

# Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future

Morag Ferguson · Ismail Rabbi · Dong-Jin Kim ·  
Melaku Gedil · Luis Augusto Becerra Lopez-Lavalle ·  
Emmanuel Okogbenin

Received: 4 October 2011 / Accepted: 21 November 2011 / Published online: 23 December 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** The advent of relatively low-cost, massively parallel, high-throughput genome sequencing and the resultant availability of high density markers are revolutionizing the ways in which molecular markers can be applied to plant breeding. With the availability of the draft cassava genome sequence, the cassava community is poised to take advantage of these new tools. Here we review the development of molecular markers applied to cassava breeding and describe the achievements that have been made using predominantly simple sequence repeat (SSR) markers. At this time of change, we report on the curation of 3,367 published and unpublished SSR primer pairs and provide a non-redundant database. We also describe ways in which new tools, particularly single nucleotide polymorphism (SNP) markers, can be applied to the development of high density maps and to fine mapping, association mapping, gene discovery, transcript profiling, inbred line development and the

prediction of heterosis, gene mining in wild species and introgressions, and genome-wide approaches, including marker-assisted recurrent selection (MARS) and genomic selection (GS). Where applicable we describe how these tools are already being applied for amassing genetic gain in cassava.

**Keywords** Single nucleotide polymorphism (SNP) markers · Marker-assisted breeding · Simple sequence repeat (SSR) markers · *Manihot esculenta*

## Abbreviations

ABC-QTL	Advanced backcross QTL mapping scheme
AFLPs	Amplified fragment length polymorphism
BAC	Bacterial artificial chromosome
CBSD	Cassava brown streak disease
cDNA	Complimentary DNA
CGM	Cassava green mite
CIAT	Centro Internacional de Agricultura Tropical
CMD	Cassava mosaic disease
DArT	Diversity Array Technology
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ESTs	Expressed sequence tags
GCP	Generation Challenge Program
GEBV	Genomic estimate of breeding value
GS	Genomic selection
IITA	International Institute of Tropical Agriculture
LD	Linkage disequilibrium
MARS	Marker-assisted recurrent selection
MAS	Marker-assisted selection
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
PPD	Post-harvest physiological determination
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA

Communicated by: Nigel Taylor

M. Ferguson (✉) · D.-J. Kim  
International Institute of Tropical Agriculture (IITA),  
c/o International Livestock Research Institute,  
P.O. Box 30709, Nairobi 00100, Kenya  
e-mail: m.ferguson@cgiar.org

I. Rabbi · M. Gedil  
IITA,  
Oyo Road,  
Ibadan, Nigeria

L. A. B. Lopez-Lavalle  
International Center for Tropical Agriculture (CIAT),  
Km 17, Recta Cali-Palmira,  
AA6713 Cali, Colombia

E. Okogbenin  
National Root Crops Research Institute,  
PMB 7006, Umuahia 440001, Abia State, Nigeria

RE	Restriction enzyme
RFLP	Restriction fragment length polymorphism
RGCs	Resistance gene candidates
SEC	Southern, eastern and central
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat

## Introduction

Cassava breeding efforts throughout the world have made significant impact on cassava production, particularly in terms of disease tolerance, yield and quality improvements (Legg et al. 2006; Nweke et al. 2002), however it is anticipated that additional gains in breeding efficiency, which would translate into genetic gain, could be made through the application of advanced molecular breeding technologies. This has been demonstrated in highly researched crops such as maize (Eathington 2005) and rice (Ragot and Lee 2007). Cassava is a difficult crop to breed, due to its intrinsic heterozygosity, variable flowering time, low seed set, and long breeding cycle combined with the agricultural goals of diverse use of the products and growth in harsh environment conditions both in terms of biotic and abiotic stresses (Jennings and Iglesias 2001).

While the use of molecular markers in cassava breeding is in its infancy (Okogbenin et al. 2007), they have been applied in plant breeding of other crops in several ways. Marker-assisted selection (MAS) can be applied if markers are either directly associated with a trait (a functional gene with a tractable phenotype) or closely associated with known genes of interest. This is the classical form of MAS, in which either single genes or quantitative trait loci (QTL) may be selected. With the advent of next-generation sequencing (NGS) technologies (454 pyrosequencing, Solexa, and SOLiD), the associated proliferation of single nucleotide polymorphism (SNP) markers, and the possibility of genotyping-by-sequencing (GBS), approaches based on genome-wide marker scans, such as marker-assisted recurrent selection (MARS) or genomic selection, may be applied for more rapid genetic gain in highly qualitative traits such as yield or drought tolerance (Heffner et al. 2011).

Molecular breeding in cassava would confer several advantages, including: (1) more accurate gene-based selection, which would allow, for example, the pyramiding of resistance genes; (2) enhanced genetic gain in quantitative traits through predictive modeling (MARS and genomic selection); (3) reduced breeding population size, which would allow breeders to work on a larger number of genotypes simultaneously; (3) reduced amount of time to product delivery; and (4) preemptive breeding in environments where particular stresses, such as cassava mosaic disease

(CMD) or cassava brown streak disease (CBSD), are not currently present, but pose a significant threat.

The successful application of molecular markers does however rely on the availability of genomic technologies. In recent years the availability of genomic resources for cassava has increased substantially, most notable through the sequencing of the cassava genome (<http://www.phytozome.net/cassava>; Prochnik et al. 2011). Here we review the current availability and ongoing applications of molecular markers to cassava research and breeding, highlight new genomic resources, and discuss and speculate on the implications of NGS on molecular breeding strategies for cassava.

## Availability of Molecular Markers

Molecular markers, such as random amplified polymorphisms (RAPDs) and restriction length polymorphisms (RFLPs), were first used in cassava to study the genetic diversity within the genus *Manihot* (Marmey et al. 1993). Later, amplified fragment length polymorphism (AFLPs) were used to understand genetic differentiation in cassava (Elias et al. 2000; Fregene et al. 2000; Roa et al. 1997). The first genetic linkage map of cassava utilized AFLP, RAPDs and isozymes (Fregene et al. 1997). For the past decade, however, these marker types have largely been replaced by simple sequence repeat (SSR) markers.

### Simple Sequence Repeat (SSR) Markers

Microsatellites (Litt and Luty 1989) or SSRs (Tautz et al. 1986) are two-, three- or four-nucleotide tandem repeat units. They reflect genomic points of variation within a species, which is more highly variable when the repeat number is ten or greater (Queller et al. 1993). SSRs are largely co-dominant, multi-allelic, and dispersed throughout the genome and can be multiplexed on semi-automated systems (Varshney et al. 2005). In cassava, several groups have developed a few thousand SSR markers from expressed sequence tags (ESTs) and enriched genomic DNA libraries (Chavarriaga-Aguirre et al. 1998; Mba et al. 2001; Raji et al. 2009; Sraphet et al. 2011; Tangphatsomruang et al. 2008). As the SSR resources were developed independently by different research groups, several SSR primer pairs of different names target the same SSR. To identify duplicates, cassava SSR information was curated using the recently developed cassava genome assembly as a reference.

*The Identification of a Non-redundant SSR Dataset* The mapping of polymerase chain reaction (PCR) products onto the genome was used to identify redundancy in the SSR collection. A total of 6,752 cassava SSR primer

sequences (forward and reverse primers for 3,367 SSRs) were compiled from various sources (Table 1) and queried using *BlastN* against the cassava genome sequence (*manihot\_esculenta\_147*, 12,977 sequences, 532.5 Mb, 3/31/11; [ftp://ftp.jgi-psf.org/pub/JGI\\_data/phytozome/v7.0/Mesculenta/assembly/](ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Mesculenta/assembly/)). From these, 5,402 SSR primers aligned to genomic scaffolds. There were many cases in which the forward and reverse primers of the same SSR were located on either different scaffolds or multiple scaffolds. In 2,079 cases, primers were designated as pairs based on SSR ID and distance between the primers. The average size of the PCR fragments from the paired SSRs was 302 bp. A total of 1,917 SSRs (92.2%) produced a PCR product less than 500 bp, an optimal size for routine PCR testing. Eighty-eight SSRs (4.7%) appeared to produce a PCR product larger than 1 kb, which is indicative of an intron.

The PCR products of 716 SSRs overlapped with those of other SSRs within a 1 kb range and were thus designated as duplicates. The remaining 1,363 SSRs appeared to be unique SSRs. The 716 duplicates were consolidated into a set of 312 primer pairs. These representatives together with the unique SSRs generated a set of 1,675 SSRs validated by genome comparison. To identify those SSRs that were not present in the cassava genome sequence, all 6,752 SSR primer sequences were blasted against the NCBI *M. esculenta* nucleotide ESTs (80,631 sequences; 40.8 Mb). This query identified 2,169 SSRs. Of these, 1,526 had already been identified from the curation process using the genome sequence, meaning a total of 643 SSR were additionally identified by EST blast analysis. Of these newly validated SSRs, 341 were unique and 302 were duplicates. The duplicates were curated into 130 non-redundant SSRs.

