Chapter 10

Perspectives on the Application of Next-generation Sequencing to the Improvement of Africa's Staple Food Crops

Melaku Gedil, Morag Ferguson, Gezahegn Girma, Andreas Gisel, Livia Stavolone and Ismail Rabbi

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61665

Abstract

The persistent challenge of insufficient food, unbalanced nutrition, and deteriorating natural resources in the most vulnerable nations, characterized by fast population growth, calls for utilization of innovative technologies to curb constraints of crop production. Enhancing genetic gain by using a multipronged approach that combines conventional and genomic technologies for the development of stress-tolerant varieties with high yield and nutritional quality is necessary. The advent of next-generation sequencing (NGS) technologies holds the potential to dramatically impact the crop improvement process. NGS enables whole-genome sequencing (WGS) and re-sequencing, transcriptome sequencing, metagenomics, as well as high-throughput genotyping, which can be applied for genome selection (GS). It can also be applied to diversity analysis, genetic and epigenetic characterization of germplasm and pathogen detection, identification, and elimination. High-throughput phenotyping, integrated data management, and decision support tools form the necessary supporting environment for effective utilization of genome sequence information. It is important that these opportunities for mainstreaming innovative breeding strategies, enabled by cutting-edge "Omics" technologies, are seized in Africa; however, several constraints must be addressed before the benefit of NGS can be fully realized. African breeding programs must have access to high-throughput genotyping facilities, capacity in the application of genome selection and marker-assisted breeding must be built and supported by capacity in genomic analysis and bioinformatics. This chapter demonstrates how interventions with NGS-enabled innovative strategies can be applied to increase genetic gain with insights from the Consortium of International

Agricultural Research (CGIAR) in general and the International Institute of Tropical Agriculture (IITA) in particular.

Keywords: Next-generation sequencing, genotype by sequencing, genome selection, plant breeding, genetic gain, developing countries

1. Introduction

Africa is the region with the highest prevalence of hunger and malnourishment. The persistent challenge of insufficient food, unbalanced nutrition, and deteriorating natural resources in the most vulnerable nations, characterized by fast population growth, calls for utilization of innovative technologies to curb constraints of crop production. Major revitalization of agricultural research in Africa is needed to underpin necessary increases in sustainable productivity in anticipation of the increase in population and changes in climate. Since many of the clonally propagated crops grown in Africa, such as cassava, yams, bananas, and plantains, and seed crops, such as cowpea, tef, sorghum, and millet, are not commonly consumed as food outside of the region, researchers in Africa have the responsibility to devise innovative breeding strategies for these crops. African agriculture is characterized by subsistence farming by smallholder farmers growing various locally adapted crops, many of which are considered understudied or "orphan" crops. These crops are vital for providing nutrition and income to resource-poor farmers, particularly in the face of confounding climatic and soil constraints. A regular supply of high-yielding nutritional varieties that respond to the changing biotic and abiotic stress environment is required. Conventional plant breeding has contributed tremendously to increased crop yields; however, the rate of genetic gain over the past few decades has been relatively slow for a number of reasons, including the lengthy breeding cycle, a characteristic of many clonally propagated crops [1]. Enhancing genetic gain entails a multifaceted approach of combining conventional and new technological advances [2,3].

The Consortium of International Agricultural Research, abbreviated as CGIAR, in collaboration with partners, is spearheading agricultural biotechnology research in Africa [4]. Several consortium research programs (CRP) are performing collaborative research on more than a dozen staple food crops of developing countries, including vegetatively propagated root, tuber, and banana (RTB), about seven grain legumes, and four dryland cereals. These crops support the livelihood of hundreds of millions of resource-limited farmers and traders in developing nations. The vegetatively propagated RTB crops (cassava, yam, potato, sweet potato, banana, and plantain) share many breeding challenges, including pathogen transmission from one generation to the next, polyploidy, low fertility and multiplication rates, and long breeding cycles. These can best be addressed by exploiting synergies across crops and technologies to increase genetic gain per unit time. Furthermore, the attainable yield potential of extensively studied crops such as rice, maize, wheat, and soybean are considerably lower in developing countries owing to unique production constraints in Africa calling for unique

http://dx.doi.org/10.5772/61665

intervention, including genomics. Declining costs of DNA sequencing have triggered a surge in research on crops of local or regional importance and, with time, should translate into increased yields and yield stability, thus reducing the reliance on a smaller number of major crops [2,5–7].

This chapter initially outlines current and prospective genomic resources pertaining to Africa's staple crops, and then discusses how genomics strategies in the era of high-throughput next-generation sequencing technologies are being applied to increase genetic gain in developing countries with insights from CGIAR in general, and IITA in particular.

2. NGS-based omics resources: Current and prospective

2.1. Whole-genome sequencing

Knowledge of a crop genome sequence is fundamental for understanding biochemical and physiological processes that govern plant traits and the way in which they respond to environments- and biotic and abiotic stresses. The rapid evolution of genome sequencing technologies [8] has resulted in an explosion of genomic information, the sequencing of a vast number of plant genomes, and opportunities to apply this to crop improvement, e.g., through the development of genome-wide marker assays [9,10]. In the rapidly changing landscape of life science technologies, a number of new disciplines have emerged, particularly for deciphering gene function and metabolic pathways; these include transcriptomics, proteomics, metabolomics, small RNAomics, epigenomics, interactomics, together with the corresponding development of bioinformatics tools and databases to support these. It is important to ensure that, as our understanding of biological processes increases, this is translated into enhanced agricultural productivity through research for development (R4D).

The genome sequences of many major world crops have been completed in the past decade, as well as a few crops of specific importance to the developing world, including cassava, yam, tef, pigeon pea, and peanut, while many still remain to be sequenced [11–13]. A drive to sequence more crop plants, particularly orphan crops of Africa, is in progress. A recent public and private sector initiative called African Orphan Crops Consortium (AOCC, http://africa-norphancrops.org/) aims to sequence, assemble, and annotate the genomes of 100 traditional African food crops.

The cost of DNA sequencing per raw million bases fell from $8,000 to $0.1 between 2001 and 2013 according to Wetterstrand, K.A. (http://www.genome.gov/sequencingcosts/) cited in [8]. With the advent of the third-generation sequencing technologies, the cost is expected to reduce still further while the speed, quality, and throughput increase exponentially. Currently, most of the staple food crops that IITA is working on have been sequenced or are being sequenced (Table 1). The focus is thus on post-genomics analysis such as genome annotation and describing gene functions as applied to crop breeding. With a fledging bioinformatics capacity, and a network of partners in advanced laboratories as well as collaboration in the CRP of CGIAR, the breeding programs in IITA are moving toward molecular breeding for enhanced

genetic gain with the aim to transfer these innovative genomics-assisted breeding schemes to our partners in the national agricultural research systems (NARS).

Species	Subspecies/ genotype	Family	Genome size (Mbp)	No. of predicted genes	Chromosome no. (2n)	Reference
Maize	*Zea mays ssp mays* B73	Poaceae	2,300	39,656	10	[15]
Soybean	*Glycine max*, variety Williams	Fabaceae	1,115	46,430	20	[16]
Cowpea	*Vigna unguiculata*	Fabaceae	620	5,888 GSRs	22	[17]
Cassava	*Manihot esculenta*	Euphorbiaceae	770	30,666	18	[18,19];
Banana	*Musa acuminata* (ssp. *malaccensis*)	Musaceae	523	36,542	22	[20]
Yam*	*Dioscorea rotundata*	Dioscoreaceae	594	21,882	20	[21]
Cacao	*Theobroma cacao* cv. *Matina*	Malvaceae	430	28,798	20	[22]

*At the time of the writing, manuscript is in preparation. Preliminary results were presented at an international conference.

Table 1. Current status of whole-genome sequences of IITA mandate crops

2.2. NGS-based genotyping and marker analysis

Massively parallel sequencing technology enabled high-throughput genotyping at an unprecedented scale. Whole-genome sequencing and re-sequencing of genome and transcriptome have yielded hundreds of thousands of single-nucleotide polymorphism (SNP) markers in several crop plants, including orphan crops. In recent years, diverse next-generation-based reduced representation protocols have been developed for the simultaneous discovery and generation of massive, genome-wide SNP data that have been applied to linkage mapping, quantitative trait locus (QTL) analysis, diversity studies, genome selection, and population genetics [14]. Protocols for reduced representation can be optimized to any species with or without a reference genome sequence [15]. The most widely used strategies for complexity reduction genotyping are restriction-site-associated DNA (RAD) [16] and genotyping by sequencing (GBS) [17], and diversity array technology (DArT)-seq, which combine complexity reduction methods and utilize a microarray platform [18]. All have been optimized for multiple plant species.

GBS protocols allow for a high level of multiplexing of up to 384 samples in one sequencing reaction, making it presently the most inexpensive and scalable assay with a library construction less complicated than RAD [19,20]. Researchers in developing countries presently focus on multiplex genotyping platforms such as GBS for genotyping cassava, yam, banana, maize,

http://dx.doi.org/10.5772/61665

and cowpea for diversity analysis and molecular breeding. However, the deployment of such SNP markers in forward breeding, where only a few specific markers are tracked, entails the selection of suitable, cost-effective assays from a wide array of genotyping platforms such as fixed arrays or flexible singleplex assays [21]. Conversion of SNPs of interest into one of the above platforms requires bioinformatics analysis pipeline to design and optimize an assay. In the CGIAR systems, the Kompetitive Allele-Specific PCR (KASP) genotyping assay is widely applied (e.g., [22]). New initiatives are being developed to establish a cost-effective genotyping hub aiming to reduce the cost of data points by fivefold. Multiplex genotyping assays such as GBS, RAD, and DArT have been successfully used to identify SNP markers associated with the trait of interest in understudied crops. Examples include disease resistance in lupin [23], pepper [24], cassava [25,26], and beans [27].

Reduced representation sequencing (RRS)-based genotyping methods have the drawback of missing mutations at the recognition site of the restriction enzymes used [19]. The use of other enzyme combinations could circumvent this problem by altering the library construction [20, 28]. In addition, the accuracy of base calling in complex polyploids and heterozygous individuals, of which there are several examples within the root and tuber staple crops of Africa, can also be problematic. Given the rapid pace of advances in both the chemistry of sequencing such as the advent of the third-generation sequencing with longer read length and shorter assay time [29] and informatics pipelines (viz. imputation), the cost and accuracy of sequence-based genotyping are anticipated to decline in the foreseeable future.

