The Cassava Genome: Current Progress, Future Directions

**Simon Prochnik · Pradeep Reddy Marri ·
Brian Desany · Pablo D. Rabinowicz ·
Chinnappa Kodira · Mohammed Mohiuddin ·
Fausto Rodriguez · Claude Fauquet · Joseph Tohme ·
Timothy Harkins · Daniel S. Rokhsar · Steve Rounsley**

Abstract The starchy swollen roots of cassava provide an essential food source for nearly a billion people, as well as possibilities for bioenergy, yet improvements to nutritional content and resistance to threatening diseases are currently impeded. A 454-based whole genome shotgun sequence has been assembled, which covers 69% of the predicted genome size and 96% of protein-coding gene space, with genome finishing underway. The predicted 30,666 genes and 3,485 alternate splice forms are supported by 1.4 M expressed sequence tags (ESTs). Maps based on simple sequence repeat (SSR)-, and EST-derived single nucleotide polymorphisms (SNPs) already exist. Thanks to the genome sequence, a high-density linkage map is currently being developed from a cross between two diverse cassava cultivars: one susceptible to cassava brown streak disease; the other resistant. An efficient genotyping-by-sequencing

S. Prochnik · D. S. Rokhsar
US DOE Joint Genome Institute,
2800 Mitchell Dr.,
Walnut Creek, CA 94598, USA

S. Prochnik
e-mail: seprochnik@lbl.gov

P. R. Marri · S. Rounsley (✉)
BIO5 Institute, University of Arizona,
1657 E. Helen St.,
Tucson, AZ 85721, USA
e-mail: steve.rounsley@gmail.com

P. R. Marri
e-mail: pradeepreddy.marri@gmail.com

B. Desany · C. Kodira · M. Mohiuddin
454 Life Sciences (Roche),
1 Commercial Street,
Branford, CT 06405, USA

B. Desany
e-mail: Brian.Desany@roche.com

C. Kodira
e-mail: chinnappa.kodira@roche.com

M. Mohiuddin
e-mail: mohammed.mohiuddin@roche.com

P. D. Rabinowicz
Institute for Genome Sciences (IGS),
University of Maryland School of Medicine,
BioPark Building II, 801 West Baltimore Street,
Baltimore, MD 21201, USA
e-mail: prabinowicz@som.umaryland.edu

P. D. Rabinowicz
Dept. of Biochemistry & Molecular Biology,
University of Maryland School of Medicine,
Baltimore, MD 21201, USA

P. D. Rabinowicz
Office of Biological and Environmental Research, DOE,
Washington, DC, USA

F. Rodriguez
Johns Hopkins University,
Baltimore, MD, USA
e-mail: f.v.rodriguez@cgiar.org

C. Fauquet
ILTAB, Donald Danforth Plant Science Center,
975 North Warson Road,
St. Louis, MO 63132, USA
e-mail: Fauquet@danforthcenter.org

(GBS) approach is being developed to catalog SNPs both within the mapping population and among diverse African farmer-preferred varieties of cassava. These resources will accelerate marker-assisted breeding programs, allowing improvements in disease-resistance and nutrition, and will help us understand the genetic basis for disease resistance.

Abbreviations

BAC	Bacterial artificial chromosome
bp	Base pair
BMGF	Bill & Melinda Gates Foundation
CBSD	Cassava brown streak disease
CIAT	International Center for Tropical Agriculture
CMD	Cassava mosaic disease
DDPSC	Donald Danforth Plant Science Center
EST	Expressed sequence tag
GBS	Genotyping-by-sequencing
GCP-21	Global Cassava Partnership
ILTAB	International Laboratory for Tropical Agriculture Biology
JGI	US Department of Energy Joint Genome Institute
kbp	Kilobase-pair
Mb	Megabase-pair
NGS	Next-generation sequencing
QTL	Quantitative trait locus
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
SV	Structural variation
WGS	Whole genome shotgun

