoro* (Terauchi & Konuma, 1994), *D. rotundata* (Mignouna *et al*., 2003), *D. alata*, *D. abyssinica*, *D. praehensilis* (Tostain *et al*., 2006), *D. japonica* (Misuki *et al*., 2005) and *D. trifida* (Hochu *et al*., 2006), are available in yams.

The bulked segregant analysis approach was successfully used for the identification of RAPD markers linked to YMV and anthracnose resistance genes (Table 5). Two RAPD markers, OPW18850

and OPX15850, closely linked in the coupling phase with the dominant YMV-resistance locus Ymv-1, were identified (Mignouna *et al*., 2002*c*). These markers successfully identified the resistance gene in resistant genotypes among a sample of 12 *D. rotundata* varieties. Similarly, two RAPD markers, OPI171700 and OPE6950, closely linked in the coupling phase with the anthracnose resistance locus, Dcg-1, were identified (Mignouna *et al*., 2002*d*). These RAPD markers will make it easier to apply for indirect selection by converting them into co-dominant PCR-based sequence characterised amplified regions (SCARs).

Conventional breeding of yam is time consuming due to various factors including the long growth cycle. The identification of DNA markers linked to key traits that affect yam yield and quality will make it possible to accelerate the gene transfer process.

Table 5. *RAPD markers linked to genetic loci associated with resistance to Yam mosaic virus (YMV) and yam anthracnose disease (YAD) in* D. rotundata *and* D. alata *respectively*

Trait	Locus	Marker
YMV resistance	YMV-1	$OPW18_{850}$
YMV resistance	YMV-1	$OPX15_{850}$
YAD resistance	Dcg-1	$OP17_{1700}$
YAD resistance	Dcg-1	$OPE6_{950}$

To date, IITA has made considerable progress in developing improved high yielding varieties with multiple pest and disease resistance, wide adaptability, and good organoleptic attributes which have been deployed in several west African countries including Nigeria, Benin, Burkina Faso, Ivory Coast, Sierra Leone, Togo and Liberia (Mignouna *et al*., 2007). Research effort in interspecific hybridization should be geared towards the genetic improvement of yam, primarily on *D. rotundata*, *D. caynensis* and *D. alata*, by transfering complementary traits from one to the other, e.g. higher carotenoid in *D.caynensis* can be transferred to *D. rotundata* by interspecific hybridization.

EST development

Most of the currently available molecular markers for the yam genome are based on AFLP and RAPD e.g. the RAPD markers reported for resistance to anthracnose disease in *D. alata* (Mignouna *et al*., 2002*a*) and those for resistance to YMV genus Potyvirus (Mignouna *et al*., 2002*b*). For marker-assisted breeding to be feasible, it is important to develop user-friendly and high throughput markers such as EST-SSRs or SNPs. The lack of DNA or EST sequences hampered fundamental studies such as gene characterization and genetic linkage mapping. The dearth of genomic data in yam species prompted IITA to take initiative in generating fundamental molecular genomic data useful to enhance the conventional yam improvement program. In an effort to develop the yam genomics resources, an initial attempt was made to sequence ESTs from a cDNA library constructed from floral tissue (Mignouna *et al*., 2003). However, the first several hundred sequences were found to be predominantly house-keeping genes, suggesting a better approach needed to be taken in construction of the cDNA library. A new project was launched recently to generate several thousand ESTs in a collaborative project between IITA and University of Virginia (USAID-Linkage fund). The objective of the project was to generate cDNA libraries from yam leaf tissues challenged with Colletotrichum gloeosporioides, the fungal pathogen responsible for yam anthracnose disease, and perform the sequencing of cDNA clones to subsequently identify ESTs with differential gene expression for marker development. In addition, SSR markers were developed in *Dioscorea alata* using sequence resources from heterologous crop species. Some of these markers were polymorphic in the test panel and are presently being tested in IITA Central Biotechnology Laboratory.

Future prospects

Array-based assays

Generation of sufficient nucleotide sequences paves the way for global gene expression analysis via microarray. The first generation of cassava chip has been developed and tested whereas the current EST project in yam is anticipated to generate sufficient ESTs to build microarray chips for transcript analysis. However, a concerted effort to generate DNA, mRNA, and protein data is essential for accelerated development of genomic tools in yam species. Presently, projects are being developed to utilize the cassava genomic resources for crop improvement. Development of genome-wide SNP markers for high throughput genotyping on BeadArray platform (Steemers & Gunderson, 2007) or similar technologies (Hyten *et al*., 2008) is on top of the list.