2.3. NGS-based gene expression analysis

Transcriptomics is the study of the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition [30]. The transcriptome includes all RNA molecules, including mRNA, rRNA, tRNA, small RNAs, and other noncoding transcribed RNA and can vary with external environmental conditions. Transcriptomics studies often try to catalog these transcripts, as well as determining the transcriptional structure of genes, in terms of their start sites, 5′ and 3′ ends, splicing patterns, and other posttranscriptional modifications. By quantifying the expression levels of specific transcripts under different conditions or development stages, transcriptomics can help to understand the functional elements of the genome, including cellular processes and biochemical signaling pathways. Two main approaches have been used: based on hybridization and sequencing. Cassava is one of the very few African staple food crop to which microarrays have been applied [31–36].

Although hybridization approaches are relatively high throughput and inexpensive compared to the alternative expression assays, they do have technical limitations and require a priori knowledge of gene transcripts. NGS with its advantages of exceptional throughput and relative affordability has now enabled sufficient depth of sequencing for the study of whole transcriptome in a comprehensive manner. This method, termed RNA-Seq (RNA sequencing), has clear advantages over other existing approaches and is fast becoming the most popular method for analysis of eukaryotic transcriptome [30]. RNA-Seq also provides a far more precise measurement of levels of transcripts and their isoforms than other methods. To date, the majority of applications of RNASeq to Africa's staple crops have focused on understanding

natural host responses to plant viruses. RNA sequencing was used to identify 700 uniquely overexpressed genes in the cassava brown streak disease (CBSD) resistant variety under cassava brown streak virus (CBSV) infection [37]. Although none of the overexpressed genes corresponded to known resistant gene orthologs, some belonged to hormone signaling pathways and secondary metabolites, both of which are linked to plant resistance. Similarly, the transcriptome of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava (12, 32, and 67 days post infection) was investigated [38]. Significantly, they found that susceptibility was mediated by transcriptome repression, rather than induction, and many R-gene homologues were repressed throughout infection in the susceptible individuals. In another study, NGS was deployed to investigate the role of miRNAs in plant growth and starch biosynthesis [39,40]. IITA and partners have completed an RNA-seq study in yam for the purpose of assembling the whole-genome sequence of *Dioscorea rotundata* and annotating predicted genes [41]. In addition, RNA-seq-based transcriptome has revealed rice genes involved in the signaling pathway for resistance to Striga [42] that may in turn shed light on the mechanism of resistance in other African crops that are vulnerable to Striga (e.g., maize, sorghum, and cowpea). Illumina-based sequencing of transcriptome from four underutilized leguminous crops has led to the development of markers for phylogenetics and comparative mapping [43]. NGS was used in modified bulk segregant RNA-seq (BSR-seq) method to clone a mutant gene in maize [44].

In addition, RNA-seq has been used successfully to address several production constraints of orphan crops [45–47], and it is envisaged that this will be a popular approach in the future. Other areas of interest for application of this technique are to understand the mechanism of Striga tolerance in maize and cowpea, yam anthracnose resistance, flowering and sex determination in yam, and drought tolerance in several crops (maize, cassava, cowpea). A single RNA-seq experiment involves taking samples at different stages of growth, tissue, and replicates. Multiplying the aforementioned factors by the number of crops and the number of traits per crops results in numerous libraries, which implies high assay cost. In this light, having in-house capacity to construct the libraries will significantly lower the cost and allow proper control of the experiment.

2.4. Bioinformatics and database

The field of bioinformatics has faced an unprecedented challenge, as a result of the new high-throughput technologies, particularly NGS, which has redefined the last decade of research in biology [48]. However, these technologies would never have made such progress without the attendant advances in the field of bioinformatics. Sequencing DNA and RNA has become so cheap and so vast that NGS is now a basic technology for many fields of research in medicine, basic research, as well as research in agriculture. In agricultural research, NGS is applied in whole-genome sequencing (WGS), whole-genome re-sequencing (WGRS), transcriptomics, metagenomics, and reduced representation sequencing for high-throughput SNP genotyping [15,21,28,29,49]. A genome sequence becomes only useful for biological applications when the genome is annotated and genes are described and their functions revealed [50]. Besides the functionality of genes, the variability of the genome of different varieties of a species is

http://dx.doi.org/10.5772/61665

important to understand the different properties a species can demonstrate [13,51]. This last point together with the functionality information is a very important opportunity to support and improve breeding activities in crops of economic importance [52].

An extensive review of NGS data analysis is beyond the scope of this chapter. An insight into the status of NGS analytical tools and cross-references (articles, books, and dedicated issues of journals) are provided in a recent review [8]. The authors classified the NGS software tools into four general categories – alignment of sequence reads, base calling, and/or polymorphism detection, de novo, and genome browsing and annotation – and cited that a gamut of packages have been developed for each category by Barba et al. [8]. Of course, as the sequencing technology evolves, the bioinformatics software tools and algorithms have to be developed to keep pace with them. Likewise, workflow and various analysis strategies and challenges have been described for metagenomics [53–55].

The focus of this chapter is the application of NGS to the improvement of crops that are the mainstay of hundreds of millions of people in the developing world. Presently, the major application of NGS is genotyping by GBS and RNA-seq in crops such as cassava, yam, maize, banana, and cowpea, among others. Using these technologies necessitated the establishment of a moderate bioinformatics platform at IITA not only to serve basic bioinformatics needs but also to support the genotyping efforts in the aforementioned crops. The platform hosts the basic bioinformatics tools such as alignment and basic sequence analysis tools. For the data analysis of NGS data, the server is equipped with tools for de novo assembly [56] and mapping [57] as well as specific needs such as genotyping by sequencing [17], transcriptomics [58], noncoding RNA (ncRNA) [59,60], DNA methylation [61,62], and metagenomics [63] as new horizons to accelerate genetic gain.

It is worthwhile to describe some applications that are routinely run in IITA to support the research activities of IITA because, ultimately, the technologies are transferred to partner national research programs. GBS is a very cost-efficient genotyping approach by reducing the complexity of the genome and increasing the number of genotypes per sequencing round. There exist several bioinformatics pipelines to clean and analyze such data. IITA installed Tassel5 [64] and GATK [65] as the most useful tools. The Tassel plug-ins are assembled to a full automatic workflow to produce a filtered variant call format (VCF) file [66]. With Tassel, the bioinformatics server of IITA is able to easily analyze more than 5,500 genotypes in parallel having approximately 1.2 TB compressed sequencing data available. The analysis runs over 2 days using at most 250 GB RAM. The analysis picks about 350,000 SNPs, which get reduced by filtering to about 170,000 high-quality SNPs, which are a reasonable number for downstream analyses such as population genetics and clustering as well as QTL analysis. The same workflow for genotyping is now applicable for different plant species, and analyses have been performed for cassava, *Dioscorea*, maize, and planned for *Musa*.

A workflow using Picard Tools and GATK is under construction and will be available for any kind of DNA sequencing data. IITA is also in the process of establishing a pipeline for the analysis of RNA-seq data using several available Illumina RNA sequencing data sets from contrasting genotypes. As a reference sequence was available, three different analyses were performed: a de novo sequence assembly to discover new unannotated genes or new alterna-

tive splice variants; mapping on the reference genome to elaborate the expression level of known, annotated genes; and the differential expression of selected genes between different genotypes. Such studies will become increasingly important for modern breeding programs since especially biotic and abiotic stresses are clearly regulated by different mechanisms other than purely genetic variations.

First experiments were conducted to study the DNA methylation profile on the model plant *Arabidopsis* to study epigenetic changes upon biotic stresses. A whole set of tools were installed and in-house scripts developed to analyze data derived from whole-genome bisulfide (BS) transformation [67]. The BS transformation converts non-methylated cysteine into a uracil and later, after polymerase chain reaction (PCR) amplification, into a thymine, whereas the methylated cysteine remains a cysteine. Since this technique is looking for single-nucleotide events and since the genomic code is "falsified," there is the need for a high-quality reference and specialized mapping strategies and statistics for the methylation calling [68]. The availability of a good-quality reference genome sequence of cassava and whole-genome re-sequencing of several clones of interest prompted DNA methylation profiling for some relevant cassava varieties. In this pilot study at IITA, currently in progress, the aim is to reveal dynamic methylation events under biotic and abiotic stresses to gain information on possible epigenetic markers for the next-generation breeding programs.

With the development of NGS noncoding RNA (ncRNA), especially the smaller species became very easy to detect, and many studies demonstrated that these ncRNAs are important players in gene regulation, regulation of DNA and histone methylation, and defense mechanisms in plants. ncRNA profiles are also important for diagnosing and characterizing virus infections in plants [69]. The virus infection triggers a defense reaction where a cascade of host ncRNA are involved, but also small interfering RNAs (siRNAs) corresponding to the viral genome are found in the plant extract. These endogenous ncRNA and the viral small RNA fragments can easily be detected by NGS. At IITA, we have the expertise and software suite of tools to search and analyze any plant ncRNAs or virus siRNAs. Again biotic and abiotic stresses in plants have a specific profile of expression of different species of ncRNA, and at IITA, we study this phenomenon to create information and tools to improve the breeding programs.

2.5. Genome editing

Genetics relies on the analysis of mutations and the phenotypic variation they cause to correlate precise sequence changes to particular genes of interest. With the help of genetic engineering techniques, desired traits can also be introduced into plants not expressing them naturally. However, the use of genetically modified crops is hindered by health, environmental, and ethical concerns. Genome editing with site-specific nucleases is the most advanced technology for precise and effective genome engineering, which promises to revolutionize applied research for crop improvement [70,71]. It involves the insertion, elimination, or replacement of a fragment of DNA at desired locations in the genome, by using engineered nucleases that create specific double-strand breaks (DSBs) and stimulate cellular DNA repair mechanisms. There are currently four classes of targetable nucleases discovered and bioengineered that are

http://dx.doi.org/10.5772/61665

used to create site-specific DSB: zinc finger nucleases (ZFNs), transcription activator–like effector nucleases (TALENs), clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) RNA-guided nucleases (RGNs), and engineered meganuclease, also known as homing endonucleases [72–75].

Over the past few years, all of the above nucleases have been used to create target-specific mutations in model and crop plants, albeit with some limitations. In all cases, a continuing issue is the delivery of all the reagents efficiently and functionally to the cells or organisms under study. The CRISPR/CRISPR-associated protein 9 (Cas9) tool seems to overcome some of the shortcomings of the other methods [76,77]. Successful examples of targetable nucleases application are reported for *Arabidopsis*, tobacco, rice, maize, soybean, barley, cabbage, and bunchgrass by using different delivery technologies, including T-DNA plasmid from Agrobacterium, protoplasts and embryonic callus manipulation, and subsequent plant regeneration [70,78–82].