In total, we generated a non-redundant set of 2,146 SSRs comprised of 1,675 curated from the genome and 471 curated from ESTs (Table 2). The curated, paired SSR sets and associated information can be found at [http://bioinformatics.iita.org/cassava\\_SSRs](http://bioinformatics.iita.org/cassava_SSRs). Although this set of SSR markers is useful, it is still limited compared to some other crop plants.

**Table 1** Simple Sequence Repeat (SSR) markers in cassava

Number of SSRs	Authors	Publication status
186	Mba et al. (2001)	Published
124	Raji et al. (2009)	Published
55	Kunkeaw et al. (2010)	Published
2190	Sraphet et al. (2011)	Published
631	M. Fregene, A. Zarate, L. Cano, Mba, V. Verdier/ C. López (CIAT)	Personal communication
190	Ferguson (IITA)	Personal communication
3376		Total

**Table 2** Simple Sequence Repeat (SSR) curation process in cassava

Curation process	No of SSRs	Note
Total number of SSR for curation	3376	
Total primer sequences	6752	5402 with Blast hits 1350 with no blast hit
Paired SSRs	2079	1917 SSR (smaller than 500 bp PCR product)
Non-redundant SSRs linked to genome	1675	1363 Singleton 312 Consolidated (from 716 duplicate SSRs)
Non-redundant SSR, with EST information only	471	341 Singleton 130 Consolidated (from 302 duplicate SSRs)
Total non-redundant SSRs	2146	

For example the Gramene database contains 19,480 SSRs (<http://www.gramene.org/markers/microsat/>), mainly from *Oryza sativa* Japonica group. The current paucity of SSRs lowers the genome resolution and limits their use for breeding and genome-wide studies.

*Curated SSR Integrated with the Genome Assembly* The draft genome and latest SSR-based genetic map (Sraphet et al. 2011) provide an opportunity to integrate the curated SSRs into the genome and genetic linkage maps. Using *BlastN*, 1,675 SSRs validated by genome sequence were located on 686 genome scaffolds, which totaled 256 Mb. The current version of the cassava genome assembly consists of 12,977 scaffolds spanning 533 Mb. This implies that about 50% of the draft cassava genome, in terms of length of SSR-containing scaffold, is linked to the curated SSR sets. Scaffold information of 1,430 putative SNPs of the cassava genome were obtained from recent SNP research (Ferguson et al. 2011). These and the curated SSRs were analyzed in relation to existing SSR-based genetic linkage maps, number of scaffolds and genome coverage. A total of 1,298 SSRs were found on the genetic linkage map, representing 253 scaffolds and 137 Mb. Results are presented in Table 3. It is anticipated that individual tracks of SSRs and SNPs will be added to the genome sequence in the online Phytozome database.

#### DArT Genotyping of Cassava and Its Wild Relatives

Prior to the development of SNP markers, marker density and the high cost per data point limited the application of molecular markers to cassava genetic resource conservation and breeding. Cassava geneticists urgently needed a set of molecular markers that more completely covered the genome, that were based on a technology platform that was

**Table 3** Integrated Cassava Simple Sequence Repeat (SSR) information

	# of scaffolds	Covered sequence regions	Note	Reference
Cassava genome assembly	12,977	533 Mbs		
Scaffold with the curated SSR	686	256 Mbs	1675 curated SSR	
Scaffold with validated SNPs			1430 SNPs	Ferguson et al. 2011
With both SNPs and curated SSR	347	194 Mbs		
Mapped scaffold	284	151 Mbs	510 SSR mapped	Sraphet et al. 2011
Mapped scaffolds with curated SSR	253	137 Mbs		
Scaffold with the validated SNP, curated SSR, and map information	177	122 Mbs	Mapped in 22 LGs	

easily accessible, and that could be scored readily in new germplasm by any member of the global cassava research community. In early 2000, a novel molecular marker technique based on micro-array DNA hybridization was developed. This technique can genotype hundreds of polymorphisms across a large number of individual plants. With the proper bioinformatic analytical tools, the Diversity Array Technology (DArT) can be used to characterize several hundreds to thousands of polymorphisms in a timely, cost-effective manner. The first developed cassava DArT array had nearly 1000 polymorphic clones with a 99.8% reproducibility (Xia et al. 2005), offering a high-throughput marker screening system at a low cost.

DArT locus informativeness was tested against the well-studied cassava SSR marker system in 436 cassava accessions at Centro Internacional de Agricultura Tropical (CIAT). It was concluded that, even though SSRs sampled significantly less loci per reaction than the DArT technology, SSRs provided greater differentiation and more effectively recovered the patterns of genetic diversification in the genus. These results indicated that DArT markers have a limited application for cassava germplasm characterization (Becerra Lopez-Lavalle per. comm.).

#### Single Nucleotide Polymorphism Markers

SNP markers and small insertions and deletions (indels) represent the most frequent form of naturally occurring genetic variation within populations (Cho et al. 1999). The identification of a high density of SNPs in cassava would dramatically facilitate progress in cassava genomics and breeding. SNPs are generally biallelic (two alleles at a locus) and in this sense are individually less informative than SSRs, which are generally multi-allelic (many alleles at a locus; Syvänen 2001). This drawback is compensated for by the abundance and suitability of SNPs to ultra-high-throughput genotyping techniques (Appleby et al. 2009; Rafalski 2002). The utilization of multi-SNP haplotypes can offset the relatively low information content of single SNP loci (Brumfield et al. 2003). An early study of sequence variation in cassava identified 136 SNPs from EST

sequences and 50 SNPs from bacterial artificial chromosome (BAC) end sequences (Lopez et al. 2005). Kawuki et al. (2009) studied sequence diversity in nine genes and identified 26 SNPs. Sakurai et al. (2007) reported the identification but no details of 2,356 SNPs (Sakurai et al. 2007). The frequency of SNPs was found to be one every 53 bp in non-coding regions and every 181 bp in coding regions, with an average of one SNP every 121 bp (Kawuki et al. 2009). Lopez et al. (2005) found on average a frequency of one SNP every 62 bp. This high frequency of SNPs is consistent with the finding from other species, such as grapevine (Salmaso et al. 2005) and maize (Ching et al. 2002).

The recent identification and validation of 1,190 SNP markers in cassava has been reported from a total of 2,954 putative EST-derived SNPs (Ferguson et al. 2011). These SNPs have been located on scaffolds of the cassava genome sequence (v.4.1). The University of Maryland's Cassava Genome Database provides 384 putative SNPs derived from genes, and 371 putative SNPs derived from the cassava physical map. A putative SNP is a nucleotide variant that has been identified from sequence data, but has not been validated and may be a result of sequencing error. An explosion in the identification of SNPs in cassava is anticipated in the near future with the plummeting cost of sequencing, which has made whole genome re-sequencing projects feasible.

A large number of known SNPs will be available for genotyping using custom-made SNP arrays. Ferguson et al. (2011) used Illumina GoldenGate (Illumina Inc., San Diego, CA) technology to assay 1,536 putative SNPs'. However, this approach is relatively inflexible in terms of the number of markers and genotypes assayed and, for this reason, may be more suited to genomic or association mapping approaches. Recently, the Generation Challenge Program (GCP) converted 1,740 SNPs in cassava for use on the KASPar platform (LGC). This system is extremely flexible in terms of the combination of numbers of markers and samples that can be genotyped and, therefore, is particularly suitable for molecular breeding applications, such as MAS



or MARS. Converted KASPar markers are available through the GCP IBM marker services (<http://marlow.iplantcollaborative.org/marker-service>). We anticipate that in the near-to-medium term most common breeding applications will rely on low cost SNPs in relatively low density formats such as KASPar. The de novo discovery of SNPs through reduced representation GBS is currently under development for cassava in a number of laboratories. Reduced representation GBS is based on reducing genome complexity with methylation-sensitive restriction enzymes (REs) and barcoded oligonucleotide adaptors, followed by high-throughput sequencing. The GBS procedure mapped roughly 200,000 and 25,000 sequence tags in maize (IBM) and barley (Oregon Wolfe Barley), respectively (Elshire et al. 2011).

#### Other Genomic Resources

ESTs are partial sequences (200–800 bp) of expressed genes randomly picked from a complimentary DNA (cDNA) library. EST data combined with full-length cDNAs form an important resource for allele mining and marker development. To date, 80,631 cassava ESTs have been deposited in GenBank (Ferguson et al. 2011; Lokko et al. 2007; Lopez et al. 2004; Sakurai et al. 2007). A sub-set of nearly 60,000 of these, filtered on quality, have been compiled into the HarvEST: Cassava database (<http://harvest.ucr.edu>). RIKEN Institute, Japan, in collaboration with CIAT have generated two further EST libraries and report Sanger, 454 and Illumina paired-end sequences (Utsumi et al. 2011). To date no RNAseq next generation sequencing data is available, although a number of projects using this technology are reported to be in progress.