F. Rodriguez · J. Tohme
CIAT,
Km 17, Recta Cali-Palmira,
Apartado Aéreo 6713 Cali, Colombia

J. Tohme
e-mail: j.tohme2@cgiar.org

T. Harkins
Life Technologies Corp,
5791 Van Allen Way,
Carlsbad, CA 92008, USA
e-mail: tim.t.harkins@gmail.com

D. S. Rokhsar
Department of Molecular and Cell Biology,
University of California,
Berkeley, CA 94720-3200, USA
e-mail: dsrokhsar@gmail.com

S. Rounsley
Dow Agrosciences,
9330 Zionsville Road,
Indianapolis, IN 46268, USA

Introduction

Cassava (*Manihot esculenta* Krantz) is grown throughout tropical Africa, Asia and the Americas. Its large, starchy, roots and edible leaves provide food for 800 M people globally, many of whom subsist on it, in part because it is drought tolerant and requires little in the way of inputs (Ceballos et al. 2010). The high starch content (20-40%) makes cassava a desirable energy source both for human consumption and industrial biofuel applications (Balat and Balat 2009; FAO 2008; Schmitz and Kavallari 2009), However, its nutritional value is limited as the roots contain little protein or micronutrients and high levels of cyanogenic compounds. The plant is also susceptible to bacterial (Boher and Verdier 1994) and insect-transmitted viral diseases (Hillocks and Jennings 2003; Patil and Fauquet 2009). While the roots can be left in the ground for many months before harvesting, post-harvest deterioration is rapid, limiting economic development for cassava farmers (Reilly et al. 2007). Cassava is an outcrossing, heterozygous species propagated clonally from stem cuttings. These properties provide barriers to the already slow process of improving yield, disease resistance and nutrient content by traditional breeding and selection.

We describe here the progress-to-date and the future goals of genomic sequencing efforts in cassava. A draft genome sequence has been generated from a single cassava accession. With this genome sequence in hand and a catalog of common genetic variants on the way, many possibilities exist for the rapid realization of cassava's potential both as a nutritional food and as a biofuel feedstock.

Cassava belongs to the family Euphorbiaceae, and the Fabid superfamily (also known as eurosids I), which includes several distantly related plants such as rosids, legumes and poplars. The availability of several Fabid genomes, in particular *Ricinus communis* [castor bean (Chan et al. 2010), which is in the same family], *Populus trichocarpa* [poplar (Tuskan et al. 2006)] and *Glycine max* [soybean (Schmutz et al. 2010)], allows researchers to take advantage of comparative genomic approaches. In addition, tools for molecular breeding are being developed by building upon the genome sequence. For example, germplasm diversity across both cultivated and wild varieties, once characterized, can be used to map valuable traits. With this map, marker-assisted selection and even genomic selection can be adopted as paradigms for generating improved cassava. Brought together, these resources should allow us to accelerate understanding of basic biology of starch accumulation, nutrient loading and resistance to diseases and pests.

Sequencing the Cassava Genome

The project to sequence the cassava genome began in 2003 with the formation of The Global Cassava Partnership (GCP-21), co-chaired by Dr. Claude Fauquet, director of the International Laboratory for Tropical Agriculture Biology (ILTAB) at the Donald Danforth Plant Science Center (DDPSC), and Dr. Joe Tohme of the International Center for Tropical Agriculture (CIAT). This, in turn, led to a 2006 proposal, by Fauquet, Tohme and 12 other leading international scientists, to the US Department of Energy Joint Genome Institute (JGI) Community Sequencing Program. The proposal was selected for a pilot project, and over the next few years a ~0.8x random shotgun sample sequence (~700k reads) was produced, half from fosmid paired-ends. As a result, nearly half the genome had been sampled, but only relatively short ~700 base-pair (bp) sequences were available.

At the 2009 International Plant and Animal Genomes conference in San Diego, USA, Steve Rounsley from the University of Arizona obtained a commitment to produce a more complete cassava sequence from 454 Life Sciences and JGI, with the encouragement of the Bill & Melinda Gates Foundation (BMGF). 454 Life Sciences and JGI chose to use 454's Genome Sequencer FLX Titanium platform to rapidly generate the DNA sequence data needed for the project. 454 Life Sciences later added longer sequences from their then-experimental FLX+ extra long read technology (http://my454. com/products/gs-flx-system/index.asp). Generation of raw sequence data was mostly completed by spring 2009, less than 90 days after the meeting in San Diego, and progress was discussed at a meeting convened by BMGF outside Paris, France in June 2009. In November of that year, the genome assembly and annotation were publicly released (available from ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v5.0/Mes culenta/), and funding was obtained from BMGF for post-genome efforts to build upon the genome data to develop breeding tools for sub-Saharan Africa.