Targetable nucleases are attractive alternative biotechnological tools for trait manipulation and breeding in crop plants. By means of targetable nucleases, mutations can be produced in a very specific manner, and known mutations can be transferred between cultivars or breeding lines without disrupting a favorable genetic background. Although genome editing approaches are relatively new and not yet widely applied, their advantage in terms of safety, robustness, speed, and precision over the classical mutagenesis and breeding is undisputable [75]. Targeted genome editing using artificial nucleases, combined with accurate gene expression analyses, has the potential to accelerate plant breeding by providing the means to modify genomes rapidly in a precise and predictable manner [71] and to restore lost traits through reverse breeding [83]. Although genome editing has not yet been applied to African staple crop species, there is no doubt that this technology will assume a great importance particularly for genetic improvement of asexually propagated crops with limited flowering ability [71].

Furthermore, technologies based on targetable nucleases offer the opportunity to overcome the major concerns of the general public about transgenic crops since the organism with the edited gene do not contain the foreign DNA. In particular, the absence of extra copies of DNAs upon nonhomologous end joining (NHEJ)-mediated gene knockout makes the final plant comparable with those arising from natural mutations. However, the development of dedicated international legislations is required to effectively promote a wide application of genome editing technologies for crop improvement [70,84]. As knowledge is gained about plant genome organization and gene functions are revealed, the potential of genome editing could be mainstreamed to broaden the genetic base of crops.

2.6. Targeting Induced Local Lesions in Genomes (TILLING) and NGS-based mutation detection

One of the factors contributing to slow genetic gain in breeding of vegetatively propagated crops is the narrow genetic base of the source population. This is a result of clonal propagation as opposed to sexual reproduction, which limits recombination. TILLING (Targeting Induced Local Lesions in Genomes) [85,86] provides an alternative approach for creating novel variation in these crops [87,88]. Rare alleles harbored in germplasm collections and wild

species can be accessed by TILLING and EcoTILLING by sequencing. TILLING may lead to the development of functional markers for screening-associated traits through marker-assisted selection (MAS). The technique of TILLING using high-throughput mutation discovery has already been applied successfully to more than 20 plant species [89].

A wide spectrum of mutation detection assays, ranging from heteroduplex analysis with high-pressure liquid chromatography (HPLC), screening with labeled primers, electrophoresis, microarray, the use of fluorescent dye-labeled primers assayed on ABI genetic analyzer have been used. However, these methods are generally slow, costly, and labor intensive. Application of NGS has been shown to be a cost-effective mutation detection system by re-sequencing the gene of interest in mutagenized plants [90,91]. The availability of genome sequence enables the use of reverse genetic approaches to identify mutations in specific target genes, thereby accelerating the generation of novel phenotypes. Comparative genome analysis methods offer the opportunity to select target genes involved in biosynthetic pathways and networks of traits/phenotypes of economic importance. The use of multidimensional pooling of DNA samples enables screening of DNA pools for multiple independent mutations in any target gene using NGS, which provides a cost-effective assay. This has led to the discovery of rare mutations in rice and wheat, termed TILLING by sequencing [92], tef [93], and in animals [94]. Different sample pooling schemes for NGS, which further enhance the power of NGS in processing multiple samples in parallel have been developed [95]. In light of the rapidly evolving sequencing technology together with a plethora of sample pooling schemes, combined with bar coding, it is feasible and imperative to apply TILLING by sequencing to understudied crops of Africa. A direct application of NGS to detect mutant regions in a segregating population of rice has been demonstrated in a method called MutMap [96].

2.7. QTL identification

This section discusses how NGS can be used to enhance QTL analysis. Following the advent of first-generation molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP), numerous studies in many crop species were launched to identify QTL, but for quantitative traits, affected by polygenes with small effects, limited success was attained in terms of application [97]. One of the explanations [98] for the limited exploitation of QTLs is the issues associated with the acquisition and summarizing of plethora of QTL information.

The rapid advance in next-generation sequencing technologies and the wide array of ultrahigh-throughput and cost-effective genotyping platforms have created a multitude of new possibilities for QTL mapping using large early-generation populations and high-density markers. Variants of NGS-based QTL identification methods, such as X-QTL, MutMap, QTL-seq, SHOREmap, and NGM, have been reviewed elsewhere [99]. Among the various NGS-based QTL mapping approaches, QTL-seq, the whole genome re-sequencing-based mapping of QTL [100], can successfully be applied to dissect key quantitative traits underlying biotic and abiotic stresses in major African staple food crops such as cassava, yam, tef, and legumes. One of the essential requirements for QTL-seq is the availability of a quality reference genome and

http://dx.doi.org/10.5772/61665

mapping populations. The technique has been applied to rice where the whole genomes of two pooled rice DNA samples with contrasting phenotypes each in F2 and recombinant inbred line (RIL) populations were re-sequenced, after which the short reads were aligned to the reference sequence to calculate an SNP index. QTL were declared at positions where the SNP were different from the reference and had an SNP index value of 1. The analysis uses careful filtering of spurious SNPs. Conventional QTL mapping verified the candidate QTLs detected by the QTL-seq, and the method was validated by simulation analysis. QTL-seq has also been used in cucumber to map a QTL involved in flowering trait [101]. Likewise, the deployment of QTL-seq for rapid identification and fine mapping of QTLs was reported in chickpea [102] and sorghum [103].

In IITA, there are ongoing projects aiming to apply this technique to mapping of QTLs controlling disease resistance (e.g., anthracnose and yam mosaic virus), as well as root quality traits such as starch content. In cassava, the approach of genome-wide association study (GWAS) and conventional QTL mapping in F1 populations is being pursued to identify markers associated with key traits, including yield, dry matter, quality, and resistance to disease.

2.8. Metagenomics

Metagenomics is the direct genetic analysis of genomes contained within an entire community of organisms such as a microbial community, and makes use of NGS technologies and bioinformatics tools [104]. The advent of metagenomics has revolutionized the study of microbial ecology, evolution, and diversity. In plant pathology and virology, metagenomics has contributed to the sequencing of genomes within infected plants and has led to the detection of many RNA and DNA viruses and/or viroids. Other areas of application include ecology and epidemiology as well as functional genomics of pathogens, and the culture-independent analysis of a mixture of microbial genomes [8,105,106].

The application of metagenomics in crop improvement is discussed below in the disease diagnostics section as the majority of plant metagenomics studies, as applied to agriculture, relate to virology. However, there are substantial shotgun metagenome sequencing studies that investigate microbial communities in soil and plants and other environmental samples [105,107–109]. The challenges of analysis are being addressed gradually [55,104]. The analysis pipeline for metagenomics follows major steps such as raw data quality checking, filtering, assembly, taxonomic classification, abundance estimation, and relative quantification of taxons [53,54].

With growing experience in NGS data analysis and a fledging bioinformatics critical mass, IITA and partners are moving toward the application of meta-omics (-genomics, -transcriptomics, and -proteomics). In the context of African agriculture, the rapidly evolving field of metagenomics will have a significant impact in revealing the diversity of microorganisms, and in describing the relationship between host-associated microbial communities and host phenotype. The declining cost of sequencing and the associated analytical tools will likely create the opportunity to develop cost-effective and efficient diagnostic kits to address the challenge of multiple infections (pathogenic races and strains) in the major crops such as

cassava [110], banana [111], and yams [112]. Survey of the incidence and distribution of viruses infecting these crops makes it one of the important tools for understanding the microbial genetics, physiology, and community ecology. The benefit of metagenomics extends to agriculturally important microbes, both disease causing and beneficial, in plant and animal production.

3. Application to crop improvement

3.1. Molecular breeding

The role of molecular markers in facilitating selection has substantially increased in the past three decades. The rapid accumulation of genomic resources provides researchers with an unprecedented wealth of information to access and manipulate genetic variation that is useful for crop improvement [113]. Genomics-assisted breeding is expected to enhance the accuracy and efficiency of breeding programs to deliver superior cultivars for sustainable agriculture. The ultrahigh throughput and decreasing cost of genotyping have elicited concepts such as genomics-assisted breeding [52] and breeding-assisted genomics [114]. Currently, the new paradigm among the Consortium of International Agricultural Research Centers (www.cgiar.org) is to mobilize "Omics" and bioinformatics-enabled interventions to assess the level of available genetic variation, to broaden the genetic bases by creating new intra- and inter-species variations, to construct new cultivars with combinations of desirable and novel traits in more efficient and effective selection schemes. The ultimate goal is to accelerate genetic gain, which will contribute to improved food and nutritional security, in an environmentally sustainable way, in low-income countries.

The unprecedented scientific and technological progress in the fields of genomics and bioinformatics can successfully be harnessed to benefit smallholder farmers in developing countries. In the face of limited agricultural inputs in developing countries, genetic improvement can play a crucial role in raising crop productivity in an environmentally sustainable way. Spurred by steadily declining costs of genotyping and unparalleled progress in computational abilities, modern genomic tools and processes are being used to devise an efficient and effective breeding strategy. The prominent constraints to breeding progress are slow genetic gain, complex traits, and genotype by environment interaction. Besides these generic constraints, neglected crops of Africa were affected by a paucity of genomic information until the dawn of NGS.

It is now feasible to access genome-wide nucleotide variation by re-sequencing the whole genome of thousands of accessions or by deploying one of the complexity reduction methods to generate high-density, genome-wide SNP markers associated with key agronomic traits attributed to quality, resilience to climate change, and biotic stresses. These technological advances led to the design of experimental populations involving multiple parents, in addition to the classical genetic mapping within specific biparental crosses. An overview of IITA's (and CGIAR's) activities in addressing crop productivity and other agricultural problems has been documented [4].

http://dx.doi.org/10.5772/61665

Evidence is emerging that the massive availability and accessibility of genomic resources and data management tools are paving the way for the deployment of innovative technologies to accelerate genetic gain. A number of recent reviews analyze the potential benefit of the Omics technologies to agricultural productivity and highlight various limitations that need to be addressed [19,27,52,115].

The two major approaches in the new paradigm of molecular breeding are (1) MAS for highly heritable traits and (2) GS for complex traits. These approaches involve the genotypic screening of large numbers of individuals at an early stage, selection at the seedling stage, and extensive phenotypic evaluation of fewer materials at a later stage. This reduced breeding cycles and the cost of multi-environment testing. Strategies such as GS also allow simultaneous selection for multiple traits through a selection index [52,116–119].

Broadly, there are two approaches to exploit QTLs. The first application is to detect large-effect QTLs with linkage or association analysis, whereas approaches such as GS utilize the computation of an individual breeding value based on genome-wide marker genotype, without taking into consideration the single small-effect QTLs in the prediction model.

Numerous reviews, opinion articles, and research papers have addressed the benefit, challenges, and prospect of GS crystallized in a recent review [113]. The salient features of GS include benefits such as increased gain from selection, reduced breeding cycles, and thus reducing cultivar development costs. Other advantages include utilization of genome-wide markers, afforded by ultrahigh-throughput NGS assays (compared to predecessor approaches to estimate breeding values), as well as the ability to target multiple traits for multiple environments. In clonally propagated crops, an additional advantage is the use of historical phenotype data to refine the prediction model.