Identification of candidate genes by comparative genomics

In order to overcome the paucity of gene level knowledge in yam, approaches such as the identification of RGAs can be deployed to identify genes involved in plant defence (Moroldo *et al*., 2008). The rapid accumulation of genome sequence data sparked an array of functional genomics tools that are being employed to understand the complex pathways involved in host plant-pathogen interaction. In the absence of yam genome sequences, such homology-based identification of RGAs can be utilized as a short-cut method for the identification of gene-targeted marker of economically important diseases such as YMV and yam anthracnose.

Application of comparative genomics will further allow the transfer of knowledge from thoroughly-studied model plants to cassava and yam. Discovery of genes involved in flowering in model plants such as Arabidopsis have been successfully utilized to identify homologous genes in garlic (Rotem *et al*., 2007) and in cauliflower (Saddic *et al*., 2006). Similar approaches can be adopted for discovery of genes regulating the flowering signaling pathways in yam.

DNA bar-coding

Species identification in the genus *Dioscorea*, the most important and the largest in the family, has remained a daunting task and the consequences of domestication on species identification has been described above. In IITA, there is a growing interest to apply DNA barcoding not only to address the issues with mislabeling and understanding inter-specific crosses, but also to get an insight into the ongoing domestication process in *Dioscorea*.

Reverse genetics: Tilling

Application of Tilling seems very prudent for yam researchers, capitalizing on advances in functional genomics of model plants. Knowledge of gene function in highly investigated plants sheds light on the genetic mechanism and pathways of key physiological traits in under-researched crops such as yam. Some of the traits that can be targeted by tilling could be resistance to diseases such as YMV, flowering, and dormancy.

Association mapping

The Genetic Resources Unit at IITA has about 3200 accessions of the major food yam species including *D. rotundata*, *D. alata*, *D. bulbifera*, *D. cayenensis*, *D. dumentorum* and *D. esculenta* in a field collection, with a partial backup *in vitro*, which can be used in breeding programmes. Similarly, thousands of cassava accessions consisting of landraces and improved clones are available. Low cost, high throughput genotyping technologies could allow identification of gene/QTLs in these accessions.

Bioinformatics in Africa

African researchers working on thoroughly-studied crops such as rice, wheat, maize, soybean will have the best genomic resources at their finger-tips provided they have an internet connection. To take advantage of publicly accessible web resources including databases, online software,

publications and multimedia learning materials, African scientists and students need institutional support and considerable internal and external funding. Like other fields of science, bioinformatics is lagging way behind in SSA, owing, partly, to poor or non-existent internet connection. Fast and broad internet connection is the key to successful online and in-silico research. Research in molecular biology is slowly taking foothold in Africa. Any molecular biology research needs to be augmented by bioinformatics and online tools. Apparently, the bioinformatics field seems to be largely neglected in SSA. Aside from South Africa (e.g. SANBI), we have no knowledge of a university offering undergraduate or graduate degree, a certificate, or other degrees with emphasis on bioinformatics. When short-term training workshops are organized sporadically, the resource persons and training materials come from advanced labs in N America and Europe (e.g. workshop held at University of Ibadan).

According to the web site of West African Bioinformatics Research Initiative (WABRI) Ilorin, Nigeria (www.wabri.org), there is an initiative on the promotion of teaching and research activities in the fields of bioinformatics, the development of databanks and software, the organization of courses, seminars, and conferences with international collaboration. While the initiative is the only one of its kind in W. Africa, it will, undoubtedly, be stifled by the frequent power failures and erratic internet connection that are commonplace in Nigeria. African Society for Bioinformatics and Computational Biology (ASBCB) is another thriving professional association promoting the advancement of bioinformatics in Africa (www.asbcb.org).