Status of Genome Project

The draft genome sequence of cassava was obtained by a whole genome shotgun (WGS) strategy (Fig 1a). In this approach, nuclear genomic DNA is isolated and broken randomly into short fragments, which are then sorted into bins of various sizes. The ends of these fragments are sequenced and the resulting short sequence reads computationally reassembled to reconstruct long stretches of genomic sequence. Assembled sequences are called "contigs" if they represent unbroken (i.e., contiguous) stretches of the genome, and "scaffolds" if they contain one or more gaps due to either unsampled sequence or repetitive regions that are difficult to reconstruct from short sequence reads. These gaps are represented as runs of the unknown nucleotide "N", with the number of Ns roughly corresponding to the length of the gap.

Assembling the cassava genome is computationally challenging, for two reasons. Firstly, repetitive sequences, often transposon-related, are interspersed between the compact genes of cassava, just as they are in most plant genomes (Gill et al. 2008). Since copies of a given repetitive element may be very similar at the sequence level, it is difficult to reconstruct each copy, especially with short sequence reads. Fortunately, the 454 FLX Titanium and experimental FLX+ sequences are typically 400- and 700-nt long, respectively, making them long enough to assemble many repetitive segments and genes together on scaffolds spanning hundreds of thousands of base pairs (see below).

Secondly, cassava, as an outbreeding species, has allelic variation, both SNPs and structural polymorphisms (deletions, insertions, and inversions involving hundreds or thousands of nucleotides), which can complicate the derivation of a single "reference" sequence for each locus. In an attempt to minimize the number of heterozygous alleles that the assembler would have to deal with, DNA from a partially inbred cassava cultivar was provided for the genome sequencing project. This line, named AM560-2, was developed at CIAT and is reported as being an S3-derived inbred from the cassava cultivar MCOL-1505 (J. Tohme, unpublished results).

The cassava genome spans an estimated 770 Mb (Awoleye et al. 1994) in N=18 chromosomes. A total of 22.4 billion bp of raw sequence data was generated (Table 1), enough to cover the genome ~29 times. This redundancy is necessary to assemble the reads into contigs and scaffolds (Fig 1a). These reads were assembled using Roche's GS *de novo* assembler (Newbler) v. 2.5 (Quinn et al. 2008) into 12,977 scaffolds that span a total of 532.5 Mb. This version 4 assembly is somewhat short of the estimated genome size, but, based on an analysis of the unassembled reads, the missing sequence is dominated by repetitive sequences such as transposable elements, ribosomal RNA genes and centromeres. Although the genome assembly is in nearly 13,000 pieces, half of it is captured in only 487 scaffolds, each longer than 258 kbp and containing 49 or more genes.

Previously, only eight cassava transposons were described in public databases (Kapitonov and Jurka 2008). With the assembled genome, a more complete characterization of repetitive content is possible by scanning the assembly for sequences that occur many times (Table 2). Over a third (37.5%) of the assembly was annotated as repetitive (Table 2), dramatically expanding the catalog of cassava transposable elements. This portion of the genome is hidden (in the jargon, "masked") when the genome is scanned for protein-coding genes.