Given the long cycle of breeding, African staple crops such as cassava are set to benefit from GS approaches [117,118,120], where preliminary results have indicated reduced time of breeding cycle and reasonable prediction accuracy in some traits. Various ways of refining the prediction models via repeated phenotypic evaluations are being considered. Fig. 1 depicts a 1-year GS-based breeding cycle that is underway at IITA, Nigeria. The challenge in this breeding scheme is, however, the situation of erratic flowering in some lines, which hinders recombination of selected clones due to failure to flower. Addressing the biology of flowering using genomics tools is imperative. In cereals, current studies are investigating at least two key applications of GS in maize and wheat breeding programs – predicting the genotypic values of individuals for potential release as cultivars and predicting the breeding value of candidates in rapid cycle populations. Prediction accuracy is affected by genetic relatedness of the populations and the heritability of the trait, where the prediction accuracy is lower in complex traits [121].

Utilization of molecular technologies that have revolutionized commercial crop breeding can be used as a proof of concept for adoption of such genomics-based prediction methodologies [122,123] to improve trait performance in other less-studied crops [115,116]. These approaches are being adopted in crops of importance in developing countries such as in maize and wheat [121], rice [124], pulses (legumes) [11], cassava [118,120], cowpea [125], lentil [126], soybean

[127,128], and pigeon pea [129]. With respect to the best practice for GS, various models are being put forward [113]. Below is the rapid cycling breeding scheme for cassava, a long cycle clonally propagated crop (Figure 1).

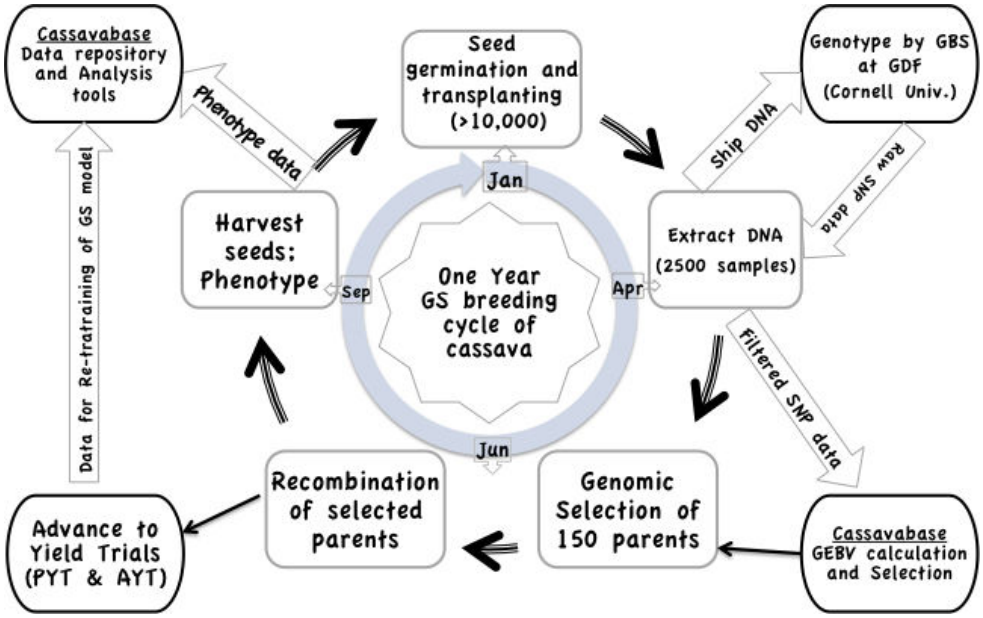

Figure 1. An overview of genomic selection-based annual breeding cycle implemented for cassava at the International Institute of Tropical Agriculture (IITA) in Nigeria. In June, crossing blocks are planted with parents selected using genomic selection and crosses made between September and November. Mature seeds are germinated and transplanted in January under irrigation. DNA is extracted from seedlings in March for genotyping by sequencing at the Genomic Diversity Facility (GDF). Raw SNP data are released to "Cassavabase" for further processing. Genomic-estimated breeding values (GEBVs) are then calculated and used to select candidate parents for the next recombination cycle. The remaining clones are also evaluated in clonal evaluation yield trials for variety development as well as for re-training the GS prediction model. **Cassavabase** (www.cassavabase.org): A bioinformatics infrastructure that integrates phenotypic data from field trials, genotypic data, as well as statistical tools in a single, user-friendly, web-based, and reliable database [130]. Breeders can use the intuitive web-based interphase to calculate genomic-estimated breeding values (GEBVs) of individuals by selecting a training population for modeling and estimating genomic-estimated breeding values of selection candidates (http://cassavabase.org/solgs). **GDF:** Genomic Diversity Facility (http://www.biotech.cornell.edu/brc/genomic-diversity-facility) provides expertise and state-of-the-art support for genotyping by sequencing (GBS) projects, including project optimization, library production, DNA sequencing, and data analysis.

It has now become evident that with advances in genotyping, fueled by NGS, phenotyping has become the rate-limiting step in genomics-enabled breeding. Concomitant development in phenotyping speed and precision is pivotal to associate genome with phenome [131] and to enable routine cost-effective high-throughput precision phenotyping. Approaches to increase throughput and quality of phenotyping range from automated and mechanized field experiment management, digital data capture, improved sample tracking methods, to deployment of ground-based and aerial advanced technologies in imaging and remote sensing [132–135].

http://dx.doi.org/10.5772/61665

Precision phenotyping has led to accelerated genetic gain by increasing heritability, mainly through reducing environmental variation [116,131], and reduced cost of trait measurement. Furthermore, robust and standardized screening protocols and the establishment of phenotyping hubs for abiotic (drought, nutrient use efficiency) and biotic (pest and disease hotspots) stresses are key elements for precision phenotyping to dissect the genetics of quantitative traits.

Leveraging existing data management and decision support tools to accommodate new data types and analytical tools, including digitized data collection (e.g., personal digital assistant (PDA), electronic field books) and sample tracking using bar codes, will be keys to the ultimate success of genomics breeding in developing countries.

3.2. Genetic resource management and utilization

Genebanks play an important role in safeguarding crop genetic diversity against the ongoing loss. They provide genetic variation for breeding for continued adaptation to changing environmental conditions and consumer demands [136,137]. The recent progress in DNA sequencing technologies that require less investment for generating large data is an opportunity to further investigate genetic variation maintained in the large germplasm collections held in trust by the CGIAR and increase the efficiency of genebanks. The 11 genebanks of the CGIAR conserve over 666,000 accessions of mainly food crops [138]. The International Institute of Tropical Agriculture (IITA) maintains over 28,000 accessions of major food crops of Africa, namely cowpea (*Vigna unguiculata*), cassava (*Manihot esculenta*), yam (*Dioscorea* spp.), soybean (*Glycine max*), bambara groundnut (*Vigna subterranea*), maize (*Zea mays*), and plantain and banana (*Musa* spp.). The aforementioned, including other important crops in developing countries [e.g., finger millet (*Eleusine coracana*), tef (*Eragrostis tef*), enset (*Ensete ventricosum*), grass pea (*Lathyrus sativus*) and their wild relatives], were considered understudied [2]. Large-scale characterization of all accessions and other genetic stocks is imperative to stimulate their utilization in breeding programs [139,140].

Traditionally, genebanks have used morphological descriptors for germplasm characterization; however, these are highly influenced by environmental conditions and different stages of plant development [141]. Moreover, the number of descriptors can be quite limited, thus greatly reducing the power to distinguish consanguineous varieties [142]. Molecular marker technologies have been widely applied for characterization and utilization of germplasm in genebanks [143]. However, the marker systems used prior to the advent of NGS, which sample a subset of the genome, have restricted applications mainly because of their limited abundance in the genome. NGS has enabled marker analysis at a much higher density. NGS-based genotyping, such as GBS, has been used for genetic diversity assessment of cultivated yam and its wild relatives [144] and cocoa [145], as well as other crop species. Breeding programs in the public and private sector deploy whole-genome fingerprinting of inbreds, to get an insight into the haplotype-level genetic diversity [116,140,146].

The advance in sequencing technologies is an advantage for efficient sequencing of large collections that include poorly studied species in genebanks with larger analytical power than the conventional molecular marker systems. Diversity assessments per se have huge utility in terms of germplasm utilization, such as definition of heterotic groups that enable breeders to make decisions in planning crosses for the population development. In addition to diversity assessment, NGS-based technologies are likely to impact further analysis of genetic variation, in terms of characterization of functional genetic diversity [148] and can be applied to pre-breeding activities to boost utilization of genetic resources in breeding programs [29,52,147].

NGS can also be applied to enhance management aspects of the genebanks, including identifying duplicates and identification of mislabeled accessions, both of which are common challenges in genebanks [148]. Diversity assessments using NGS could help guide the need for further targeted germplasm collection and improve the development of subsets of the collection, also referred to as core or minicore or diversity research sets, that would further improve the efficient utilization of germplasm for cultivar development.

A strong genomics and bioinformatics platform will greatly facilitate essential elements of genebank management, particularly the verification of accession identity, characterization of duplicates in the collection, and diversity analysis. Furthermore, rapid genotyping methods (e.g., GBS and WGS) will be essential for allele mining and large-scale association of genotype–phenotype, which are taken together with methods of developing trait-specific subsets, also referred to as core or mini core or diversity research sets, to greatly enhance the value of the collections for breeding and research. In particular gene pool, enhancement (pre-breeding) will be strengthened in terms of both base broadening within a species and use of crop wild relatives for the integration of key traits. Such approaches can be applied not only to staple crops but also to obtain rapid advances in the improvement of underutilized and under-researched but important crops such as cocoyam, winged bean, and African yam bean.

3.3. Breeding data management

The adoption of new Omics technologies by breeding programs in developing countries can contribute to the enhancement of breeding efficiency. There is a growing effort to harness advances in bio-computational methods and information and communication technology (ICT) to successfully utilize diverse phenotypic, environmental, genomic, and other metadata to provide decision support tools at various stages of the breeding pipeline. Modern breeding schemes such as GS and MAS involve a deluge of genotype data such as GBS-derived SNP markers, advanced statistical analysis to compute GEBV, and large amounts of high-throughput phenotype information, all of which require efficient informatics tools, automated data analysis pipelines, and decision-making tools for analysis and integration. Efficient utilization of such unprecedented volumes of genotypic, phenotypic, and other data entails development of informatics, database, and decision support tools.

http://dx.doi.org/10.5772/61665

Access to affordable genotyping platform by scientists in developing countries has been realized through various bilateral research-for-development projects. However, it is inconceivable to make progress without modern breeding tools and management processes that will facilitate the integration, analysis, and decision-making tools. One initiative that aims at providing some of these tools is the breeding management system (BMS) developed and promoted by the integrated breeding platform (IBP) (https://www.integratedbreeding.net/breeding-management-system). The service of BMS is delivered by IBP regional hubs that are strategically located throughout developing countries and hosted by partner research institutions such as IITA in Nigeria. The hubs provide support for adoption, customization, and use of BMS and related services, mainly through capacity building, technical support, and crop-specific expertise. Presently, IBP comprises ready-to-use information and tools for over 10 crops, including diagnostic markers and trait dictionaries.

