academicJournals

Vol. 14(2), pp. 96-107, 14 January, 2015
DOI: 10.5897/AJB2014.14113
Article Number: 4E2A61249618
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Identification of genes that have undergone adaptive evolution in cassava (*Manihot esculenta*) and that may confer resistance to cassava brown streak disease

M. Ferguson[1], T. M. Koga[2], D. A. Johnson[2], K. A. Koga[2], G. N. Hirsch[2], L. A. Becerra Lopez-Lavalle[3,4] and W. Messier[2]*

[1]International Institute of Tropical Agriculture (IITA), c/o International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi 00100, Kenya.
[2]Evolutionary Genomics Inc., 1801 Sunset Place, Suite C, Longmont, CO, 80501, USA.
[3]International Center for Tropical Agriculture (CIAT), Cali, Colombia.
[4]CGIAR Research Program for Root Tubers and Bananas.

Received 18 August, 2014; Accepted 22 December, 2014

Cassava (*Manihot esculenta*) is a vital food security crop and staple in Africa, yet cassava brown streak disease (CBSD) and cassava mosaic disease result in substantial yield losses. The aim of this study was to identify genes that have undergone positive selection during adaptive evolution, from CBSD resistant, tolerant and susceptible *M. esculenta* varieties and inter-specific hybrids, as well as a wild cassava species. Transcriptomes of 13 genotypes were sequenced and three genes with strong positive selection were detected (designated as EG2771, EG964 and EG5651). Sequence variation for candidate genes in 18 different cassava genotypes was examined in relation to known response to CBSD and whitefly infection. Although, we cannot ascribe a selection pressure that was responsible for the observed positive selection with complete certainty at this stage, given the congruence of the pattern of particular alleles of our positively selected genes and the pattern of disease resistance of the cassava varieties we examined, it is likely that some protein variants coded by alleles of EG2771 and EG964 may be associated with CBSD and whitefly resistance responses. This warrants further investigation. Other alleles of our positively selected genes were likely influenced by domestication or some other unknown selective pressure.

Key words: Cassava, cassava brown streak disease (CBSD), resistant, tolerant, susceptible.

INTRODUCTION

The starchy roots of cassava (*Manihot esculenta* Crantz) form one of the most important food staples in many countries of Sub-Saharan Africa (Herrera Campo et al., 2011). It is largely agreed that cassava was domesticated from a single wild species, *M. esculenta* ssp. *flabellifolia* (Allem, 1994; 2002; Olsen, 2004), although this may have

*Corresponding author. E-mail: wmessier@evolgen.com. Tel: 303-862-3222.

occurred more than once over time, and in distant sites across the Brazilian states of Rondónia, Mato Grosso and Goiás (Allem, 2002). The exact location(s) of cassava domestication is still debated (Lebot, 2009). Cassava was introduced into Africa by Portuguese traders around 1550, where it spread relatively slowly. Cassava cultivation and consumption in Africa only became significant in the late 1800s (Jones, 1959). Africa is now the largest producer of cassava with approximately 13 million hectares (ha) harvested in 2011, compared to 19 million ha harvested globally, although average yields are among the lowest (nearly 12t/ha, compared to 18t/ha in China) (FAO, 2013). Two virus-caused diseases, cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) cause significant yield losses in many countries in Africa. CMD is caused by cassava mosaic geminiviruses (CMGs) (Fauquet et al., 2003; Legg, 2008), and is found wherever cassava is grown in Africa (Thresh and Cooter, 2005). A particularly virulent combination of CMGs resulted in a widespread pandemic in the Great Lakes region (Uganda, Kenya and Tanzania) from the late 1980s, which was largely controlled through the dissemination and growth of resistant varieties (Legg et al., 2006). The pandemic was associated with up to 100-fold increases in the abundance of the whitefly vector (Legg and Ogwal, 1998). CBSD is caused by two viruses in the family Potyviridae, genus *Ipomovirus* (Mbanzibwa et al., 2009; Monger et al., 2001; 2010; Winter et al., 2010). The disease was initially restricted to lowland coastal areas of Kenya, Tanzania and Mozambique, as well as the surrounds of Lake Malawi in Tanzania and Malawi (Hillocks et al., 1999; Nichols, 1950; Legg and Raya, 1998), but from 2004 the spread of CBSD to higher altitudes and inland to countries surrounding the Great Lakes region was reported. CBSD is now considered a major threat to food security in Uganda and its spread is threatening West Africa. Both of these virus diseases are transmitted by the whitefly, *Bemisia tabaci* (Gennadius) (Maruthi et al., 2005; Maruthi et al., 2014) and disseminated through infected cuttings of this vegetatively propagated crop (Legg et al. 2011). The most realistic approach to reducing losses to CMD and CBSD is the use of host-plant resistance or deployment of less-susceptible cultivars (Hillocks et al., 2003). The urgency for farmer-preferred varieties that combine CBSD and CMD resistance is acute.