While a substantial portion of the assembled scaffold sequence is represented in "gaps" (~113 Mb), almost all cassava genes are captured in the contigs. Of the putative transcripts

Fig. 1 **a** Overview of whole genome shotgun sequencing and assembly. Starting with plant material, many genomes' worth of DNA is extracted, purified, fragmented, pooled by length and sequenced to a high level of redundancy with the aim of sequencing every region of the genome so that the chromosomal sequence can be generated (assembled) by overlapping fragments that have (near-)identical sequences. Longer range, paired-end sequence information is used to bridge sections of the genome that are not unique (repeats) and impossible to resolve by this approach **b** The phytozome genome browser (http://www.phytozome.net/cassava) provides a portal for accessing, browsing, searching and downloading all available cassava sequence and annotation data and for comparative plant genomic analysis
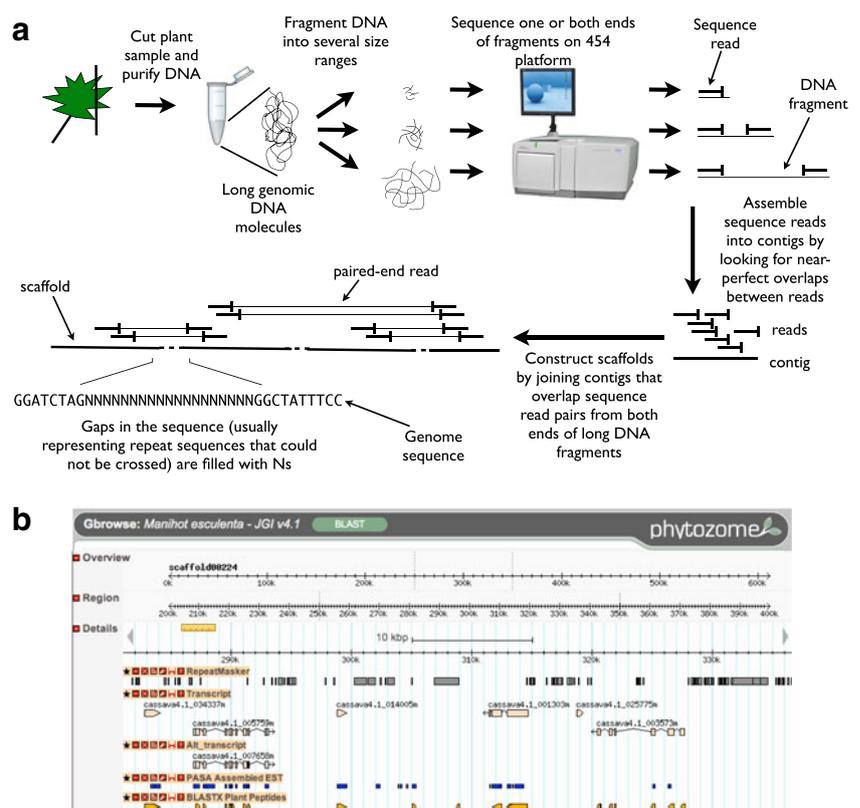
represented in an assembly of publicly available cassava EST sequences (http://cassava.igs.umaryland.edu/blast/db/EST_asmbl_and_single.fasta), 96% can be mapped to the genome assembly. It can be estimated that the remaining portion of the genome is largely repetitive and non-gene-coding. Consistent with this, the fractions of the estimated genome size (~31%) and WGS reads (~36%) that do not appear in the assembly are approximately equal, despite low read error rates. The scaffolds obtained have yet to be assigned to chromosomes, as this requires genetic markers with known sequence. However, an 88% complete genetic map comprising 23-linkage groups includes the genetic locations of 284 scaffolds (Sraphet et al. 2011).

With the gene-rich portion of the genome in hand, the next step is to identify the protein-coding genes and the exons that comprise them. This is achieved computationally by aligning sequences from mRNA fragments (ESTs) to the genome, as well as looking for regions with homology to known proteins from other plant species. The 80,459 Sanger ESTs from Genbank were augmented by a new set of 2.7 million reads from leaf and root libraries, generated by 454 Life Sciences using the FLX Titanium platform. While half

Table 1 Cassava genomic and mRNA sequence data

Sequence type	Source	Technology	Sequences generated	Notes
Genome shotgun	Roche	454 Titanium	39,259,112	
		454 FLX Plus (experimental)	10,785,244	
	JGI	454 Titanium	21,581,680	
		Sanger	723,958	
	U. Maryland & UC Davis	Sanger	75,748	BAC-end seq.
Expressed sequence tags (ESTs)	Roche	454 Titanium	1.51 M reads (leaf)	0.30 M after removing chloroplast and rDNA sequences
		454 Titanium	1.19 M reads (root)	
	various	Sanger	80,459	