In today's Omics era, web-based, peer-reviewed molecular databases and web servers abound [149]. An annual issue of the journal "Nucleic Acid Research" is dedicated to databases and web servers and documents a wide spectrum of databases, including a substantial number on plant databases. A comprehensive list of genomic resources (platforms and databases) relevant to genomics-enabled crop improvement, including genome sequences of crop plants, has been published recently [12]. Table 2 provides a partial list of deployed or planned breeding-relevant technology and tools currently in use. The Kazusa marker database [150] features genomics and genetics information for 10 plant species, whereas SolGenomics is a portal for several solanaceous plant species [130]. These and other breeders' toolboxes such as Soybase and MaizeGDB can serve as a starting point for comparative analysis of orphan crops with limited genomic resources.

Developments of several other similar and complementary custom-made breeding toolboxes are underway in various projects implemented in developing countries. A concerted effort by multidisciplinary teams, galvanized by various consortium research programs (CRPs), including national programs, are diligently working on development of pipelines for connecting diverse types of data to appropriate analytical tools and for processing imaging and remote sensing phenotype data.

The multidisciplinary nature of modern plant breeding/genetic research is underpinned by acquisition, analysis, and utilization of "big data" not only from field trials but also from laboratory analyses. Laboratory analysis includes analytical chemistry for profiling nutritional content and other metabolites, which entails efficient data management system. Moreover, high-density genome-wide marker data generated from next-generation sequencing for marker–trait associations as well as whole-genome expression profiling are increasingly being utilized for crop improvement pipelines. A comprehensive open-access database comprising phenotype and marker data, trial design, and analysis pipeline is a must-have to aid in streamlined integration of various data from plant breeding, including phenotypes recorded from field trials; genotypic data, gene expression, and analytical chemistry requires reliable and user-friendly database. Such a database must also have inbuilt quantitative genetics analysis tools/pipelines that would allow breeders to not only store and retrieve raw data but also calculate breeding values and selection index, design crosses, as well as field trials.

Moreover, discovery research such as QTL mapping can be done on the database through implementation of genetic mapping methods.

Project/Host	Database/Tool	Purpose	URL	Remark/ Reference
Integrated breeding platform	Breeding management system (BMS*)	Tools for Crop information management Nursery and trial management Statistical analysis Marker-assisted breeding	https:// www.integratedbreeding.net/	Current regional hubs: 4 in Africa, 3 in Asia
Cassavabase	NextGen cassava breeding project; Boyce Thompson Institute for Plant Research	Breeders toolbox; maps and markers; genes; phenotypes; genome sequences	http:// www.cassavabase.org/	Implemented based on SolGenomics
SolGenomics	Sol Genomics Network, Boyce Thompson Institute for Plant Research	Tomato, pepper, potato, coffee, Nicotiana, Petunia, and other solanaceous plants	http://solgenomics.net/	[159]
Soybase	USDA, Soybean Genetics Database Iowa State University	Soybean breeder's toolbox and database including genome sequences, maps, markers, genetic stocks (including mutants)	http:// www.soybase.org/	[160]
MaizeGDB	USDA funded maize genetics and genomics database	Community-oriented informatics service featuring genome browser, maps, locus, gene, QTL, diversity, metabolic pathways and others	http://maizegdb.org/	[161]
Phytozome	Department of Energy's Joint Genome Institute	The Plant Comparative Genomics portal for sequenced and annotated green plant genomes and phylogenetics	http:// phytozome.jgi.doe.gov/ pz/portal.html	[162]
Kazusa	Kazusa DNA Research Institute	SSR markers and linkage maps for 10 plant species	http:// marker.kazusa.or.jp	[158]

*BMS, hosted by IITA as a regional hub for integrated breeding platform (IBP), is a suite of interconnected software specifically designed to help breeders manage their day-to-day activities through all phases of their breeding programs.

Note: Other CGIAR-driven initiatives include Genomic and Open-source Breeding Informatics Initiative (GOBII), Integrated Genotyping Service and Support at Biosciences eastern and central Africa (BECA)/International Livestock Research Institute (ILRI), and Shared Industrial-Scale High-Throughput Genotyping Facility for delivering high-density genomics breeder's tools and low-cost genotyping services.

Table 2. Partial list of crop- or project-specific databases and breeder's toolboxes relevant to breeders in developing nations that are in use or in progress.

4. Disease diagnostics and monitoring

Plant diseases are caused by a wide array of pathogens, including viruses, bacteria, and fungi. A combination of techniques, including microscopy, serological [e.g., enzyme-linked immu-

http://dx.doi.org/10.5772/61665

nosorbent assay (ELISA)], and molecular (e.g., PCR) techniques, are used in detection and identification of pathogens associated with major diseases of African food staples. Conventional methods of virus diagnostics, using antibodies and PCR, often lack the sensitivity to detect viruses that exist in low abundance and emerging viruses with unknown genomes. Therefore, next-generation deep sequencing approaches and bioinformatics analysis can be used for de novo assembly of virus and viroid genomes, to perform reliable characterization and diagnostics of known and unknown viruses and viroids [112,154,155]. In the wake of NGS technologies, powerful and high-throughput novel approaches, such as metagenomics, have been developed and widely used to analyze nucleotide sequence of microbial populations in plant samples (see section 2.8) [8,105,156]. In particular, deep sequencing of small RNA families such as short interfering RNAs (siRNAs) can be used to identify and reconstruct any DNA or RNA virus genome and its microvariants with the help of bioinformatics tools [155,157]. Furthermore, the application of NGS can be extended to insect vectors for discovery and characterization of insect viruses [109].

The potential use of NGS technologies for diagnostic programs in quarantine and certification of some fruits have been demonstrated (reviewed in [8]). Existing diagnostics tools that are deployed in several clonally propagated crops (cassava, yam, banana) for quarantine monitoring during exchange of planting material can be enhanced using NGS. In IITA, diagnostic tools have been combined with digital data capture tools for real-time surveillance and rapid diagnosis. This has been put to use for monitoring pathogens of cassava and banana in East Africa.

5. Conclusions: Prospects and perspectives

The productivity of staple food crops of hundreds of millions of people in developing countries is stagnating or diminishing as natural resources are depleted as a result of overcultivation and poor resource management, among other factors. Genetic improvement is heralded as the best option to enhance crop productivity, resilience to climate effects, and nutritional quality. The effective and efficient application of advanced biosciences tools and products holds substantial promise for enhanced agricultural productivity, improved livelihoods, and better prospects for food and nutrition security in Africa, where less-studied crops are grown as staples [114,115,158]. Genomics-enabled breeding will enable scientists to more effectively tap into the wealth of genetic variation in landraces and wild relatives for novel traits.

Next-generation sequencing has evolved to the third generation of sequencing technology and boasts even longer read length, shorter run time, and lower cost per unit data [21]. Applications of NGS are broadening at a remarkable pace from whole-genome sequencing and re-sequencing to transcript sequencing, metagenomics, and methylome sequencing. Thus, the application of NGS in agriculture is now vital to breeding, diagnosis, evolution, ecology, and basic functional genomics. SNP markers are already becoming the predominant marker types in modern breeding strategies [21,29]. Additional outcomes include the dissection of biochemical and genetic mechanisms or metabolic pathways underlying agronomically important traits, leading to a better understanding of how the genome and phenome are related [114].

The ultrahigh-throughput capacity of NGS platforms and the commercial scale of automated pipelines make it cheaper to outsource genotyping services such as GBS and RAD. Capital investment in state-of-the-art genomics facilities in all laboratories is not prudent for several reasons. However, establishment of shared resources at regional and subregional center of excellence, such as BECA, is fully recognized by stakeholders. The West Africa Biotechnology Initiative (WABI), copromoted by IITA and subregional organizations such as CORAF/WECARD (West and Central African Council for Agricultural Research and Development), is promoting such an idea and mobilizing resources toward this goal. This is likely to reduce turnaround times for GBS samples, and raise the quality of cDNA libraries.

Mainstreaming this highly promising but complex and rapidly evolving next-generation breeding scheme entails continuous training and effective information sharing. Although recent scientific progress heralded the era of molecular breeding, most public sector researchers in Africa are far from harvesting the fruit of the technological advances.

Reasons for this range from limited awareness of the technological advances to lack of adequate infrastructure, knowledge, and limited resources that are required to make use of markers in crop breeding. In recent times, that trend is changing as research institutions operating in Africa (international, regional, and national systems) strive relentlessly to accelerate the adoption and application of advanced biosciences tools in support of the region's agricultural transformation. WABI is striving to establish a center of excellence to promote the adoption of biotechnology to enable innovative approaches, resulting in increased crop yield. Availability of training and service platforms in various subregions of Africa (e.g., West and Central, East and South) will not only make it more affordable and accessible to the users and trainees in the continent but also focus more on the needs that are specific to the region's research.

Developing in-house capacity for GBS data analysis pipeline, NGS library construction, and automated DNA extraction is fundamental for routine applications of GS/MAS in breeding programs. The spectacular diffusion of ICT throughout Africa, particularly mobile phone technology and smart devices, paves the way for access to web-based education and genomic resources. Given the poor connectivity in developing countries, however, developing Internet-free databases and tools is necessary in the interim.

Efficient data management systems are a prerequisite for applying genomic information by international, national, and private sectors involved in improving the rate of genetic gain in crops. WGS and assembly require advanced instruments, skilled personnel, and strong computational capacities. It also requires improvement of assembly and continual annotation of genes as more and more information is generated by whole-genome re-sequencing or functional genomics. Integration of genomics information with other phenotypic and environment data also requires strong skill in programming and database development. Moreover, processing of big data requires basic programming skills in order to automate routine data manipulation and processing needs. Thorough knowledge in bioinformatics will afford the ability to apply comparative genomics with the aim of extending the power of genomics to orphan crops with little DNA sequence information.

http://dx.doi.org/10.5772/61665

The bioinformatics infrastructure at IITA can serve as a model for similar start-up bioinformatics units at the national program. Such platform hosts most of the standard bioinformatics tools to deal with any kind of sequence analysis, including shotgun and targeted DNA/RNA sequences. Importantly, analysis pipeline for GBS data is very essential for routine application of genomics in selection schemes.