Breeding programs were launched independently in Tanzania (Jennings, 1957) and Madagascar (Cours, 1951; Cours et al., 1997). Cassava breeding started in 1935 at Amani in Tanzania to develop cultivars resistant to both CMD and CBSD (Jennings, 1957). Initially, a large number of cassava cultivars were evaluated; however, high enough levels of resistance or tolerance to CMD to form the basis of a breeding program were not found. Emphasis was therefore given to transferring resistance to cassava from related species. Tree-like species used were *Manihot glaziovii* Muell-Arg., *Manihot dichotoma* Ule., *Manihot catingae* Ule., and 'tree' cassava, believed to be a natural hybrid between *M. esculenta* and *M. glaziovii* (Jennings and Iglesias, 2002). *M. glaziovii* or Cerea rubber was introduced as a plantation crop for rubber production. *Manihot melanobasis* Muell-Arg. and *Manihot saxicola* Lanj, two shrub-like species from Surinam (Hillocks and Jennings, 2003) were also used. These latar two species are now regarded as forms of *M. esculenta* subsp. *flabellifolia* (Allem, 2002)]. The hybrids derived from *M. glaziovii* and *M. melanobasis* were the most promising in terms of resistance/tolerance to CMD and CBSD and priority was given to these. The most tolerant of these is known as 'Kaleso' in Kenya and as 'Namikonga' in Tanzania (Hillocks and Jennings, 2003) and has recently been shown to be resistant to (U)CBSV with 'Kiroba' being tolerant and 'Albert' being susceptible (Bouvaine and Gowda, 2014). 'Namikonga' is infected by the virus and shows reduced leaf symptoms but the virus titer is kept low and the onset of root necrosis is delayed until at least 12 to 18 months after planting. Namikonga has existed under disease pressure since it was developed in the 1930s or 40s and resistance has so far proved to be durable. The mode of inheritance of resistance has not been resolved.

Surveys conducted in Tanzania and Mozambique indicated that some local cultivars showed tolerance to CBSD, and it is likely that natural selection has occurred in areas of high disease pressure. 'Nachinyaya' was identified in 1995 from southern Tanzania (Thresh, 2003) with delayed onset of necrosis (Hillocks, 2003). Recently, screening of germplasm has been extended to Mozambique, Kenya, Uganda and Tanzania, but few resistant varieties have been identified. The International Center for Tropical Agriculture (CIAT) has also been screening wild cassava species for useful variation in terms of pest and disease resistance, nutritional attributes and abiotic stress tolerance (Carabalí et al., 2010a; Carabalí et al., 2010b).

Whiteflies (Hemiptera: Aleyrodidae) are the most important cassava pests in the Americas and Africa and to a lesser extent Asia. Twelve species including *Aleurotrachelus socialis* and *B. tabaci* are serious pests of cassava (Bellotti, 2002; Herrera Campo et al., 2011). Whiteflies consume large quantities of phloem sap from cassava leaves, which causes leaf yellowing and premature leaf drop. Large populations of whiteflies can severely reduce the quality and mass of cassava roots and therefore have a substantial impact on global food security (Bellotti and Arias, 2001). For example, *A. socialis*, which is a serious pest of cassava in northern South America and Brazil, can cause a root yield reduction of over 79% in a single year (Bellotti and Arias, 2001; Vargas and Bellotti, 1983). Prolonged whitefly feeding alone (in the absence of viruses) can cause serious economic losses due to yield reductions of more

than 50% (Byrne and Bellows, 1991); thus in Brazil alone, accumulated losses between 1995 and 2000 neared $1 billion per year (Oliveira et al. 2001).

The whitefly *B. tabaci* is also a vector for the viruses that cause CMD and CBSD in Africa (Herrera Campo et al., 2011; Hillocks et al., 2002; Legg et al., 2011). CMD epidemics are correlated with high population densities of *B. tabaci* at the epidemic front (Herrera Campo et al., 2011; Legg et al., 2011). Given the strict correlation of virus epidemics and high whitefly population levels, pyramiding host-plant resistance to whiteflies with resistance to geminiviruses should be the optimal strategy for enhancing cassava productivity globally (Lokko et al., 2006; Patil and Fauquet, 2009). To better screen for new sources of variation, and efficiently transfer useful variation into elite varieties, it is advantageous to identify molecular markers associated with traits of interest, or better still identify gene(s) that influence those traits. A number of approaches have been used to aid gene discovery. These include QTL mapping, association mapping and map-based cloning. Another approach, which we have used here, is to identify the genes that have undergone adaptive evolution (positive selection) in response to a natural or imposed selection pressure (Messier and Stewart, 1997).