While bioinformatics in IITA is at an infant stage, it has the capacity to offer training good enough to spark interest in those who have a knack for computers. With this view, IITA offered an introductory workshop to beginners with no prior knowledge of bioinformatics. The workshop covered basic concepts of bioinformatics and commonly used public databases; steps and tools for primer designing; commonly used similarity search tools such as BLAST; and phylogenetic analysis. The cassava genome consortium, in collaboration with IITA scientists, has a plan to train young scientists in genomics and bioinformatics tools applicable to crop improvement.

Conclusion

To meet the demand for the rapidly growing population and to mitigate the impact of climate change that has exacerbated the global food crisis, Africa has no choice but to adopt the most innovative plant breeding strategies that integrate the latest innovations in biosciences with conventional breeding practices. Molecular plant breeding has the potential to accelerate the process of crop improvement (Varshney *et al*., 2005; Moose & Mumm, 2008). Arguably, the emerging genetic and genomic resources will assist breeders in devising more efficient strategies aimed at improving multiple traits in a more precise way (Brady & Provart, 2007; Collard & Mackill, 2008; Xu & Crouch, 2008). As cassava will be joining the few other crops and model plants with complete genome sequences available, the cassava research community will have to consider applying the most appropriate genomic-based, cutting-edge technologies to improve the productivity of this major staple crop. Understandably, the cost of genomics-assisted breeding is prohibitive even for developed countries. However, the new initiative by the GCP to form a molecular breeding platform has the potential to realize the adoption of molecular breeding by developing countries. The importance of yam in Sub-Saharan Africa justifies the application of advance genomic tools for germplasm enhancement. It is worthwhile for stakeholders to initiate yam genome sequencing.

References

Abraham K, Arnau, G. 2007. Use of DNA Markers for Genetically Improving the Productivity, Palatability, Storability and Dry Matter Content of Tubers of Greater Yam. In *Mid-term report of IFCPAR Project Number 3000-B1*. New Delhi, India: IFCPAR. 26 pp.

Anderson J V, Delseny M, Fregene M A, Jorge V, Mba C, Lopez C, Restrepo S, Soto S, Piegu B, Verdier V, Cooke R, Tohme J, Horvath D P. 2004. An EST resource for cassava and other species of Euphorbiaceae. *Plant Molecular Biology* **56**:527–539.

Arnau G, Nemorin A, Maledon E, Abraham K. 2009. Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics* **118**:1239–1249.

Awoleye F, van Duren M, Dolezel J, Novak F. 1994. Nuclear DNA content and *in vitro* induced somatic polyploidization cassava (*Manihot esculenta* Crantz) breeding. *Euphytica* **76**:195–202.

Bennett M D, Leitch J. 2005. Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany* **95**:45–90.

Brady S, Provart N J. 2007. Extreme breeding: Leveraging genomics for crop improvement. *Journal of the Science of Food and Agriculture* **87**:925–929.

Ceballos H, Iglesias C A, Perez J C, Dixon A G O. 2004. Cassava breeding: Opportunities and challenges. *Plant Molecular Biology* **56**:503–516.

Chu E P, Figueiredo-Ribeiro R C L. 1991. Native and exotic species of Dioscorea used as food in Brazil. *Economic Botany* **45**:467–479.

Collard B C, Mackill D J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical transactions of the Royal Society of London, Series B, Biological Sciences* **363**:557–572

Daniell H, Wurdack K J, Kanagaraj A, Lee S B, Saski C, Jansen R K. 2008. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. *Theoretical and Applied Genetics* **116**:723–737.

De C R, Guerra M. 2002. Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. *Hereditas* **136**:159–168.

Dixon, A, Akoroda M., Okechukwu R, Ogbe F, Ilona P, Sanni L, Ezedinma C, Lemchi J, Ssemakula G, Yomeni M O, Okoro E, Tarawali G. 2008. Fast track participatory approach to release of elite cassava genotypes for various uses in Nigeria's cassava economy. *Euphytica* **160**:1–13.

Dixon A G O, Bandyopadhyay R, Coyne D, Ferguson M, Shaun R, Ferris B, Hanna R, Hughes J, Ingelbrecht I, Legg J, Mahungu N, Manyong V, Mowbray D, Neuenschwander P, Whyte J, Hartmann P, Ortiz R. 2003. Cassava: From poor farmer's crop to pacesetter of African rural development. *Chronica Horticulturae* **43**:8–15.