A number of microarrays have been developed for cassava, although their application appears to be limited. A Euphorbiaceae microarray was developed by James Anderson and consisted of a unigene set of 19,015 cDNAs from leafy spurge (Anderson et al. 2007). A cassava unigene microarray was produced by Valerie Verdier (Lopez et al. 2004). A 60-mer oligonucleotide microarray representing 20,840 cassava genes was used to study storage root formation and stress response (Yang et al. 2011). A similar array was developed with approximately 11,000 probes for transcriptome analysis (Ingelbrecht et al. 2008).

#### The Application of Molecular Markers to Diversity Assessments

Several studies have characterized the genetic diversity in cassava gene pools with the aim of aiding genetic resource conservation and breeding programs. One of the first attempts using molecular markers on a global scale looked

at genetic diversity, at differentiation and for potential heterotic groups in cassava (Fregene et al. 2003). The genetic diversity was assessed in 283 cassava landraces from Tanzania (163), Nigeria (29) and the Neotropics (Brazil, Colombia, Peru, Venezuela, Guatemala, Mexico and Argentina; 91) using 67 marker loci. The high levels of genetic diversity they found in all countries was unexpected, considering the probable center of domestication along the southern rim of the Amazon Basin and the later expansion into other regions of the Neotropics, Africa and Asia (Olsen and Schaal 2001) that would have been expected to produce a founder effect of reduced diversity and an increase in genetic differentiation. The authors attributed the observed high levels of diversity to spontaneous recombination in farmer's fields. Levels of molecular marker diversity in cassava landraces from Africa and several Neotropical countries were also found to be similar to those reported in Brazilian landraces (Beeching et al. 1993; Fregene et al. 2000).

Fregene et al. (2003) observed both a separation between Neotropical and African landraces and a more pronounced substructure in the African accessions as compared to the Neotropical landraces. Landraces from Guatemala were particularly highly differentiated from other regions. This general structure agrees with a previous AFLP marker study of 29 African and 11 Neotropical landraces (Fregene et al. 2000).

A larger diversity study of approximately 2,300 accessions from the International Institute of Tropical Agriculture (IITA), Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) and CIAT gene banks using 30 SSR markers was conducted (Hurtado et al. 2008). While this is not a global representation of cassava diversity, as only 2.4% of the varieties were from southern, eastern or central Africa (<http://gcpcr.grinfo.net/>), the study did observe a separation of accessions from Africa and Latin America/Asia. It also revealed a separation of some Nigerian accessions from the rest of Africa, and some Guatemalan accessions from other Latin American samples.

Until 2010, only relatively small-scale genetic diversity assessments of cassava in southern, eastern and central (SEC) Africa had been undertaken using a range of molecular markers (Benesi 2005; Fregene et al. 2000; Kizito et al. 2005; Zacarias et al. 2004). Using 26 SSR markers, Kawuki (2009) examined the nature and extent of genetic variation within a group of 1,401 cassava varieties from seven countries in SEC Africa: Tanzania (270 genotypes), Uganda (268), Kenya (234), Rwanda (184), Democratic Republic of Congo (DRC; 177), Madagascar (186) and Mozambique (82) (Kawuki 2009). This study revealed somewhat uniformly high levels of diversity across the region. It also revealed a subtle diversity sub-structure, with farmer-varieties from southern and eastern coastal areas (Madagascar, Mozambique and Tanzania) clustering together and germplasm from

the Democratic Republic of Congo, Uganda and Rwanda clustering together. This could reflect two routes of introduction of cassava into Africa, one through West Africa during the 1700s (Carter et al. 1992; Jones 1959) and the other through the east African coastline in the 1750s, where it was then introduced to Madagascar and to mainland Africa (Jones 1959; Langlands 1966).

Recently, 1,190 SNP markers were used to genotype 53 cassava varieties from the Americas, Asia, West Africa and SEC Africa (Ferguson et al. 2011). This study shows a similar structure in the germplasm as that observed by Fregene et al. (2003) using SSR markers. Although germplasm tended to cluster together on a regional basis, the Americas, West Africa and SEC Africa, other groups contained germplasm of mixed origin. A slightly larger genetic diversity was found in the Americas (0.3488) compared to that in Africa (0.3357), both with a sample size of 22. These relatively uniform levels of diversity were consistent with those observed by Fregene et al. (2003). Mean observed heterozygosity from SNP markers was lower than that observed by Fregene et al. (2003). This was attributed to the biallelic nature of SNPs as opposed to the multi-allelic nature of SSRs.

Molecular markers have been applied to a number of other purposes related to diversity in cassava. Chavarriaga-Aguirre et al. (1999) used SSR markers to identify duplicates in the CIAT core collection (Chavarriaga-Aguirre et al. 1999). Similarly, Ferguson et al. (2011) demonstrate the use of SNPs to detect duplicates in the IITA Genebank collection. SNP and SSR markers were used to make inferences about cassava's putative wild progenitor and to analyze variation in its natural populations (Olsen and Schaal 2001; Olsen 2004).

Another obvious, but so far unexploited, use of genetic diversity data is delineation of heterotic grouping in cassava to maximize heterosis for hybrid breeding. Heterosis, defined as increased performance of hybrid progeny compared to their parents, is known to be a function of genetic divergence among cultivars and is more highly expressed in outbreeding species (heterozygous) than in inbreeding ones (Becker 1993). There is an acute need to utilize the existing molecular diversity and genetic differentiation data for classifying and delineating heterotic groups in cassava as was done at the turn of the last century in maize (Crow 1998). Assigning germplasm to genetically divergent heterotic groups is fundamental for optimum exploitation of heterosis through crossing of complementary lines or populations. Traditionally, pedigree analysis, agro-morphological differentiation, measurement of heterosis and combining ability analysis in diallel crosses have been employed to establish heterotic patterns and grouping. This can be quite demanding, considering the large number of pair-wise crosses required. Molecular markers such as SSRs and SNPs can be

used to determine genetic relatedness and thus assist in the selection of parents required for experimental studies to identify heterotic groups. Theoretical and empirical research in maize and other crops have identified linear association between marker-based genetic distance and heterosis (Reif et al. 2003). Being a clonally propagated species, heterosis can be effectively exploited as breeders need to create a superior genotype once yet maintain it indefinitely.

### The Application of Markers to the Discovery of Marker-Trait Associations

*Genetic Linkage Maps* Genetic linkage maps form the basis of many approaches to the discovery of both major genes and QTLs (Lander and Botstein 1989) that in turn can be applied to conventional MAS. A genetic map also offers a framework for carrying out evolutionary and comparative genomic studies (Ahn and Tanksley 1993) and contributes to understanding the organization and dynamics of genomes, such as landscapes of linkage disequilibrium (LD) (Flint-Garcia et al. 2003; Sewell et al. 1999).

Due to cassava's highly heterozygous nature, inbreeding depression during selfing (Rojas et al. 2011), long growing cycle and low seed number per cross, genetic linkage mapping has traditionally been carried out in outbreeder full-sib (F1) families that are genetically fixed through vegetative propagation. A similar approach has been used in other perennial and clonally propagated crops such as tea (Hackett et al. 2000), rhodesgrass (Ubi et al. 2004) and banana (Hippolyte et al. 2010). Linkage maps are calculated using a double pseudo-testcross mapping strategy to create separate parental linkage maps in order to account for two independent parental meioses before simultaneous analysis (Grattapaglia and Sederoff 1994). Table 4 summarizes the salient characteristics (population type, sample size, number and types of mapped markers, number of linkage groups, map coverage and average distance between adjacent markers) of available cassava genetic maps. The first cassava genetic map was generated from a full-sib cross and consisted of 168 markers, mainly RFLPs and RAPDs, as well as a few SSRs and isoenzymes (Fregene et al. 1997). This linkage map was based on an F1 segregating population (known as the K family) of two geographically divergent parents (♀TMS30572 × ♂CM2177-2). The female parent TMS30572, from Africa and with resistance to CMD, was derived by introgressing *M. glaziovii* into *M. esculenta* (Hahn et al. 1980), while the male parent was an elite CIAT cultivar with no resistance to CMD but a high photosynthetic rate. Currently, CIAT and Diversity Arrays Technology Pty Ltd are collating and analyzing DArT marker data for 150 F1 progeny of the K family with the aim of saturating the K family genetic

**Table 4** Summary of published genetic linkage maps of cassava

Study	Population type and sample size	Marker numbers	Categories	Linkage groups	Map coverage (cM)	Average distance (cM)
Fregene et al. 1997	F1 ( $n=90$ ) TMS30572 $\times$ CM 2177-2	168	RFLP(132) RAPD(30) SSR(3) Isozymes(3)	20	932	7.9
Okogbenin et al. 2006	F2 ( $n=268$ ) TMS30572 $\times$ CM 2177-2	100	SSRs	22	1236	17.9
Kunkeaw et al. 2010	F1 ( $n=58$ ) Rayong 90 $\times$ Rayong 5	137	AFLP(119) SSR(18)	33	1059	8
Chen et al. 2010	F1 ( $n=210$ ) SC6 $\times$ Mianbao	355	AFLP(231) SSR(41) SRAP(48) ESSR(35)	18	1707	4.8
Kunkeaw et al. 2011	F1 ( $n=100$ ) Huay Bong 60 $\times$ Hanatee	211	AFLP(137) SSR(74)	20	1178	5.6
Sraphet et al. 2011	F1 ( $n=100$ ) Huay Bong 60 $\times$ Hanatee	510	SSRs	23	1420	4.54
Whankaew et al. 2011	F1 ( $n=100$ ) Huay Bong 60 $\times$ Hunatee	303	SSRs	27	1328	5.8

map. This should enable scaffolds of the current cassava genome sequence to be anchored on the K family genetic map (Becerra Lopez-Lavalle per. comm.).