Table 2 *M. esculenta* v. 4 assembly and repeat statistics

Total assembled scaffold length	532.5 Mb
Total assembled contig sequence length	419.5 Mb (21% gaps)
Total number of scaffolds	12,977
Half of the assembly is in scaffolds longer than	258.1 kbp (487 scaffolds)
Annotated repetitive portion of assembly	37.5%
Gypsy-related sequence	140 Mb (10M reads)
rDNA sequence	54 Mb (3.9 M reads, 6 k copies)
polygalacturonase gene sequence	12 Mb (850 k reads, 3 k copies)

of the leaf EST reads represented rDNA and chloroplast transcripts and were not useful for annotating the nuclear genome, the remaining 1.4 M reads were used to improve gene prediction. In addition to this mRNA data, predicted protein sequences from castor bean, *Arabidopsis*, rice, soybean and *Populus* were mapped to the cassava assembly to help predict gene loci. To date we have predicted 30,666 protein-coding loci in cassava by combining ESTs, peptide homology to other species, and the statistical sequence patterns common to plant genes (e.g., intronic splice signals; Table 3). Over a third (11,526) of the protein-coding loci are supported by ESTs, which also provide evidence for a total of 3,485 alternative splice forms (Table 3).

The gene content of cassava is broadly similar to castor bean, Arabidopsis, soybean and rice (Table 3). The cassava genome sequence and annotation can be readily aligned to those of related plants, allowing comparative genomic analysis of genome structure and function. For example, a global comparison of cassava and castor bean genomes reveals extensive colinearity (an example region is shown in Fig. 2a) and a paleo-genomic duplication in cassava (Fig. 2b) that took place relative to castor bean (Chan et al. 2010). However, not all genes are found in two copies in cassava, suggesting that some initially duplicated genes have been subsequently lost. Conversely some paralogous families have more than two members, suggesting additional gene duplications that resulted in increased family size.

The cassava genome and annotation can be viewed, browsed, compared with other genomes, and downloaded for custom analysis at the Phytozome comparative plant genomics portal (www.phytozome.net, Fig. 1b; Goodstein et al. 2012). Extensive instructions are available to help users find their way around the site (http://phytozome.net/help.php).

Genetic Variability and Diversity in Cassava

Although cassava originated in South America and was exported to Africa and Asia, its population structure is poorly understood relative to better studied crops such as maize and rice. An understanding of genetic variation allows the development of robust systems of markers for mapping and breeding, including the characterization of germplasms that might provide useful alleles (Edwards and Batley 2010). Initial marker development in cassava has relied upon simple sequence repeats (SSRs), such as microsatellite sequences (Raji et al. 2009; Roa et al. 2000), as well as ~2,000 SNPs identified in expressed sequence tags (Ferguson et al. 2011; Tangphatsornruang et al. 2008). Known SSR and SNP markers, however, are sparsely distributed across the cassava genome and may not be ideal for either fine-mapping or inexpensive large-scale assays. In the future, this can be addressed using newer methods for marker discovery which rely on increasingly inexpensive next-

Table 3 Protein-coding annotation comparisons to cassava

	cassava v. 4.1	castor bean (TIGR v. 0.1)	*Arabidopsis thaliana* Columbia (TAIR10)	soybean (JGI v. 1)	rice (MSU v. 6.0)
Protein-coding gene loci	30,666	31,221	27,416	46,367	40,838
Alternate transcripts	3,485	0	7,970	9,420	10,420
Genes supported by one or more ESTs over >75% of length	11,526	3,262	16,931	9,431	18,364
Avg. number of exons/transcript	5.1	4.2	5.3	6.0	4.3
Median exon length (bp)	148	142	155	142	163
Median intron length (bp)	166	169	99	185	176
Transcripts with Pfam domain annotation (KOG orthology)	20,641 (12,307)	16,720 (9,321)	19,419 (12,184)	34,065 (20,601)	20,766 (9,933)