Such an effort demands full engagement and transformation in the policy of national programs and other stakeholders. As expressed in previous views [52,159], relevant short-term and long-term training and institutional capacity building should be intensified. Academic institutions need to revise their curricula to develop expertise in NGS data analysis and bioinformatics. The participation of the fledging private sector also needs to be boosted.

It is clear that certain activities such as efficient DNA extraction and associated databases and decision-making breeding tools may need to operate at local levels; other activities such as GBS, SNP genotyping for forward breeding, NGS, and training may need to operate at regional levels; and curation of whole crop databases and development of analysis tools may operate at global levels. It is vital that communication occurs at all of these levels and across levels, including international institutes, NARS, and universities, and that the system remains responsive to the rapidly changing scientific environment, if NGS is to close the yield gap of staple crops in Africa.

6. Acronyms

AOCC; African Orphan Crops Consortium

BMS; Breeding management system

BS; Bisulfide

Cas9; CRISPR-associated protein 9

CBSD; Cassava brown streak disease

CBSV; Cassava brown streak virus

CGIAR; The Consortium of International Agricultural Research

CORAF/

WECARD; West and Central African Council for Agricultural Research and Development

CRISPR; Clustered regularly interspaced short palindromic repeat

CRP; Consortium research programs

DArT; Diversity Array Technology

DSB; Double-strand breaks

GBS; Genotyping by sequencing

GDF; Genomic Diversity Facility

GEBV; Genomic-estimated breeding value

GS; Genome selection

GWAS; Genome-wide association study

IBP; Integrated breeding platform

ICT; Information and communication technology

IITA; International Institute of Tropical Agriculture

KASP; Kompetitive Allele-Specific PCR

MAS; Marker-assisted selection

NARS; National agricultural research systems

ncRNA; Noncoding RNA

NGS; Next-generation sequencing

NHEJ; Nonhomologous end joining

PDA; Personal digital assistant

QTL; Quantitative trait loci

R4D; Research for development

RAD; Restriction-site-associated DNA

RGN; RNA-guided nucleases

RRS; Reduced representation sequencing

RTB; Root, tuber, and banana

siRNA; small interfering RNA

SNP; Single nucleotide polymorphism

TALENs; Transcription activator–like effector nucleases

TILLING; Targeting Induced Local Lesions in Genomes

WGS; Whole-genome sequencing

ZFN; Zinc finger nuclease

WABI; The West Africa Biotechnology Initiative (WABI),

Author details

Melaku Gedil[1*], Morag Ferguson[1], Gezahegn Girma[1], Andreas Gisel[1,2], Livia Stavolone[1,3] and Ismail Rabbi[1]

*Address all correspondence to: m.gedil@cgiar.org

1 International Institute of Tropical Agriculture, Ibadan, Nigeria

2 Institute for Biomedical Technologies – CNR, Bari, Italy

3 Institute for Sustainable Plant Protection – CNR, Bari, Italy

References

[1] Gedil M, Sartie AM. Perspectives on molecular breeding of Africa's main staple food crops – cassava and yam. Asp Appl Biol 2010;96:123–36.

[2] Varshney RK, Glaszmann J-CC, Leung H, Ribaut J-MM. More genomic resources for less-studied crops. Trends Biotechnol 2010;28:452–60. doi:10.1016/j.tibtech.2010.06.007.

[3] Ribaut J-M, de Vicente MC, Delannay X. Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 2010;13:213–18.

[4] Gedil M, Tripathi L, Ghislain M, Ferguson M, Ndjiondjop M, Kumar PL, et al. Biotechnology success stories by the Consultative Group on International Agriculture Research (CGIAR) system. In: Wambugu F, Kamanga D, editors. Biotechnology in Africa: Emergence, Initiative and Future. Springer, Heidelberg. 2014, p. 95-114.

[5] Armstead I, Huang L, Ravagnani A, Robson P, Ougham H. Bioinformatics in the orphan crops. Brief Bioinform 2009;10:645–53.

[6] Varshney RK, Ribaut J-MM, Buckler ES, Tuberosa R, Rafalski JA, Langridge P. Can genomics boost productivity of orphan crops? Nat Biotechnol 2012;30:1172–76.

[7] Tadele Z, Esfeld K, Plaza S. Applications of high-throughput techniques to the understudied crops of Africa. Asp Appl Biol 2010;96:233–40.

[8] Barba M, Czosnek H, Hadidi A. Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses 2014;6:106–36. doi: 10.3390/v6010106.

[9] Rius M, Bourne S, Hornsby HG, Chapman MA. Applications of next-generation sequencing to the study of biological invasions. Curr Zool 2015;61:488–504.

[10] Nybom H, Weising K, Rotter B. DNA fingerprinting in botany: past, present, future. Investig Genet 2014;5:1-35. doi:10.1186/2041-2223-5-1.

[11] Bohra A, Pandey MK, Jha UC, Singh B, Singh IP, Datta D, et al. Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. Theor Appl Genet 2014;127:1263–91. doi:10.1007/s00122-014-2301-3.

[12] Dhanapal AP, Govindaraj M. Unlimited thirst for genome sequencing, data interpretation, and database usage in genomic era: the road towards fast-track crop plant improvement. Genet Res Int 2015;2015:1–15.

[13] Michael TP, Vanburen R. Progress, challenges and the future of crop genomes. Curr Opin Plant Biol 2015;24:71-81.

[14] Rowe HC, Renaut S, Guggisberg A. RAD in the realm of next-generation sequencing technologies. Mol Ecol 2011. 20:3499-3502. doi:10.1111/j.1365-294X.2011.05197.x.

[15] Leggett RM, MacLean D. Reference-free SNP detection: dealing with the data deluge. BMC Genomics 2014;15:S10. doi:10.1186/1471-2164-15-S4-S10.

[16] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 2008;3:1–7. doi:10.1371/journal.pone.0003376.

[17] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 2011;6:1–10. doi:10.1371/journal.pone.0019379.

[18] Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol Biol 2012;888:67–89. doi:10.1007/978-1-61779-870-2_5.

[19] He J, Zhao X, Laroche A, Lu Z, Liu H, Li Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding 2014;5:1–8. doi:10.3389/fpls.2014.00484.

[20] Hamblin MT, Rabbi IY. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: a study in cassava (*Manihot esculenta*). Crop Sci 2014;54:2603–08. doi:10.2135/cropsci2014.02.0160.

[21] Thomson MJ. High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2014;2:195–212. doi:10.9787/PBB.2014.2.3.195.

[22] Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 2014;33:1–14. doi:10.1007/s11032-013-9917-x.

[23] Yang H, Tao Y, Zheng Z, Li C, Sweetingham M, Howieson J. Application of next-generation sequencing for rapid marker development in molecular plant breeding: a

case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genomics 2012;13:318. doi:10.1186/1471-2164-13-318.

[24] Devran Z, Kahveci E, Özkaynak E, Studholme DJ, Tör M. Development of molecular markers tightly linked to Pvr4 gene in pepper using next-generation sequencing. Mol Breed 2015;35:101. doi:10.1007/s11032-015-0294-5.

[25] Rabbi IY, Hamblin MT, Kumar PLL, Gedil MA, Ikpan AS, Jannink JL, et al. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Res 2014;186:87–96. doi:10.1016/j.virusres.2013.12.028.

[26] Rabbi I, Hamblin M, Gedil M, Kulakow P, Ferguson M, Ikpan AS, et al. Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. Crop Sci 2014;54:1384–96. doi:10.2135/cropsci2013.07.0482.

[27] Hart JP, Griffiths PD. Genotyping-by-sequencing enabled mapping and marker development for the potyvirus resistance allele in common bean. Plant Genome 2015;8:1-14. doi:10.3835/plantgenome2014.09.0058.

[28] Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One 2013;8:1–9. doi:10.1371/journal.pone.0054603.

[29] Egan AN, Schlueter J, Spooner DM. Applications of next-generation sequencing in plant biology. Am J Bot 2012;99:175–85. doi:10.3732/ajb.1200020.

[30] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63. doi:10.1038/nrg2484.

[31] Lopez C, Jorge V, Piégu B, Mba C, Cortes D, Restrepo S, et al. A unigene catalogue of 5700 expressed genes in cassava. Plant Mol Biol 2004;56:541–54. doi:10.1007/s11103-004-0123-4.

[32] Anderson JV, Delseny M, Fregene MA, Jorge V, Mba C, Lopez C, et al. An EST resource for cassava and other species of Euphorbiaceae. Plant MolBiol 2004;56:527–39. doi:10.1007/s11103-004-5046-6.

[33] Reilly K, Bernal D, Cortés DF, Gómez-Vásquez R, Tohme J, Beeching JR. Towards identifying the full set of genes expressed during cassava post-harvest physiological deterioration. Plant Mol Biol 2007;64:187–203. doi:10.1007/s11103-007-9144-0.

[34] Yang J, An D, Zhang P. Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. J IntegrPlant Biol 2011;53:193–211.

[35] Sakurai T, Plata G, Rodríguez-Zapata F, Seki M, Salcedo A, Toyoda A, et al. Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific

expansions in gene families related to stress response. BMC Plant Biol 2007;7:66. doi: 10.1186/1471-2229-7-66.

[36] Utsumi Y, Tanaka M, Morosawa T, Kurotani A, Yoshida T, Mochida K, et al. Transcriptome analysis using a high-density oligomicroarray under drought stress in various genotypes of cassava: an important tropical crop. DNA Res 2012;19:335–45. doi: 10.1093/dnares/dss016.

[37] Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One 2014;9:e96642. doi:10.1371/journal.pone.0096642.

[38] Allie F, Pierce EJ, Okoniewski MJ, Rey C. Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. BMC Genomics 2014;15:1–30.

[39] Chen X, Xia J, Xia Z, Zhang H, Zeng C, Lu C, et al. Potential functions of microRNAs in starch metabolism and development revealed by miRNA transcriptome profiling of cassava cultivars and their wild progenitor. BMC Plant Biol 2015;15:1–11. doi: 10.1186/s12870-014-0355-7.

[40] Zeng C, Wang W, Zheng Y, Chen X, Bo W, Song S, et al. Conservation and divergence of microRNAs and their functions in Euphorbiaceous plants. Nucleic Acids Res 2009;38:981–95. doi:10.1093/nar/gkp1035.