Egesi C N, Asiedu R, Ude G, Ogunyemi S, Egunjobi J K. 2006. AFLP marker diversity in water yam (*Dioscorea alata* L.). *Plant Genetic Resources* **4**:181–187

Fregene M A, Angel F, Gomez R, Rodriguez F, Chavarriaga-Aguirre P, Roca W, Tohme J, Bonierbale M W. 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz) 555. *Theoretical and Applied Genetics* **95**:431–441.

H**ochu I, Santoni S, Bousalem M. 2006**. Isolation, characterization and cross-species amplification of microsatellite DNA loci in the tropical American yam *Dioscorea trifida. Molecular Ecology Notes* **6**:137–140

Hyten D L, Song Q, Choi I Y, Yoon M S, Specht J E, Matukumalli L K, Nelson R L, Shoemaker R C, Young N D, Cregan P B. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theoretical and Applied Genetics* **116**:945–952.

Ingelbrecht I, Bak S, Gorodkin J, Jørgensen K, Winter S, Raji A, Gedil M, Hearne S, Ferguson M, Anderson J V, Møller B L, Dixon A G. 2009. New Tools For Genetic Mapping, Fingerprinting, And Gene Discovery In *Cassava; Progress On Marker Development And A Cassava-Specific Oligo DNA Microarray*. Plant and Animal Genome XVII conference, 10–14 Jan 2009, Town and Country Convention Center, San Diego, CA, USA.

Lebot V. 2008. *Tropical Root and Tuber Crops Cassava, sweet potato, yams and aroids*. Crop Production Science in Horticulture Series, 17. MPG Books Group.

Lokko Y, Anderson J V, Rudd S, Raji A, Horvath D, Mikel M A, Kim R, Liu L, Hernandez A, Dixon A G, Ingelbrecht I L. 2007. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reporter* **26**:1605–1618.
Lokko Y, Okogbenin E, Mba C, Dixon A, Raji A, Fregene M. 2007. Cassava. In *Genome Mapping and Molecular Breeding in Plants*, *Vol. 3, Pulses, Sugar and Tuber Crops*, Ch. 14, pp. 249–266. Ed. C Kole. Berlin, Heidelberg: Springer-Verlag.
Malapa R, Arnau G, Noyer J L, Lebot V. 2005. Genetic diversity of the greater yam (*Discorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. As revealed with AFLP markers. *Genetics Resources and Crop Evolution* **52**:919–929.
Mba R E C, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M. 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theoretical and Applied Genetics* **102**:21–31.
Mignouna H D, Dansi A, Zok S. 2002*a*. Morphological and isozymic diversity of the cultivated yams (*Discorea cayenensis/Dioscorea rotundata* complex) of Cameroon. *Genetic Resources and Crop Evolution* **49**:21–29.
Mignouna H D, Mank R A, Ellis T H N. 2002*b*. A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. *Theoretical and applied Genetics* **105**:716–725.
Mignouna H D, Mank R A, Ellis T H N. 2002*c*. A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. *Theoretical and applied Genetics* **105**:726–735
Mignouna H D, Abang M M, Onasanya A, Asiedu R. 2002*d*. Identification and potential application of RAPD markers for anthracnose resistance in water yam (*Dioscorea alata*). *Annals of Applied Biology* **141**:61–66.
Mignouna H D, Dansi A. 2003. Yam (*Dioscorea* spp.) domestication by the Nago and Fon ethnic groups in Benin. *Genetic Resources and Crop Evolution* **50**:519–528.
Mignouna, H D, Abang M M, Wanyera N W, Chikaleke V A, Asiedu R, Thottappilly G. 2005. PCR marker-based analysis of wild and cultivated yams (*Dioscorea* spp.) in Nigeria: genetic relationships and implications for *ex situ* conservation. *Genetic Resources and Crop Evolution* **52**:755–763.
Mignouna H D, Abang M M, Asiedu R. 2007. Yams. In *Genome mapping and molecular breeding Vol. 3: Pulses, Sugar and Tuber Crops*, pp. 271–296. Ed. C Kole. Heidelberg, Berlin, New York, Tokyo: Springer.
Misuki I, Tani N, Ishida K, Tsumura Y. 2005. Development and characterization of microsatellite markers in a clonal plant, *Dioscorea japonica* Thunb. *Molecular Ecology Notes* **5**:721–723.
Moose S P, Mumm R H. 2008. Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* **147**:969–977.
Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, B.De B, Guichard V C, Brunaud V, Le C, Scalabrin I S, Testolin R, Di G G, Morgante M, dam-Blondon A F. 2008. A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. *BMC Plant Biology* **8**:66.
Nassar N. 2000. Cytogenetics and evolution of cassava (*Manihot esculenta* Crantz). *Genetics and Molecular Biology* **23**:1003–1014.
Neuwinger H D. 1996. *African ethnobotany: poisons and drugs: chemistry, pharmacology, toxicology*. London, UK: Chapman and Hall.
Ojulong H, Labuschagne M T, Fregene M, Herselman L. 2008. A cassava clonal evaluation trial based on a new cassava breeding scheme. *Euphytica* **160**:119–129.
Okechukwu R U, Dixon A G O. 2008. *Cassava variety release system in Nigeria: a successful fast-track approach*. Global Cassava Partnership-First Scientific Meeting: Cassava Meeting the challenges of the New Millennium, 21–25 July 2008, Ghent University, Ghent, Belgium. 102 pp.