The protein-coding gene annotations in cassava are compared to castor bean, *Arabidopsis*, soybean and rice

a   **b**
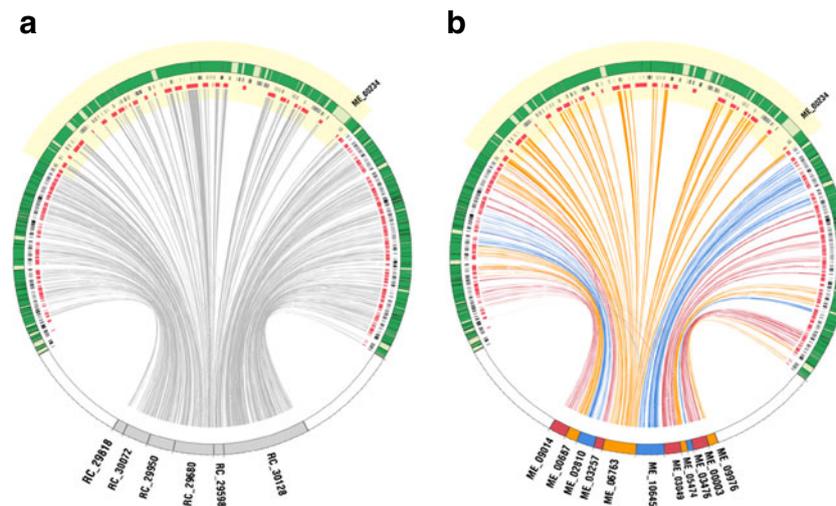
Fig. 2 a Circos plot showing cassava-castor bean colinearity. A sample 2.7 Mb region of the cassava genome was aligned against castor bean scaffolds with the Promer tool from Mummer v3.2, and visualized with Circos. Colored segments of the circle represent cassava while greyscale represents corresponding segments of castor bean genome. Green blocks in outer ring are contigs within the 2.7 Mb scaffold. The second ring represents repeat content in grey blocks, and the inner ring represents genes in red blocks. Grey lines linking cassava segments to castor bean segments represent homology at the protein level, and the fact that these lines do not cross over each other indicates colinearity. The central portion of the cassava scaffold, highlighted in yellow, is zoomed in to show more detail. **b** The same 2.7 Mb cassava scaffold as in (a; top) but aligned against other regions of the cassava genome that are highly similar (bottom). This demonstrates the presence of a duplication in the cassava genome

generation sequencing (NGS) and can provide greater power for high density SNP discovery.

Future Directions

Several resources are currently available for improving the genome sequence assembly and are being integrated with the draft genome sequence. The first is a bacterial artificial chromosome (BAC)-based physical map from Pablo Rabinowicz at the Institute for Genome Sciences, University of Maryland and Mincheng Luo at the University of California, Davis (funded by the Generation Challenge Programme). Additional sequencing and computational gap closure are planned, as well as the generation of a dense genetic map to tie scaffolds together at the chromosome scale. This will provide a more accurate substrate for gene models, and a platform for robust systems of markers for studying natural variation and for guiding breeding programs.

Characterizing genetic variation in cassava is a significant part of the cassava genome effort. Whole genome sequencing projects of several key cultivars are already underway, including parents of populations developed for genetic mapping of tolerance to cassava brown streak disease (CBSD) – a disease that is having a devastating effect in East Africa. A large mapping population derived from two parents ("Albert", a disease susceptible variety and "Namikonga", a disease resistant variety; Ferguson et al – this issue) is being developed, and with inexpensive genotyping-by-sequencing (GBS), a robust,

high resolution genetic map for cassava will be available in the next few years. Two additional sequencing efforts are already underway. These are generating half a dozen whole genome sequences as well as GBS information for hundreds of cultivars (including some wild relatives) chosen to sample traits as widely as possible and maximize diversity. The goal is to obtain a catalog of SNPs segregating within farmer varieties useful for African environments. These can then be used to accelerate breeding programs that will address disease and nutrition concerns, ultimately improving cassava's quality as a crop.