Nearly a decade later, a second SSR-based genetic map consisting mainly of an F2 population derived from a single F1 plant was developed (Okogbenin et al. 2006). More recently, several maps with between 137 and 510 markers comprised of AFLPs and SSRs (both genomic and EST-derived) have been published (Chen et al. 2010; Cortés et al. 2002; Kunkeaw et al. 2010; Kunkeaw et al. 2011; Marín Colorado et al. 2009; Sraphet et al. 2011; Whankaew et al. 2011). Most of these maps were developed from two specific bi-parental crosses, Huay Bong 60  $\times$  Hunatee and TMS 30572  $\times$  CM 2177–2, which together account for five out of the seven current cassava genetic maps. In spite of this common map derivation and the fact that many SSRs are present in multiple maps, cassava lacks a unified consensus map. A number of SSR- and SNP-based genetic maps of cassava are under development, which should increase marker map density.

#### Mapping of Quantitative Trait Loci (QTL)

Molecular markers have been widely used for mapping QTL underlying agronomic traits in many crops. In cassava, QTL controlling cyanogenic glucoside accumulation and dry

matter content (Kizito et al. 2007; Whankaew et al. 2011), plant architecture and productivity (Boonchanawiwat et al. 2011; Okogbenin et al. 2008; Okogbenin and Fregene 2003), bacterial blight (Jorge et al. 2000; Jorge et al. 2001; Lopez et al. 2007; Wydra et al. 2004), wound response in post-harvest physiological deterioration (PPD) (Cortés et al. 2002), carotene levels (Marín Colorado et al. 2009), CMD (Akano et al. 2002) and CBSD (Kulembeka 2011) have been reported. Table 5 provides a summary of some published QTL studies in cassava, including target traits, parents and population type, and markers used.

Unfortunately in many cases, a small sample size and limited number of markers has led to poor resolution of QTL markers. For example, of the nine QTL related to productivity component traits identified by Okogbenin et al. (2008), seven had QTL intervals of between 16 and 44 centi-Morgans (cM), despite the fact that this study employed the largest population size of all the studies, 268 individuals. In addition, as described above, most studies were conducted on the ‘K family’ or derivatives thereof (Table 5). This bi-parental population represents an extremely small proportion of the global cassava diversity, and the identified QTL may have little relevance to QTL segregation in other populations, thus limiting the scope of inference and the application in marker-assisted selection. A more comprehensive dissection of genetic architecture requires development of multiple populations that represent

**Table 5** A summary of QTL studies in cassava

Study/Trait	Parents <sup>a</sup>	Markers used	Comments/Key findings
Jorge et al. 1999 -CBB	TMS30572 × CM2177-2 F1 ( <i>n</i> =150)	68 RFLPs (female map) and 15 (male-derived map)	Identified 8 QTLs in female map (explaining 9–20%) and 4 QTLs on Male map (explaining 10.7–27.1% of variance)
Akano et al. 2002 -Cassava Mosaic Disease (CMD)	TME3 × TMS30555 F1 ( <i>n</i> =80)	186 SSRs	Used bulk segregant analysis (BSA) approach, with 40 susceptible and 40 resistant One SSR (SSRY28) was tightly linked with QTL. Trait was scored qualitatively
Okogbenin and Fregene 2003 -Plant architecture and productivity traits	TMS30572 × CM2177-2 F1 ( <i>n</i> =144)	Fregene et al. 1997 (250 markers)	Reported 114, most are environment specific Several genomic segments QTLs for multiple traits co-localize All QTLs identified for fresh root yield were associated with plant architecture traits
Wydra et al. 2004 -CBB	TMS30572 × CM2177-2 BC1 ( <i>n</i> =111)	63 SSRs	Eleven markers were identified, explaining between 16% and 33.3% of phenotypic variance of area under disease progress curve
Kizito et al. 2007 -CNP and dry matter content (DMC)	MCOL1684 × THAI1 F2 ( <i>n</i> =199)	110 SSRs	Two QTLs controlling CNP and six QTL controlling DMC
Akinbo et al. 2007 -Cassava Anthracnose Disease (CAD)	TME117 × TMS92/0326 F1 ( <i>n</i> =60)	53 RAPDs	Chi-square test revealed a monogenic dominant gene action. Used BSA 2 RAPDs (OPAF2 and OPF06) explained 33.7 and 27.4% ( <i>P</i> <0.05) of the CAD resistance
Lopez et al. 2007 -Cassava bacterial blight (CBB)	TMS30572 × CM2177-2 F1 ( <i>n</i> =150)	10 defence-related genes, 2 BACs containing resistance gene candidates (RGCs) and 340 other regular markers	Uncovered two new QTLs. The largest QTL in the BAC containing RCG explained 61.6% of phenotypic variance
Okogbenin et al. 2008 -Yield components	TMS30572 × CM2177-2 F2 ( <i>n</i> =268)	122 SSRs	Nine QTLs that together explained 33% for fresh foliage, 43.5% for dry root yield and 36% for harvest index
Whankaew et al. 2011 -Cyanogenic potential (CNP)	Hunatee × HuayBong60 F1 ( <i>n</i> =100)	303 SSRs	Five QTLs identified, the largest QTL, CN09R1 explains 26% variance

<sup>a</sup> This population has been extensively used in mapping studies at CIAT

a larger sample of the available genetic variation in the species (Holland 2007). A feasible alternative to the creation of multiple bi-parental populations is the adoption of newer QTL dissection approaches, such as association mapping (Buckler and Thornsberry 2002).

*Genome-Wide Association Mapping* Association mapping or LD mapping has been proposed as an efficient way to determine the genetic basis of complex traits (Abdurakhmonov

and Abdurakhmonov 2008). Association mapping relies on germplasm samples and does not require the development of bi-parental populations. Compared to conventional linkage mapping, association mapping takes advantage of historical LD between genes coding for a trait and closely linked markers for mapping. In comparison, the classical F1-based QTL mapping population is characterized by a small number of recombination events per chromosome (Abdurakhmonov and Abdurakhmonov 2008; Nordborg and



Tavaré 2002; Stich et al. 2006). Association mapping thus has the potential to provide greater map resolution. In addition, the use of diverse germplasm in association mapping enables many alleles and traits to be evaluated simultaneously (Stich et al. 2006).

The applicability and resolution of association mapping and other modern breeding approaches such as genomic selection depend on the extent and structure of LD within the population under consideration. LD is the nonrandom association of alleles at different loci and is affected by the breeding system of the species (selfing versus outcrossing), population structure and genome-wide recombination patterns (Flint-Garcia et al. 2003). Rapid decay in LD, a common feature in outbreeding crops (and cassava is not expected to be an exception) means that a substantially greater number of genetic markers are needed to detect linkage between a marker and a causal locus (Yu and Buckler 2006). To our knowledge, no information is available on genome-wide or intragenic LD in cassava germplasm. This information is required to set the stage for genome-wide scans that will uncover associations between molecular markers and important agronomic traits.