The Adapted Traits Platform, designed at Evolutionary Genomics, Inc. (EG) performs high throughput molecular evolution analysis (Messier and Sikela, 2001) to identify positively selected genes based on Ka/Ks analysis (Li et al., 1985; Li 1993). Ka/Ks analysis was originally developed to document the role of positive selection on known protein coding genes. This analysis is used here as a novel tool to pinpoint previously unsuspected positively-selected genes from large transcriptome data sets, in an attempt to identify cassava gene candidates for CBSD resistance.

Here, we analyze the genomes of wild species, interspecific hybrids and cultivated cassava, *M. esculenta*, that have a range of responses to CBSD, CMD and whitefly, as well as other traits related to domestication in terms of storage root formation, to identify genes that have undergone positive selection and adaptive evolution. The goal of this study was to discover candidate cassava genes that can be used to help control pest and disease resistance in cassava, particularly CBSD; these genes can now be further validated.

MATERIALS AND METHODS

Germplasm

Leaf samples were collected in a Cryo Express Dry Shipper (Taylor-Wharton) from 10 cassava varieties that produce storage roots, one wild species (*M. glaziovii*) that does not produce storage roots, and four inter-specific hybrids that produce small fibrous storage roots. One young leaf was sampled from each of one plant per genotype. The fifth fully expanded leaf from the top of the plant was sampled where possible. Table 1 provides information regarding the germplasm used, including pedigree details (where known), and response to CMD and CBSD (where known). *M. esculenta* leaf samples were either collected from a screenhouse or field, 'tree' cassava was collected at the Sugarcane Research Station, Kibaha, Tanzania, 'homestead' cassava was collected from homesteads close to Dar-es-Salaam and *M. glaziovii* was collected from a wild population at Sirawi in Tanzania. Both wild and cultivated cassava germplasm from Latin American germplasm was obtained from CIAT's germplasm collection. These materials were selected because of their contrasting varietal response to CBSD and to whitefly (*A. socialis*) infestation, a fundamental part of our experimental design. The germplasm chosen represents the best source of resistance/tolerance to CBSD that is currently available. Where pedigree information is available, lines were selected that do not share a common resistant ancestor and therefore reflect different sources of resistance. Namikonga, Nachinyaya, Mkombozi, NDL06/132, Kiroba and AR40-6 are the resistant parents of QTL mapping populations where Albert, AR37-80 and TMS4(2)1425 form susceptible parents. Other 'tree cassava' accessions were selected to represent the evolutionary continuum, and lines from South America for whitefly resistance.

RNA extraction and sequencing

Total RNA from African germplasm was extracted at the ILRI-BecA hub using a cetyl trimethyl ammonium bromide (CTAB) method. Total RNA from CIAT-collected germplasm was extracted using RNeasy Plant Mini Kit (Qiagen). RNA quality control and characterization were conducted at Evolutionary Genomics, Inc. (Longmoht, CO, USA). One microgram total RNA was run on a 0.75% agarose/TBE gel and visualized using Gel Red Nucleic Acid Stain (Biotium, Hayward, CA). Samples which had distinct ribosomal RNA bands and minimal smearing were deemed suitable for use as template for cDNA synthesis. cDNA libraries were constructed and subsequent transcriptome sequencing from the cDNA libraries was performed under contract by SeqWright DNA Technology Services (Houston, TX) using the Roche 454 GS-FLX Titanium platform. cDNA synthesis and Sanger sequencing of candidate genes were conducted at Evolutionary Genomics, Inc. (Longmoht, CO). cDNA was synthesized from 1µg of total RNA using the 1st Strand cDNA Synthesis Kit (Epicentre, Madison, WI). 3'-RACE CDS Primer A (Clontech, Mountain View, CA) was used as the poly dT primer for the reaction. Primer sequences were designed against homologous regions of putative target genes with Tms between 57 and 60°C using the OligoAnalyzer (IDT, Coralville, IA). Primers were purchased from Integrated DNA Technologies (IDT, Iowa-USA). PCR reactions were performed in 25 µL reactions containing 60 mM Tris-SO4 (pH 8.9), 18 mM Ammonium Sulfate, 2.0 mM Magnesium Sulfate, 0.2 mM each dNTP, 0.4 µM each Forward and Reverse primer, 0.5 Units Platinum® *Taq* Hi Fidelity (Invitrogen, Carlsbad, CA) in a Applied Biosystems Veriti thermocycler (Applied Biosystems, Carlsbad, CA). General PCR conditions were 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 57 - 60°C (depending on primer Tm) for 30 s, 68°C for 1 min, and a final 68°C extension step for 10 min. PCR products were then loaded onto 0.75 - 1.5% agarose gels and visualized using Gel Red Nucleic Acid Stain (Biotium, Hayward, CA).