In addition to describing SNPs, these whole genome sequencing ventures will allow mapping of genomic structural variation (SV), including gene presence/absence, local duplications and transposon activity. These genetic changes are increasingly recognized as important components of heritable variation (Lam et al. 2010; Springer et al. 2009). Rare SNPs/SVs will still be elusive, but these programs should be able to identify nearby linked markers, analogous to the human HapMap (http://www.hapmap.org).

High-throughput genotyping (the determination of the alleles present at a defined set of marker loci) is now possible, with test assays underway in cassava. With the declining cost of sequencing, there is much enthusiasm for genotyping based on reduced representation sequencing (or GBS), in which loci defined by their proximity to restriction sites are assayed en masse (Elshire et al. 2011). Combined with phenotypic information, the lower cost genotyping will allow deeper sampling of diverse germplasm, the simplification of QTL (quantitative trait locus) mapping for traits such as drought or disease

resistance segregating in crosses between arbitrary parents, and, possibly, whole genome association mapping.

Germplasm from hundreds of African cassava cultivars will be characterized in this approach, allowing marker-assisted breeding schemes to be developed for improving nutrient content as well as tolerance of both drought and viral cassava mosaic disease (CMD) and CBSD. We anticipate developing tens of thousands of markers across the genome. In addition to mapping and breeding, genome-enabled genetic approaches can also open the door to a mechanistic understanding of the underlying biology. As data is generated and analyzed, a genome variation database will be developed, thus providing breeding tools to aid farmers in improving cassava, with a special focus on increased resistance to CBSD.

References

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

Balat M, Balat H (2009) Recent trends in global production and utilization of bio-ethanol fuel. Appl Energ 86:2273–2282

Boher B, Verdier V (1994) Cassava bacterial blight in Africa: the state of knowledge and implications for designing control strategies. Afr Crop Sci J 2:505–509

Ceballos H, Okogbenin E, Pérez JC et al (2010) Cassava. In: Bradshaw JE (ed) Root and tuber crops, handbook of plant breeding, vol 7. Springer, New York, pp 53–96

Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species Ricinus communis. Nat Biotechnol 28:951–956

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8:2–9

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

FAO (2008) Cassava for food and energy security. FAO Newsroom. http://www.fao.org/newsroom/en/news/2008/1000899/index.html. Cited 19 Nov 2011

Ferguson ME, Hearne SJ, Close TJ et al (2011) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor Appl Genet

Gill N, Hans CS, Jackson S (2008) An overview of plant chromosome structure. Cytogenet Genome Res 120:194–201

Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res (In Press)

Hillocks RJ, Jennings DL (2003) Cassava brown streak disease: a review of present knowledge and research needs. Int J Pest Manag 49:225–234

Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet 9:411–412, author reply 414

Lam HM, Xu X, Liu X et al (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059

Patil BL, Fauquet CM (2009) Cassava mosaic geminiviruses: actual knowledge and perspectives. Mol Plant Pathol 10:685–701

Quinn NL, Levenkova N, Chow W et al (2008) Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. BMC Genomics 9:404

Raji AA, Anderson JV, Kolade OA et al (2009) Gene-based microsatellites for cassava (Manihot esculenta Crantz): prevalence, polymorphisms, and cross-taxa utility. BMC Plant Biol 9:118

Reilly K, Bernal D, Cortes DF et al (2007) Towards identifying the full set of genes expressed during cassava post-harvest physiological deterioration. Plant Mol Biol 64:187–203

Roa AC, Chavarriaga-Aguirre P, Duque MC et al (2000) Cross-species amplification of cassava (Manihot esculenta) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. Am J Bot 87:1647–1655

Schmitz PM, Kavallari A (2009) Crop plants versus energy plants-on the international food crisis. Bioorg Med Chem 17:4020–4021

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Springer NM, Ying K, Fu Y et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734

Sraphet S, Boonchanawiwat A, Thanyasiriwat T et al (2011) SSR and EST-SSR-based genetic linkage map of cassava (Manihot esculenta Crantz). Theor Appl Genet 122:1161–1170

Tangphatsornruang S, Sraphet S, Singh R et al (2008) Development of polymorphic markers from expressed sequence tags of Manihot esculenta Crantz. Mol Ecol Resour 8:682–685

Tuskan GA, Difazio S, Jansson S et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596–1604