#### Gene Discovery

Cassava production in developing countries is beset by a multitude of pests and diseases (Ceballos et al. 2004; Dixon et al. 2003). It is important that research identifies sources of resistance or tolerance to these continuously evolving stresses and devises strategies to efficiently deploy them in a range of germplasm. Genome-wide surveys have resulted in the identification of about 150 resistance gene analogues (RGAs) in *Arabidopsis* (Meyers et al. 2003; Tan et al. 2007), about 500 in rice (Zhou et al. 2004) and about 400 in poplar (Tuskan et al. 2006). In the pre-genome sequence era, degenerate primers were successfully used to isolate RGAs, resulting in thousands of NBS-LRR (nucleotide binding site-leucine rich repeat) like partial sequences (Bai et al. 2002; Budak et al. 2006; Chen et al. 2007; Gedil et al. 2001; van der Linden et al. 2004) and prompting the formation of dedicated databases (Sanseverino et al. 2010). Candidate genes retrieved by this method were successfully used to develop and map molecular markers co-segregating with disease resistance traits (Calenge et al. 2005a; Calenge et al. 2005b; Moroldo et al. 2008; Qiu et al. 2007) and, in many cases, were found to map close to major resistance gene QTL (Gebhardt et al. 2006; Gebhardt and Valkonen 2001; Speulman et al. 1998).

A previous study in cassava reported the isolation of 12 classes of resistance gene candidates (RGCs), of which two full-length protein coding sequences were identified and mapped on the framework cassava linkage map (Lopez et al. 2003). Using a similar comparative approach, a study

carried out at IITA led to the isolation and characterization of over 500 partial sequences of NBS-LRR-type R genes in cassava and its relatives (*M. glaziovii*, *M. brachyandra*, *M. epruinosa*, *M. tripartita*, and castor bean, *Ricinus communis*) in the Euphorbeaceae family (Gedil et al. submitted). More than half (353 sequences or 64%) of the total sequences had open-reading frames (ORF) uninterrupted with stop codons, whereas the rest of the sequences did not have an ORF, which implies that the genes are not expressed and are pseudogenes that have no functional role. Both TIR (toll interleukin 1 receptor) and non-TIR sub-families were observed by phylogenetic analysis. Multiple sequence alignment (MSA) revealed that the newly identified sequences showed similarity to domains/motifs of known R genes, identifying them as candidate R genes. The candidate sequences matched many homologous R genes in the draft cassava genome with high sequence similarity. This finding furnishes fundamental knowledge about RGAs in cassava and wild *Manihot* species. It is anticipated that understanding the structure, localization, function, variation, and evolution of resistance genes, in combination with other gene and/or genetic mapping approaches, will enable the development of functional gene-targeted markers for use in molecular resistance breeding and of novel strategies for anticipatory and durable disease control (Lawson et al. 2010).

#### The Application of Markers to Cassava Breeding

##### Marker-Assisted Selection (MAS)

Breeding a new variety of cassava usually takes 10 years, due to its long growth cycle (12–18 months). MAS can dramatically increase the precision of selection, leading to more rapid genetic gain fewer cycles of phenotypic evaluation and, thus, reducing the time for varietal development. Under the current molecular breeding scheme used for cassava, varieties could be tested for release in six years. Additionally, use of MAS in the seedling stage dramatically reduces population sizes, making breeding more economical and allowing breeders to work on a larger number of populations.

The only known applications of molecular breeding in cassava are selection for CMD and cassava green mite (CGM) resistance in CIAT and National breeding programs. MAS has rapidly facilitated the breeding for CMD2-mediated resistance in Latin America (in the absence of the pathogen) and in Africa, where the disease is most prevalent (Blair et al. 2007). Other field evaluations have indicated that the markers RME1 and NS158 are excellent predictors of CMD resistance (Okogbenin et al. 2007). CMD resistance was introgressed into improved elite CIAT

lines using markers. These are now referred to as the CR-series (CR families). Two markers (NS1009 and NS346) associated with CGM resistance have been used in MAS (Okogbenin, per. comm.). Combining CMD and CGM resistance, another set of families was developed using markers and denoted as the AR series. Both AR and CR genotypes of the same set were shared and distributed in vitro to African National Agricultural Research Services (NARS) through activities supported under the GCP. This has been a significant achievement, given the susceptibility of Latin American germplasm to CMD. Of the more than 1,000 genotypes introduced into African NARS, several genotypes have been integrated into multi-stage breeding activities, some to the point of varietal selection by farmers in Nigeria, Tanzania, Ghana and Uganda. One of the Latin American-derived cassava varieties, CR41-10 (UMUCASS 33) was released in Nigeria in 2010 and represents the first Latin American variety officially released in Africa. The markers are being used to transfer CMD resistance into desirable genetic backgrounds of East African farmer-preferred varieties with CBSD tolerance in order to combine resistance to both viral diseases.

The threat of CMD requires improved resistance and enhanced durability and has prompted further screening for new sources of CMD resistance in Nigeria. Molecular marker analysis has identified a new source of CMD resistance (*CMD3*) in TMS 97/2205, an IITA-developed variety found to show high CMD resistance in different ecologies that have high to very high disease pressure in Nigeria (Egesi et al. 2007). The near immunity of this variety to CMD has been attributed to the combined effects of *CMD2* and *CMD3* loci. Gene pyramiding, the process of combining several genes together into a single phenotype (Collard and Mackill 2008), has been initiated for CMD resistance breeding using both *CMD2* and *CMD3* genes for enhanced durability and stability of CMD resistance.

Results from MAS-bred CGM genotypes indicate variation in response to the pest. Progenies selected with the markers for CGM resistance tended to show good resistance to the pest in East Africa in contrast to the moderate tolerance observed for CGM in West Africa (Okogbenin, unpublished). The phenotypic differences between African sub-regions could be due to variation in CGM pressure, which was higher at the Umudike test site in Nigeria than in Mtwara and Chambeze, Tanzania, and Namulonge, Uganda.

#### Gene Mining of Wild Relatives for Gene Pool Development, using MAS

Wild *Manihot* germplasm offers a wealth of useful genes for agronomic cassava. Key target breeding traits have been discovered in wild accessions of cassava, including high levels of protein in *M. esculenta* sub spp. *flabellifolia*, *M.*

*peruviana* and *M. tristis* (CIAT 2004), low amylose corn starch (3–5%) or waxy starch in *M. crassiseipala* and *M. chloristicta*, and delayed PPD in an interspecific hybrid between cassava and *M. walkerae* (Bertram 1993). Moderate to high levels of resistance to CGM, whiteflies and the cassava mealybug have been found in interspecific hybrids of *M. esculenta* sub spp. *flabellifolia*. The use of wild species in breeding programs is restricted by linkage drag, requiring pre-breeding activities. However, the use of molecular markers to introgress a single target region of the genome can save two to four backcross generations (Frisch et al. 1999). It is possible that the genetic potential of wild relatives can be released by an advanced backcross QTL mapping scheme (ABC-QTL) in which markers are used for both foreground and background selection (Tanksley and McCouch 1997). ABC-QTL has been used at CIAT to introgress genes for protein content, waxy starch and delayed PPD. Genotypes with QTL of interest and minimum donor parent genome were selected and used for generating advanced backcross populations (Blair et al. 2007).

In the case of introgressing a naturally occurring mutant that creates granule-bound starch in wild relatives, a highly targeted approach was adopted. Sequencing of the glycosyl transferase region of the GBSSI gene from wild relatives and two cassava accessions resulted in the identification of four SNPs that differentiated wild accessions from cassava. These were used to develop allele-specific molecular markers for MAS (Blair et al. 2007). Such allele-specific markers may be used to select genotypes that harbor the recessive mutant gene for future selfing to recover waxy starch. This approach represents an innovative molecular tool to accelerate the introgression of favorable alleles from wild relatives into cassava.

#### Estimation of Heterozygosity and Development of Partial Inbreds

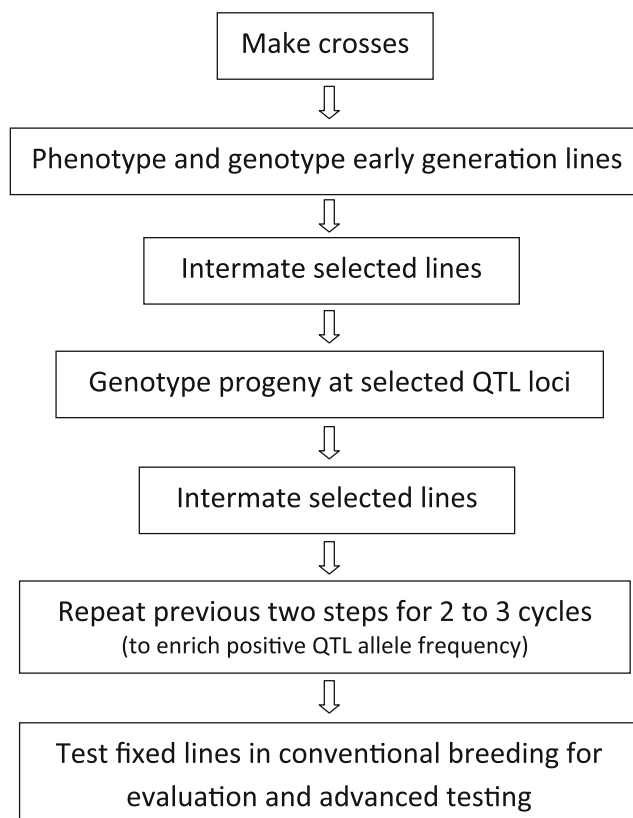
Cassava genotypes are heterozygous and show extensive segregation in F1 breeding populations, making breeding for complex traits very uncertain. Inbred lines are preferred as parents since they do not have the confounding effect of dominant traits masking recessive ones and carry less genetic load. The breeding value (for the trait) of these homozygous  $S_1$  genotypes doubles (if the assumption of heterozygosity for the trait in the elite  $S_0$  genotype holds true). At CIAT efforts are underway to reduce high segregation in the seedling nurseries and to minimize MAS cost at the early stages of the breeding scheme. To do this, favorable alleles were fixed at six target marker loci for both CMD and CGM. Castro et al. (unpublished) showed that after one generation of selfing ( $S_0$  to  $S_1$ ), markers revealed that the general reduction of heterozygosity was 50% by inbreeding effect. Selection for inbreeding tolerance is