[41] Tamiru M, Natsume S, Takagi H, Babil PK, Yamanaka S, Lopez-Montes A, et al. Whole genome sequencing of Guinea yam (*Dioscorea rotundata*). First Glob. Conf. Yam, Accra, Ghana: International Institute of Tropical Agriculture; Ibadan, Nigeria. 2013. p.20

[42] Mutuku JM, Yoshida S, Shimizu T, Ichihashi Y, Wakatake T, Seo M, et al. The WRKY45-dependent signaling pathway is required for resistance against Striga parasitism. 2015;168: 1152-1163. doi:10.1104/pp.114.256404.

[43] Chapman MA. Transcriptome sequencing and marker development for four underutilized legumes. Appl Plant Sci 2015;3:1400111. doi:10.3732/apps.1400111.

[44] Liu S, Yeh C-T, Tang HM, Nettleton D, Schnable PS. Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 2012;7:e36406.

[45] Goyer A, Hamlin L, Crosslin JM, Buchanan A, Chang JH. RNA-seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. BMC Genomics 2015;16:472. doi:10.1186/s12864-015-1666-2.

[46] Humbert S, Subedi S, Cohn J, Zeng B, Bi Y-M, Chen X, et al. Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses. BMC Genomics 2013;14:3. doi:10.1186/1471-2164-14-3.

[47] Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP. In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res 2009;16:353–69. doi:10.1093/dnares/dsp023.

[48] Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014;30:418–26. doi:10.1016/j.tig.2014.07.001.

[49] Grover CE, Salmon A, Wendel JF. Targeted sequence capture as a powerful tool for evolutionary analysis. Am J Bot 2012;99:312–19. doi:10.3732/ajb.1100323.

[50] Reeves GA, Talavera D, Thornton JM. Genome and proteome annotation: organization, interpretation and integration. J R Soc Interface 2009;6:129–47. doi:10.1098/rsif.2008.0341.

[51] Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, et al. Insights into the maize pan-genome and pan-transcriptome. Plant Cell 2014;26:121–35. doi:10.1105/tpc.113.119982.

[52] Varshney RK, Terauchi R, McCouch SR. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 2014;12:1–8. doi:10.1371/journal.pbio.1001883.

[53] Bzhalava D. Bioinformatics for viral metagenomics. J Data Mining Genomics Proteomics 2013;4:3–7. doi:10.4172/2153-0602.1000134.

[54] Lindner MS, Renard BY. Metagenomic abundance estimation and diagnostic testing on species level. Nucleic Acids Res 2013;41:1–8. doi:10.1093/nar/gks803.

[55] Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 2014;5:209. doi:10.3389/fpls.2014.00209.

[56] Schatz MC, Witkowski J, McCombie WR. Current challenges in de novo plant genome sequencing and assembly. Genome Biol 2012;13:243. doi:10.1186/gb4015.

[57] Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. Bioinformatics 2012;28:3169–77. doi:10.1093/bioinformatics/bts605.

[58] Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rätsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods 2013;10:1185–91. doi:10.1038/nmeth.2722.

[59] Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 2014;157:77–94. doi:10.1016/j.cell.2014.03.008.

[60] Bond DM, Baulcombe DC. Small RNAs and heritable epigenetic variation in plants. Trends Cell Biol 2014;24:100–07. doi:10.1016/j.tcb.2013.08.001.

[61] Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet 2012;13:705–19. doi:10.1038/nrg3273.

[62] Meaburn E, Schulz R. Next generation sequencing in epigenetics: insights and challenges. Semin Cell Dev Biol 2012;23:192–99. doi:10.1016/j.semcdb.2011.10.010.

[63] Teeling H, Glöckner FO. Current opportunities and challenges in microbial metagenome analysis – a bioinformatic perspective. Brief Bioinform 2012;13:728–42. doi: 10.1093/bib/bbs039.

[64] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics 2007;23:2633–55.

[65] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–98. doi:10.1038/ng.806.

[66] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–58. doi:10.1093/bioinformatics/btr330.

[67] Li Y, Tollefsbol TO. DNA methylation detection: bisulfite genomic sequencing analysis. Methods Mol Biol 2011;791:11–21. doi:10.1007/978-1-61779-316-5_2.

[68] Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. Nat Methods 2012;9:145–51. doi:10.1038/nmeth.1828.

[69] Hagen C, Frizzi A, Kao J, Jia L, Huang M, Zhang Y, et al. Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses. Arch Virol 2011;156:1209–16. doi:10.1007/s00705-011-0979-y.

[70] Araki M, Ishii T. Towards social acceptance of plant breeding by genome editing. Trends Plant Sci 2015;20:1–5. doi:10.1016/j.tplants.2015.01.010.

[71] Andersen MM, Landes X, Xiang W, Anyshchenko A, Falhof J, Østerberg JT, et al. Feasibility of new breeding techniques for organic farming. Trends Plant Sci 2015;20:426–34. doi:10.1016/j.tplants.2015.04.011.

[72] Carroll D. Genome engineering with zinc-finger nucleases. Genetics 2011;188:773–82. doi:10.1534/genetics.111.131433.

[73] Carroll D. A CRISPR approach to gene targeting. Mol Ther 2012;20:1658–60. doi: 10.1038/mt.2012.171.

[74] Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 2013;14:49–55. doi:10.1038/nrm3486.

[75] Osakabe Y, Osakabe K. Genome editing with engineered nucleases in plants. Plant Cell Physiol 2015;56:389–400. doi:10.1093/pcp/pcu170.

[76] Fichtner F, Urrea Castellanos R, Ülker B. Precision genetic modifications: a new era in molecular biology and crop improvement. Planta 2014;239:921–39. doi:10.1007/s00425-014-2029-y.

[77] Kumar V, Jain M. The CRISPR-Cas system for plant genome editing: advances and opportunities. J Exp Bot 2015;66:47–57. doi:10.1093/jxb/eru429.

[78] Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, et al. Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 2012;161:20–27. doi:10.1104/pp.112.205179.

[79] Zhang F, Maeder ML, Unger-Wallace E, Hoshaw JP, Reyon D, Christian M, et al. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. Proc Natl Acad Sci U S A 2010;107:12028–33. doi:10.1073/pnas.0914991107.

[80] Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. Nucleic Acids Res 2013;41:e188. doi:10.1093/nar/gkt780.

[81] Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, et al. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 2009;459:437–41. doi:10.1038/nature07992.

[82] Marton I, Zuker A, Shklarman E, Zeevi V, Tovkach A, Roffe S, et al. Nontransgenic genome modification in plant cells. Plant Physiol 2010;154:1079–87. doi:10.1104/pp.110.164806.

[83] Palmgren MG, Edenbrandt AK, Vedel SE, Andersen MM, Landes X, Østerberg JT, et al. Are we ready for back-to-nature crop breeding? Trends Plant Sci 2015;20:155–64.

[84] Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biol 2014;14:327. doi:10.1186/s12870-014-0327-y.

[85] Comai L, Henikoff S. TILLING: practical single-nucleotide mutation discovery. Plant J 2006;45:684–94. doi:10.1111/j.1365-313X.2006.02670.x.

[86] McCallum CM, Comai L, Greene EA, Henikoff S. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiol 2000;123:439–42. doi:10.1104/pp.123.2.439.

[87] Waugh R, Leader DJ, McCallum N, Caldwell D. Harvesting the potential of induced biological diversity. Trends Plant Sci 2006;11:71–79. doi:10.1016/j.tplants.2005.12.007.

[88] Mba C, Afza R, Jankowicz-Cieslak J, Bado S, Matijevic M, Huynh O, et al. Enhancing genetic diversity through induced mutagenesis in vegetatively propagated plants. In: Shu Q, editor. Induc. Plant Mutat. Genomics Era, Food and Agriculture Organization of the United Nations, Rome: 2009, p. 262–65.

[89] Tadele Z, Mba C, Till BJ. TILLING for mutations in model plants and crops. In: Jain S, Brar S, editors. Mol. Tech. Crop Improv. 2nd ed., Rome, Springer Netherlands; 2010, p. 307–32.

[90] Gilchrist E, Haughn G. Reverse genetics techniques: engineering loss and gain of gene function in plants. Briefings Funct Genomics Proteomics 2010;9:103–10. doi: 10.1093/bfgp/elp059.

[91] Till BJ, Zerr T, Comai L, Henikoff S. A protocol for TILLING and Ecotilling in plants and animals. NatProtoc 2006;1:2465–77. doi:10.1038/nprot.2006.329.

[92] Tsai H, Howell T, Nitcher R, Missirian V, Watson B, Ngo KJ, et al. Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiol 2011;156:1257–68. doi:10.1104/pp.110.169748.

[93] Zhu Q, Smith SM, Ayele M, Yang L, Jogi A, Chaluvadi SR, et al. High-throughput discovery of mutations in tef semi-dwarfing genes by next-generation sequencing analysis. Genetics 2012;192:819–29. doi:10.1534/genetics.112.144436.

[94] Pan L, Shah AN, Phelps IG, Doherty D, Johnson EA, Moens CB. Rapid identification and recovery of ENU-induced mutations with next-generation sequencing and paired-end low-error analysis. BMC Genomics 2015;16:1–13. doi:10.1186/s12864-015-1263-4.

[95] Marroni F, Pinosio S, Morgante M. The quest for rare variants: pooled multiplexed next generation sequencing in plants. Front Plant Sci 2012;3:1–9. doi:10.3389/fpls.2012.00133.

[96] Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 2012;30:174–78.

[97] Bernardo R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 2008;48:1649–64. doi:10.2135/cropsci2008.03.0131.

[98] Salvi S, Tuberosa R. The crop QTLome comes of age. Curr Opin Biotechnol 2015;32:179–85. doi:10.1016/j.copbio.2015.01.001.

[99] Xu F, Sun X, Chen Y, Huang Y, Tong C, Bao J. Rapid Identification of major QTLs associated with rice grain weight and their utilization. PLoS One 2015;10:e0122206. doi:10.1371/journal.pone.0122206.

[100] Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 2013;74:174–83. doi:10.1111/tpj.12105.

[101] Lu H, Lin T, Klein J, Wang S, Qi J, Zhou Q, et al. QTL-seq identifies an early flowering QTL located near Flowering Locus T in cucumber. Theor Appl Genet 2014;127:1491–99. doi:10.1007/s00122-014-2313-z.

[102] Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Laxmi, et al. Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. DNA Res 2015:1–11. doi:10.1093/dnares/dsv004.

[103] Han Y, Lv P, Hou S, Li S, Ji G, Ma X, et al. Combining next generation sequencing with bulked segregant analysis to fine map a Stem Moisture Locus in Sorghum (*Sorghum bicolor* L. Moench). PLoS One 2015;10:e0127065. doi:10.1371/journal.pone.0127065.

[104] Thomas T, Gilbert J, Meyer F. Metagenomics – a guide from sampling to data analysis. Microb Inform Exp 2012;2:3. doi:10.1186/2042-5783-2-3.