Okogbenin E, Marin J, Fregene M. 2006. An SSR-based molecular genetic map of cassava. *Euphytica* **147**:433–440.
Olayide S O. 1972. Agricultural Productivity and Increased Food Production under economic development in Nigeria. *Proceedings of the Annual Conference of the Nigerian Economic Society.*
Raven P, Fauquet C, Swaminathan M S, Borlaug N, Samper C. 2006. Where next for genome sequencing? *Science* **311**:468.
Rotem N, Shemesh E, Peretz Y, Akad F, Edelbaum O, Rabinowitch H D, Sela I, Kamenetsky R. 2007. Reproductive development and phenotypic differences in garlic are associated with expression and splicing of LEAFY homologue gaLFY. *Journal of Experimental Botany* **58**:1133–1141.
Saddic L A, Huvermann B, Bezhani S, Su Y, Winter C M, Kwon C S, Collum R P, Wagner D. 2006. The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* **133**:1673–1682.
Sakurai T, Plata G, Rodriguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M. 2007. Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology* **7**:66.
Sautour M, Mitaine-Offer A-C, Lacaille-Dubois M-A. 2007. The Dioscorea genus: a review of bioactive steroid saponins. *Journal of Natural Medicines* **61**:91–101
Scarcelli N, Tostain S, Vigouroux Y, Agbangla C, Daïnou O, Pham J L. 2006. Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Molecular Ecology* **15**:2421–2431.
Steemers F J, Gunderson K L. 2007. Whole genome genotyping technologies on the BeadArray platform. *Biotechnology Journal* **2**:41–49.
Tamiru M, Becker H C, Maass B L. 2007. Genetic Diversity in yam germplasm (*Dioscorea* spp.) from Ethiopia and their relatedness to the main cultivated *Dioscorea* species assessed by AFLP Markers. *Crop Science* **47**:1744–1753.
Terauchi R, Chikaleke V, Thottappilly G, Hahn S K. 1992. Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical and Applied Genetics* **83**:743–751.
Terauchi R, Kahl G. 1999. Mapping of the *Dioscorea tokoro* genome: AFLP markers linked to sex. *Genome* **42**:752–762.
Terauchi R, Konuma A. 1994. Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. *Genome* **37**:794–801.
Tostain S, Agbangla C, Scarcelli N, Mariac C, Daïnou O, Berthaud J, Pham J -L. 2007. Genetic diversity analysis of yam cultivars (*Dioscorea rotundata* Poir) in Benin using simple sequence repeat (SSR) markers. *Plant Genetic Resources: Characterization and Utilization* **5**:71–81.
Tostain S, Scarcelli N, Brottier P, Marchand J -L, Pham, J -L, Noyer J -L. 2006. Development of DNA microsatellite markers in tropical yam (*Dioscorea* sp.). *Molecular Ecology Notes* **6**: 173–175.
Varshney, R K, Graner A, Sorrells M E. 2005. Genomics-assisted breeding for crop improvement. *Trends in Plant Science* **10**:621–630.
Yunbi X, Crouch J H. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science* **48**:391–407.