biased by the differences in homozygosity levels of segregating partially inbred genotypes. Markers can be used to estimate heterozygosity in selfed lines to permit co-variance correction in the selection of phenotypically vigorous genotypes. Molecular markers are presently being used to assess homozygosity of selfed populations in the development of partial inbred lines in Africa. The markers will be used to determine regions in the genome that are particularly related to the expression of heterosis and to measure genetic distances among inbred lines such that crosses can be conducted with higher probabilities of success (Ceballos et al. 2004). It is hoped that this will improve the prospects of developing seed technology for cassava involving the use of seeds as propagules and of hybrid development.

#### Marker Assisted Recurrent Selection (MARS) for Complex Traits

MAS is useful for pyramiding genes of relatively large effect, such as disease resistance genes (Jia et al. 2002; Komori et al. 2003; Murai et al. 2001). However, most agronomic traits are quantitative in nature, and their manipulation has been challenging due to a complexity of interactive factors such as epistasis, pleiotropy and genotype-by-environment interaction. Breeding for complex traits is expensive due to the need for highly replicated phenotyping trials over several environments. This is driving the quest for a MAB approach that increases precision of selection and reduces the requirement for phenotyping.

MARS is a MAB strategy for forward breeding of genes and QTLs for relatively complex traits (Crosbie et al. 2006; Eathington 2005; Ragot et al. 2000; Ribaut and Betran 1999). Here, QTL mapping is performed in the F1 from a biparental cross in which both parents contribute favorable alleles with the ideal genotype being a mosaic of beneficial alleles from both parents (Ragot and Lee 2007). Several generations of crossing and genotypic selection are done for each phenotyping trial. MARS is essentially a genotype construction process that leads to an increase in the frequency of beneficial alleles and to the development of individuals having the best haplotype combination at selected loci in the genome. The principle can be extended to multi-parental populations where favorable alleles come from more than two parents (Peleman and van der Voort 2003). A typical MARS scheme is illustrated in Fig. 1. Under the GCP - Cassava Challenge Initiative, African breeding programs have initiated MARS for drought tolerance breeding in cassava. SSR and SNP markers are used to identify QTLs and then to identify important allele combinations through three cycles of selection, which is only then followed by phenotyping.



**Fig. 1** MARS scheme (Adapted from Integrated Breeding Platform – Generation Challenge Programme)

#### Genomic Selection for Complex Traits

A proposed alternative approach to dealing with multiple loci conferring small effects is referred to as ‘genomic selection’ (GS) (Meuwissen et al. 2001). This approach is facilitated by new high-throughput genotyping and novel statistical methods. Unlike traditional MAS, which relies on knowledge of individual loci associated with a specific trait, GS uses all marker data as predictors of performance, thus enabling the selection for multiple loci of small genetic effect (Jannink et al. 2010). Essentially, breeding populations are extensively genotyped (possibly using GBS) to give full genome coverage and phenotyped to create models that calculate genomic estimates of breeding values (GEBVs). GEBVs are then used as a criterion for selecting candidate parents. These values can then be used for selection within a breeding population, without the need for phenotypic evaluation. Simulation and empirical studies indicate that GS can lead to a considerable increase in the rate of genetic gain while dramatically reducing the need for phenotypic evaluation. Benefits of genomic selection are being experienced in the dairy cattle industry (Hayes et al. 2011). In addition, though GEBVs do not show the effects of underlying genes, simulation studies calculate that it is remarkably accurate (Habier et al. 2007; Zhong et al. 2009).

This new breeding approach has been comprehensively reviewed (Heffner et al. 2009; Jannink et al. 2010) and should provide significant benefits in breeding for some highly quantitative traits in cassava.

## Conclusions

Previously, molecular markers, predominantly SSR markers, have been used in cassava to understand genetic diversity and differentiation in populations, to map QTLs associated with large effect genes and to mine genes from wild species. In the past, marker numbers have been a limitation. The advent of relatively low-cost, massively-parallel, high-throughput genome sequencing has made high-density SNP discovery feasible. Simultaneously, relatively low cost, flexible, low to medium density SNP genotyping technologies have been developed. It is envisaged that these platforms, at least in the near to medium term, will serve the cassava molecular breeding community for short-term breeding applications such as MAS and MARS. This has heralded a new era for the application of molecular markers to plant breeding. As GBS and whole genome re-sequencing become more available, they are likely to enable the application of genome-wide marker selection, particularly useful for more complex traits. With the availability of the cassava genome sequence, the cassava community is poised to take advantage of these new tools for rapid progress and genetic gain. As described above, we envisage the application of these tools in many different ways, including the development of high-density maps and fine mapping, association mapping, exploration of the genome sequence for gene discovery, transcript profiling, inbred line development and the prediction of heterosis, gene mining in wild species and introgressions, short-term breeding applications such as MAS and MARS, and genome-wide selection approaches such as genomic selection. Some of these applications are already underway in cassava.

## References

- Abdurakhmonov IY, Abdurkarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int J Plant Genomics*. doi:10.1155/2008/574927
- Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. *Proceedings of the National Academy of Sciences of the United States of America* 90:7980–7984
- Akano AO, Dixon AGO, Mba C, Barrera E, Fregene M (2002) Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor Appl Genet* 105:521–525
- Anderson JV, Horvath D, Chao WS, Foley ME, Hernandez A, Thimmapuram J, Lie L, Gong GL, Band M, Kim R, Mikel MA (2007) Characterization of an EST database for the perennial weed leafy spurge: an important resource for weed biology research. *Weed Science* 55:193–203
- Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. In: Gustafson JP, Langridge P, Somers DJ (eds) *Plant genomics*. Humana Press, pp 19–39
- Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res* 12:1871–1884
- Becker H (1993) *Pflanzenzüchtung*. Eugen Ulmer Verlag, Stuttgart
- Beeching JR, Marmey P, Gavada M, Noirot M, Haysom H, Hughes M, Charrier A (1993) An assessment of genetic diversity within a collection of cassava (*Manihot esculenta* Crantz.) germplasm using molecular markers. *Ann Bot* 72:515–520
- Benesi IRM (2005) Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis. University of the Free State, South Africa
- Bertram RB (1993) Application of molecular techniques to genetic resources of cassava (*Manihot esculenta* Crantz, Euphorbiaceae): interspecific evolutionary relationships and interspecific characterization. PhD dissertation. University of Maryland
- Blair M, Fregene M, Beebe S, Ceballos H (2007) Marker-assisted breeding in common beans and cassava. Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organisation of the United Nations (FAO), Rome, pp 81–115
- Boonchanawiwat A, Sraphet S, Boonseng O, Lightfoot DA, Triwitayakorn K (2011) QTL underlying plant and first branch height in cassava (*Manihot esculenta* Crantz). *Field Crops Research* 121:343–349
- Brumfield R, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18:249–256
- Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5:107–111
- Budak H, Su S, Ergen N (2006) Revealing constitutively expressed resistance genes in *Agrostis* species using PCR-based motif-directed RNA fingerprinting. *Genet Res* 88:165–175
- Calenge F, Van der Linden CG, Van de Weg E, Schouten HJ, Van Arkel G, Denancé C, Durel CE (2005a) Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. *Theor Appl Genet* 110:660–668
- Calenge F, van der Linden CG, Van de WE, Schouten HJ, Van AG, Denance C, Durel CE (2005b) Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. *Theor Appl Genet* 110:660–668
- Carter S, Fresco L, Jones P, Fairbairn J (1992) *An Atlas of cassava in Africa: historical, agroecological and demographic aspects of crop distribution*. CIAT, Cali
- Ceballos H, Iglesias CA, Perez JC, Dixon AG (2004) Cassava breeding: opportunities and challenges. *Plant Mol Biol* 56:503–516
- Chavarriga-Aguirre PP, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theor Appl Genet* 97:493–501
- Chavarriga-Aguirre P, Ma M, Tohme J, Duque M, Iglesias C, Bonierbale M, Kresovich S, Kochert G (1999) Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. *Mol Breed* 5:263–273
- Chen G, Pan D, Zhou Y, Lin S, Ke X (2007) Diversity and evolutionary relationship of nucleotide binding site-encoding disease-



- resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.). *J Biosci* 32:713–721
- Chen X, Xia Z, Fu Y, Lu C, Wang W (2010) Constructing a genetic linkage map using an F1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). *Plant Mol Biol Rep* 28:676–683
- Ching A, Caldwell K, Jung M, Dolan M, Smith O, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3:19
- Cho RJ et al (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat Genet* 23:203–207
- CIAT. Annual Report. (2004) Cali, Colombia
- Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans R Soc B Biol Sci* 363:557–572
- Cortés D, Reilly K, Okogbenin J, Beeching JR, Iglesias C, Tohme J (2002) Mapping wound-response genes involved in post-harvest physiological deterioration (PPD) of cassava (*Manihot esculenta* Crantz). *Euphytica* 128:47–53
- Crosbie TM, Eathington SR, Johnson GR, Edwards M, Reiter R, Stark S, Mohanty RG, Oyervides M, Buchler RE, Walker AK, Dodert R, Delannay X, Pershing JC, Hall MA, Lamkey KR (2006) Plant breeding: past, present and future. In: Lampkin L, Lee M (eds) *Plant breeding: the Arnel R. Hallauer International Symposium*. Blackwell, Ames
- Crow JF (1998) 90 years ago: the beginning of hybrid maize. *Genetics* 148:923–928
- Dixon AGO, Bandyopadhyay R, Coyne D, Ferguson M, Shaun R, Ferris B, Hanna R, Hughes J, Ingelbrecht I, Legg J, Mahungu N, Manyong V, Mowbray D, Neuenschwander P, Whyte J, Hartmann P, Ortiz R (2003) Cassava: from poor farmer's crop to pacesetter of African rural development. *Chronica Hort* 43:8–15
- Eathington SR (2005) Practical applications in molecular technology in the development of commercial maize hybrids. Proc. 60th Ann. Corn and Sorghum Seed Res. Conf. Washington DC, USA. American Seed Trade Association
- Egesi C, Ogbe F, Akoroda M, Ilona P, Dixon A (2007) Resistance profile of improved cassava germplasm to cassava mosaic disease. *Euphytica* 155:215–224
- Elias M, Panaud O, Robert T (2000) Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system, using AFLP markers. *Heredity* 85:219–230
- Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K et al (2011) A robust simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Ferguson M, Hearne S, Close T, Wanamaker S, Moskal W, Town C, de Young J, Marri P, Rabbi I, de Villiers E (2011) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor Appl Genet*. doi:10.1007/s00122-011-1739-9
- Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Ann Rev Plant Biol* 54:357–374
- Fregene M, Angel F, Gomez R, Rodriguez F, Chavarriaga P, Roca W, Tohme J, Bonierbale M (1997) A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theor Appl Genet* 95:431–441
- Fregene M, Bernal A, Duque M, Dixon A, Tohme J (2000) AFLP analysis of African cassava (*Manihot esculenta* Crantz.) germplasm resistant to the cassava mosaic disease (CMD). *Theor Appl Genet* 100:678–685
- Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AG, Dean R, Kresovich S (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theor Appl Genet* 107:1083–1093
- Frisch M, Bohn M, Melchinger A (1999) Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci* 39:1295–1301
- Gebhardt C, Valkonen JP (2001) Organization of genes controlling disease resistance in the potato genome. *Annu Rev Phytopathol* 39:79–102
- Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JP (2006) Marker-assisted combination of major genes for pathogen resistance in potato. *Theor Appl Genet* 112:1458–1464
- Gedil MA, Slabaugh MB, Berry S, Johnson R, Michelmore R, Miller J, Gulya T, Knapp SJ (2001) Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: genetic mapping and linkage to the downy mildew resistance gene P11. *Genome* 44:205–212
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397
- Hackett CA, Wachira FN, Paul S, Powell W, Waugh R (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). *Heredity* 85:346–355
- Hahn SK, Howland AK, Terry ER (1980) Correlated resistance of cassava to mosaic and bacterial blight diseases. *Euphytica* 29:305–311
- Hayes BJ, Bowman P, Chamberlain A, Goddard ME (2011) Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–443
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Heffner EL, Sorrells ME, Jannink J-L (2011) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Hippolyte I et al (2010) A saturated SSR/DarT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biol* 10:65
- Holland J (2007) Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* 10:156–161
- Hurtado P, Ospina C, Marin J, Buitrago C, Castelblanco W, Correa A, Alfonso P, Barrera E, Gutierrez J, Fregene M, Hearne S, Ferguson M, Alves A, Fortes-Ferreira C, De Vicente C (2008) Assessment of the diversity in global cassava genetic resources based on simple sequence repeat (SSR) markers. In: Fauquet CM (ed) *Cassava: Meeting the Challenges of the New Millennium*. Proceedings of the First Scientific Meeting of the Global Cassava Partnership. 21–25 July, 2008. Ghent, Belgium
- Ingelbrecht IL, Jorgensen K, Bak S, Gorodkin J, Raji A, Winter S, Lokko Y, Gedil M, Anderson JV, Moller B, Dixon AGO (2008) Utilization of ESTs from cassava: progress on development of EST-SSR markers and an oligo DNA microarray. First Scientific Meeting of the Global Cassava Partnership GCP-1. Cassava: Meeting the Challenges of the New Millennium. 7–21-2008
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177
- Jennings D, Iglesias C (2001) Breeding for crop improvement. In: Hillocks R, Thresh J (eds) *Cassava biology, production and utilization*. CAB International, Oxon, pp 149–166
- Jia Y, Wang Z, Singh P (2002) Development of dominant rice balst Pita rissistance gene markers. *Crop Sci* 42:2145–2149
- Jones W (1959) *Manioc in Africa*. Stanford University Press, Stanford
- Jorge V, Fregene MA, Duque MC, Bonierbale MW, Tohme J, Verdier V (2000) Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). *Theor Appl Genet* 101:865–872