[105] Lebeis SL. Greater than the sum of their parts: characterizing plant microbiomes at the community-level. Curr Opin Plant Biol 2015;24C:82–86. doi:10.1016/j.pbi.2015.02.004.

[106] Schloss PD, Handelsman J. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol 2005;6:229. doi:10.1186/gb-2005-6-8-229.

[107] Kellner H, Luis P, Portetelle D, Vandenbol M. Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. Microbiol Res 2011;166:360–68. doi:10.1016/j.micres.2010.07.006.

[108] Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K, et al. Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. Appl Environ Microbiol 2014;80:1777–86. doi:10.1128/AEM.03712-13.

[109] Liu S, Vijayendran D, Bonning BC. Next generation sequencing technologies for insect virus discovery. Viruses 2011;3:1849–69. doi:10.3390/v3101849.

[110] Legg JP, Lava Kumar P, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, et al. Cassava virus diseases: biology, epidemiology, and management. Adv Virus Res 2015;91:85–142. doi:10.1016/bs.aivir.2014.10.001.

[111] Geering ADW, Olszewski NE, Harper G, Lockhart BEL, Hull R, Thomas JE. Banana contains a diverse array of endogenous badnaviruses. J Gen Virol 2005;86:511–20. doi:10.1099/vir.0.80261-0.

[112] Filloux D, Murrell S, Koohapitagtam M, Golden M, Julian C, Galzi S, et al. The genomes of many yam species contain transcriptionally active endogenous geminiviral sequences that may be functionally expressed. Virus Evol 2015;1:vev002–vev002. doi:10.1093/ve/vev002.

[113] Heslot N, Sorrels ME, Jannink J. Perspectives for genomic selection applications and research in plants. Crop Sci 2015;55:1–30. doi:10.2135/cropsci2014.03.0249.

[114] Poland J. Breeding-assisted genomics. Curr Opin Plant Biol 2015;24:119–24. doi:10.1016/j.pbi.2015.02.009.

[115] Rivers J, Warthmann N, Pogson BJ, Borevitz JO. Genomic breeding for food, environment and livelihoods. Food Secur 2015;7:375–82. doi:10.1007/s12571-015-0431-3.

[116] Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, et al. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 2014;65:311–36. doi:10.1071/CP14007.

[117] Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH. Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theor Appl Genet 2015. 128:1647-1667. doi: 10.1007/s00122-015-2555-4.

[118] De Oliveira EJ, de Resende MDV, da Silva Santos V, Ferreira CF, Oliveira GAF, da Silva MS, et al. Genome-wide selection in cassava. Euphytica 2012;187:263–76. doi: 10.1007/s10681-012-0722-0.

[119] Ray S, Satya P. Next generation sequencing technologies for next generation plant breeding. Front Plant Sci 2014;5:1–4. doi:10.3389/fpls.2014.00367.

[120] Ly D, Hamblin M, Rabbi I, Melaku G, Bakare M, Gauch HG, et al. Relatedness and genotype environment interaction affect prediction accuracies in genomic selection: a study in Cassava. Crop Sci 2013;53:1312–25.

[121] Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb) 2014;112:48–60. doi:10.1038/hdy.2013.16.

[122] Crossa J, Pérez P, de los Campos G, Mahuku G, Dreisigacker S, Magorokosho C. Genomic selection and prediction in plant breeding. J Crop Improv 2011;25:239–61. doi: 10.1080/15427528.2011.558767.

[123] Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, et al. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci 2014;54:1476-1488. doi:10.2135/cropsci2013.03.0195.

[124] Anacleto R, Cuevas RP, Jimenez R, Llorente C, Nissila E, Henry R, et al. Prospects of breeding high-quality rice using post-genomic tools. Theor Appl Genet 2015;128:1449–66. doi:10.1007/s00122-015-2537-6.

[125] Huynh B-L, Ehlers JD, Ndeve A, Wanamaker S, Lucas MR, Close TJ, et al. Genetic mapping and legume synteny of aphid resistance in African cowpea (*Vigna unguiculata* L. Walp.) grown in California. Mol Breed 2015;35:36. doi:10.1007/s11032-015-0254-0.

[126] Kumar S, Rajendran K, Kumar J, Hamwieh A, Baum M. Current knowledge in lentil genomics and its application for crop improvement. Front Plant Sci 2015;6:1–13. doi: 10.3389/fpls.2015.00078.

[127] Jarquín D, Kocak K, Posadas L, Hyma K, Jedlicka J, Graef G, et al. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 2014;15:740. doi:10.1186/1471-2164-15-740.

[128] Bao Y, Kurle JE, Anderson G, Young ND. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol Breed 2015;35:128. doi:10.1007/s11032-015-0324-3.

[129] Pazhamala L, Saxena RK, Singh VK, Sameerkumar CV, Kumar V, Sinha P, et al. Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). Front Plant Sci 2015;6:1–12. doi:10.3389/fpls.2015.00050.

[130] Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, et al. The Sol Genomics Network (SGN) – from genotype to phenotype to breeding. Nucleic Acids Res 2014;43:D1036–41. doi:10.1093/nar/gku1195.

[131] Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet 2013;126:867–87. doi:10.1007/s00122-013-2066-0.

[132] Rousseau D, Chéné Y, Belin E, Semaan G, Trigui G, Boudehri K, et al. Multiscale imaging of plants: current approaches and challenges. Plant Methods 2015;11:1–9. doi:10.1186/s13007-015-0050-1.

[133] Bergsträsser S, Fanourakis D, Schmittgen S, Cendrero-Mateo MP, Jansen M, Scharr H, et al. HyperART: non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging. Plant Methods 2015;11:1–17. doi:10.1186/s13007-015-0043-0.

[134] Araus L, Elazab A, Vergara O, Cabrera-Bosquet L, Serret MD, Zaman-Allah M, et al. New technologies for phenotyping. In: Fritsche-Neto R, Borem A, editors. Phenomics, Springer International Publishing, Switzerland; 2015. p.1-14. doi: 10.1007/978-3-319-13677-6_1.

[135] Zaman-Allah M, Vergara O, Araus JL, Tarekegne A, Magorokosho C, Zarco-Tejada PJ, et al. Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. Plant Methods 2015;11:35. doi:10.1186/s13007-015-0078-2.

[136] Bansal KC, Lenka SK, Mondal TK. Genomic resources for breeding crops with enhanced abiotic stress tolerance. Plant Breed 2014;133:1–11. doi:10.1111/pbr.12117.

[137] Dwivedi SL, Upadhyaya HD, Stalker HT, Blair MW, Bertioli DJ, Nielen S, et al. Enhancing crop gene pools with beneficial traits using wild relatives. Plant Breed Rev 2008;30:179–230.

[138] Koo B, Pardey PG, Wright BD. The economic costs of conserving genetic resources at the CGIAR centers. Agric Econ 2003;29:287–97. doi:10.1111/j.1574-0862.2003.tb00165.x.

[139] Lee S-H, Tuberosa R, Jackson SA, Varshney RK. Genomics of plant genetic resources: a gateway to a new era of global food security. Plant Genet Resour 2014;12:S2–5. doi: 10.1017/S1479262114000513.

[140] Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol 2013;14:R55. doi:10.1186/gb-2013-14-6-r55.

[141] Elhoumaizi MA, Saaidi M, Oihabi A, Cilas C. Phenotypic diversity of date-palm cultivars(*Phoenix dactylifera* L.) from Morocco. Genet Resour Crop Evol 2002;49:483–90. doi:10.1023/A:1020968513494.

[142] Duminil J, Di Michele M. Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosyst – An Int J Deal with All Asp Plant Biol 2009;143:528–42. doi:10.1080/11263500902722964.

[143] Börner A, Khlestkina EK, Chebotar S, Nagel M, Arif MAR, Neumann K, et al. Molecular markers in management of ex situ PGR – A case study. J Biosci 2012;37:871–77. doi:10.1007/s12038-012-9250-2.

[144] Girma G, Hyma KE, Asiedu R, Mitchell SE, Gedil M, Spillane C. Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. Theor Appl Genet 2014;127:1783–94.

[145] Padi FK, Ofori A, Takrama J, Djan E, Opoku SY, Dadzie AM, et al. The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genet Genomes 2015;11:44. doi:10.1007/s11295-015-0875-9.

[146] Wallace JG, Upadhyaya HD, Vetriventhan M, Buckler ES, Tom Hash C, Ramu P. The genetic makeup of a Global Barnyard Millet Germplasm Collection. Plant Genome 2015;8:1-7. doi:10.3835/plantgenome2014.10.0067.

[147] McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, et al. Agriculture: feeding the future. Nature 2013;499:23–24. doi:10.1038/499023a.

[148] Girma G, Korie S, Dumet D, Franco J. Improvement of accession distinctiveness as an added value to the global worth of the yam (*Dioscorea* spp) genebank. Int J Conserv Sci 2012;3:199–206.

[149] Galperin MY, Rigden DJ, Fernández-Suárez XM. The 2015 Nucleic Acids Research Database Issue and molecular biology database collection. Nucleic Acids Res 2015;43:D1–5. doi:10.1093/nar/gku1241.

[150] Shirasawa K, Isobe S, Tabata S, Hirakawa H. Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. Breed Sci 2014;64:264–71. doi: 10.1270/jsbbs.64.264.

[151] Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 2010;38:D843–46. doi:10.1093/nar/gkp798.

[152] Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EKS, et al. MaizeGDB: curation and outreach go hand-in-hand. Database (Oxford) 2011;2011:bar022. doi:10.1093/database/bar022.

[153] Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012;40:D1178–86. doi:10.1093/nar/gkr944.

[154] Kehoe MA, Coutts BA, Buirchell BJ, Jones RAC. Plant virology and next generation sequencing: experiences with a Potyvirus. PLoS One 2014;9:e104580. doi:10.1371/journal.pone.0104580.

[155] Li R, Gao S, Hernandez AG, Wechter WP, Fei Z, Ling K-S. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. PLoS One 2012;7:e37127. doi:10.1371/journal.pone.0037127.

[156] Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinkiene M, et al. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol Plant Pathol 2009;10:537–45. doi:10.1111/J.1364-3703.2009.00545.X.

[157] Seguin J, Rajeswaran R, Malpica-López N, Martin RR, Kasschau K, Dolja V V, et al. De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. PLoS One 2014;9:e88513. doi:10.1371/journal.pone.0088513.

[158] Dennis ES, Ellis J, Green A, Llewellyn D, Morell M, Tabe L, et al. Genetic contributions to agricultural sustainability. Philos Trans R Soc Lond B Biol Sci 2008;363:591–609. doi:10.1098/rstb.2007.2172.

[159] Fridman E, Zamir D. Next-generation education in crop genetics. Curr Opin Plant Biol 2012;15:218–23. doi:10.1016/j.pbi.2012.03.013.

INTECH