- Jorge V, Fregene M, Vélez CM, Duque MC, Tohme J, Verdier V (2001) QTL analysis of field resistance to *Xanthomonas axonopodis* pv. *manihotis* in cassava. *Theor Appl Genet* 102:564–571
- Kawuki RS (2009) Variation in cassava (*Manihot esculenta* Crantz.) based on single nucleotide polymorphisms, simple sequence repeats and phenotypic traits. PhD thesis. Department of Plant Sciences: Plant Breeding. University of the Free State, Bloemfontein, South Africa
- Kawuki R, Ferguson M, Labuschagne M, Herselman L, Kim DJ (2009) Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol Breed* 23:669–684
- Kizito E, Bua A, Fregene M, Egwang T, Gullberg U, Westerbergh A (2005) The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. *Euphytica* 146:45–54
- Kizito E, Rönnberg-Wänstljung AC, Egwang T, Gullberg U, Fregene M, Westerbergh A (2007) Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas* 144:129–136
- Komori T, Yamamoto T, Takemori N, Kashiwara M, Matsushima H, Nitta N (2003) Fine genetic mapping of the nuclear gene, Rf-1, that restores the BT-type cytoplasmic male sterility in rice (*Oryza sativa* L.) by PCR-based markers. *Euphytica* 129:241–247
- Kulembeka HPK (2011) Genetic linkage mapping of field resistance to cassava brown streak disease in cassava (*Manihot esculenta* Crantz.) landraces from Tanzania. PhD Thesis. Department of Plant Sciences (Plant Breeding), University of the Free State, South Africa
- Kunkeaw S, Tangphatsomruang S, Smith DR, Triwitayakorn K (2010) Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. *Plant Breed* 129:112–115
- Kunkeaw S, Yoocha T, Sraphet S, Boonchanawiwat A, Boonseng O, Lightfoot D, Triwitayakorn K, Tangphatsomruang S (2011) Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihot esculenta* Crantz). *Mol Breed* 27:67–75
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Langlands B (1966) Cassava in Uganda 1860–1920. *Uganda J* 30:211–218
- Lawson WR, Jan CC, Shatte T, Smith L, Kong GA, Kochman JK (2010) DNA markers linked to the R<sub>2</sub> rust resistance gene in sunflower (*Helianthus annuus* L.) facilitate anticipatory breeding for this disease variant. *Mol Breed Online first*:1–8
- Legg J, Owor B, Sseruwagi P, Ndunguru J (2006) Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Adv Virus Res* 67:355–418
- Litt M, Luty JM (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle action gene. *Am J Hum Genet* 44:397–401
- Lokko Y, Anderson J, Rudd S, Raji A, Horvath D, Mikel M, Kim R, Liu L, Hernandez A, Dixon A, Ingelbrecht I (2007) Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Rep* 26:1605–1618
- Lopez CE, Zuluaga AP, Cooke R, Delseny M, Tohme J, Verdier V (2003) Isolation of Resistance Gene Candidates (RGCs) and characterization of an RGC cluster in cassava. *Mol Genet Genomics* 269:658–671
- Lopez C, Jorge V, Piegu B, Mba C, Cortes D, Restrepo S, Soto M, Laudie M, Berger C, Cooke R, Delseny M, Tohme J, Verdier V (2004) A unigene catalogue of 5700 expressed genes in cassava. *Plant Mol Biol* 56:541–554
- Lopez C, Piegu B, Cooke R, Delseny M, Tohme J, Verdier V (2005) Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). *Theor Appl Genet* 110:425–431
- Lopez C, Quesada-Ocampo LM, Bohorquez A, Duque MC, Vargas J, Tohme J, Verdier V (2007) Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome* 50:1078–1088
- Marín Colorado J, Ramírez H, Fregene M (2009) Genetic mapping and QTL analysis for carotene in an S1 population of cassava. *Acta Agron, Universidad Nacional de Colombia* 58:15–21
- Marmey P, Beeching J, Hamon S, Charrier A (1993) Evaluation of cassava (*Manihot esculenta* Crantz.) germplasm using RAPD markers. *Euphytica* 74:203–209
- Mba REC, Stephensen P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theor Appl Genet* 102:21–31
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell* 15:809–834
- Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di GG, Morgante M, dam-Blondon AF (2008) A physical map of the heterozygous grapevine ‘Cabernet Sauvignon’ allows mapping candidate genes for disease resistance. *BMC Plant Biol* 8:66
- Murai H, Hashimoto Z, Sharma PN, Shimizu T, Murata K, Takumi S, Mori N, Kawasaki S, Nakamura C (2001) Construction of a high-resolution linkage map of a rice brown planthopper (*Nilaparvata lugens*) resistance gene bph2. *Theor Appl Genet* 103:526–532
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends Genet* 18:83–90
- Nweke FI, Spencer DSC, Lynam JK (2002) The cassava transformation: Africa’s best-kept secret. Michigan State University Press, p 273
- Okogbenin E, Fregene M (2003) Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in cassava (*Manihot esculenta* Crantz.). *Theor Appl Genet* 107:1452–1462
- Okogbenin E, Marín J, Fregene MA (2006) An SSR-based molecular genetic map of cassava. *Euphytica* 147:433–440
- Okogbenin E, Porto M, Egesi C, Mba C, Espinosa E, Santos L, Ospina C, Marín J, Barrera E, Gutierrez J, Ekanayake C, Iglesias C, Fregene MA (2007) Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. *Crop Sci* 47:1895–1904
- Okogbenin E, Marín J, Fregene M (2008) QTL analysis for early yield in a pseudo F2 population of cassava. *Afr J Biotechnol* 7:131–138
- Olsen KM (2004) SNPs, SSRs and inferences on cassava’s origin. *Plant Mol Biol* 56:517–526
- Olsen K, Schaal B (2001) Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. *Am J Bot* 88:131–142
- Peleman J, van der Voort J (2003) Breeding by design. *Trend Plant Sci* 7:330–334
- Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2011) The cassava genome: current progress, future directions. *Tropical Plant Biol.* doi:10.1007/s12042-011-9088-z
- Qiu JW, Schurch AC, Yahiaoui N, Dong LL, Fan HJ, Zhang ZJ, Keller B, Ling HQ (2007) Physical mapping and identification of a

- candidate for the leaf rust resistance gene *Lr1* of wheat. *Theor Appl Genet* 115:159–168
- Queller D, Strassman J, Hughes C (1993) Microsatellites and Kinship. *Trends Ecol Evol* 8:285–288
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94–100
- Ragot M, Lee M (2007) Marker-assisted selection in maize: current status, potential limitation and perspectives from the private and public sectors. Marker assisted selection: current status and future perspectives in crops, livestock, forestry and fish. FAO, Rome, pp 117–150
- Ragot M, Gay G, Muller J, Durovjay J (2000) Efficient selection for adaptation to the environment through mapping and manipulation in maize. In: Ribaut J-M, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments. CIMMYT, Mexico, pp 128–130
- Raji A, Anderson J, Kolade O, Ugwu C, Dixon A, Ingelbrecht I (2009) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant Biol* 9:118
- Reif J, Melchinger A, Xia X, Warburton ML, Hoisington D, Vasal S, Srinivasan G, Bohn M, Frisch M (2003) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Sci* 43:1275–1282
- Ribaut J-M, Betran J (1999) Single large-scale marker-assisted selection (SLS-MAS). *Mol Breed* 5:531–541
- Roa AC, Maya MM, Duque MC, Tohme J, Allem A, Bonierbale MW (1997) AFLP analysis of relationships among cassava and other *Manihot* species. *Theor Appl Genet* 95:745–750
- Rojas M, Pérez J, Ceballos H, Baena D, Morante N, Calle F (2011) Analysis of inbreeding depression in eight S1 cassava families. *Crop Sci* 49:543–548
- Sakurai T, Plata G, Rodriguez-Zapata F, Seki M, Salcedo A, Toyoda A, Ishiwata A, Tohme J, Sakaki Y, Shinozaki K, Ishitani M (2007) Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biol* 7:66
- Salmaso M, Faes G, Segala C, Stefanini M, Salakhutdinov I, Zyprian E, Toepfer R, Grando MS, Velasco R (2005) Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. *Mol Breed* 14:385–395
- Sanseverino W, Roma G, De SM, Faino L, Melito S, Stupka E, Fruscianta L, Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res* 38:D814–D821
- Sewell MM, Sherman BK, Neale DB (1999) A consensus map for loblolly pine (*Pinus taeda* L.). I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. *Genetics* 151:321–330
- Speulman E, Bouchez D, Holub EB, Beynon JL (1998) Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. *Plant J* 14:467–474
- Sraphet S, Boonchanawiwat A, Thanyasiriwat T, Boonseng O, Tabata S, Sasamoto S, Shirasawa K, Isoe S, Lightfoot D, Tangphatsornruang S, Triwitayakorn K (2011) SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor Appl Genet* 122:1161–1170
- Stich B, Maurer H, Melchinger A, Frisch M, Heckenberger M, van der Voort J, Peleman J, Sorensen A, Reif J (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol Breed* 17:217–226
- Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2:930–942
- Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, Bent AF, Michelmore RW (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol* 7:56
- Tangphatsornruang S, Sraphet S, Singh R, Okogbenin E, Fregene M, Triwitayakorn K (2008) Development of polymorphic markers from expressed sequence tags of *Manihot esculenta* Crantz. *Mol Ecol Res* 8:682–685
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:1063–1066
- Tautz D, Trick M, Dover G (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322:652–656
- Tuskan GA et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604
- Ubi BE, Fujimori M, Mano Y, Komatsu T (2004) A genetic linkage map of rhodesgrass based on an F1 pseudo-testcross population. *Plant Breed* 123:247–253
- Utsumi Y, Sakurai T, Umemura Y, Ayling S, Ishitani M, Narangajavana J, Sojikul P, Triwitayakorn K, Matsui M, Manabe R-i, Shinozaki K, Seki M (2011) RIKEN cassava initiative: establishment of a cassava functional genomics platform. *Tropical Plant Biol*. doi:10.1007/s12042-011-9089-y
- van der Linden CG, Wouters DC, Mihalka V, Kochieva EZ, Smulders MJ, Vosman B (2004) Efficient targeting of plant disease resistance loci using NBS profiling. *Theor Appl Genet* 109:384–393
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23:48–55
- Whankaew S, Poopear S, Kanjanawattanawong S, Tangphatsornruang S, Boonseng O, Lightfoot D, Triwitayakorn K (2011) A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. *BMC Genom* 12:266
- Wydra K, Zinsou V, Jorge V, Verdier V (2004) Identification of pathotypes of *Xanthomonas axonopodis* pv. *manihotis* in Africa and detection of quantitative trait loci and markers for resistance to bacterial blight of cassava. *Phytopathology* 94:1084–1093
- Xia L, Peng K, Yang S, Wenzl P, Carmen de Vicente M, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* 110:1092–1098
- Yang J, An D, Zhang P (2011) Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *J Integr Plant Biol* 53:193–211
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization in maize. *Curr Opin Biotechnol* 17:155–160
- Zacarias A, Botha A, Labuschagne M, Benesi I (2004) Characterisation and genetic distance analysis of cassava (*Manihot esculenta* Crantz.) germplasm from Mozambique using RAPD fingerprinting. *Euphytica* 138:49–53
- Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182:355–364
- Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics* 271:402–415