5 µL of the PCR products was prepared for sequencing by enzymatic treatment with 2 µL ExoSAP-It (Affymetrix, Santa Clara, CA). The reactions were incubated at 37°C for 15 min followed by 15 min at 80°C to inactivate the enzymes. 2 µL of the reaction mixture was used as template for labeling with BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA) as follows: 2 µL of the template and 2 µL of 0.8 µM sequencing primer was added to a mixture of BigDye® Terminator v3.1 Ready Reaction Mix and BigDye® Terminator Sequencing Buffer in a 10 µL reaction. The reaction conditions were 96°C for 1 min followed by 25 cycles of 96°C for 10 s, 50°C for 5 s, 60°C 75 s.

Table 1. Germplasm from which RNA was extracted from leaf tissue for RNA isolation and sequencing.

Variety	Pedigree	Country of origin	Response to CBSD	Response to CMD	Response to Whitefly	Notes
Namikonga	Third (?) backcross from *M. glaziovii* onto *M. esculenta*. Same as Kaleso.	Tanzania	Resistant	Susceptible	Unknown	Very late root bulking. Poor yield. Parent of mapping population from which QTLs for CBSD resistance were derived. Genetic variation within this variety.
Mkombozi	A half sib of 92/0099S2(SM). 92/0099 is a cross 91934 X 81/00032 where 91934 is 58308 x Ogunjobi. 81/00032 is a cross between U/1421 and P-2. No further information is available on U/1421 or P-2.	Tanzania but originally from IITA, Nigeria	Initially was thought to be tolerant but recently has succumbed to CBSD.	Resistant with CMD2 gene	Unknown	Used as parent in mapping population to generate QTLs for CBSD resistance. Two years before QTL available.
AR40-6	12.5% from wild species *M. esculenta* ssp. *flabellifolia* and 50% from CMD resistant variety C39.	CIAT, Colombia	Resistant/ tolerant	Resistant with CMD2 gene	Unknown	Used as parent in mapping population to generate QTLs for CBSD. Two years before QTL available.
Nachinyaya	Unknown	Tanzania	Resistant/ tolerant	Previously tolerant but seems to be succumbing	Unknown	Used as parent in mapping population to generate QTLs for CBSD. Two years before QTL available. has M. glaziovii characteristics.
NDL06/132	Self of variety Naliendele (NDL 90/34) which is half sib of Kibaha. Kibaha has *M. flabellifolia* background.	Tanzania	Tolerant	Resistant	Unknown	Used as parent in mapping population to generate QTLs for CBSD. Two years before QTL available. Breeding line selected at ARI Naliendele in southern Tanzania.
Kiroba	Unknown	Tanzania	Tolerant	Susceptible	Unknown	Used as parent in mapping population to generate QTLs for CBSD. Two years before QTL available. has *M. glaziovii* characteristics.
Muzege	Unknown	Tanzania	Resistant/tol erant	Resistant	Unknown	Local farmer variety popular in the Kibaha district of Tanzania.
AR37-80		CIAT, Colombia	Susceptible	Resistant with CMD2 gene	Unknown	Susceptible parent of mapping population.

Table 1. Contd.

Variety	Pedigree	Country of origin	Response to CBSD	Response to CMD	Response to Whitefly	Notes
TMS4 (2)1425	58308 x Oyarugba Funfun		Susceptible	Resistant with polygenic source of resistance	Unknown	Susceptible parent of CBSD mapping population.
Albert	Unknown	Tanzania	Susceptible	Resistant	Unknown	Normal to early root bulking. Reasonable yield. Susceptible parent from which CBSD resistant QTLs currently available. Local farmer variety.
M. glaziovii	Wild species	Sirawi, Tanzania	Unknown	CMD evident in many plants, but not on the plant from which leaves were collected for RNA extraction.	Unknown	Does not produce storage roots. Imported by Germans to East Africa for rubber production. Now wild along coast of Tanzania.
Tree Cassava 1	Appears to be an interspecific hybrid between *M. glaziovii* and *M. esculenta*.	SRI, Kibaha, Tanzania	Unknown	Unknown	Unknown	Grown around homestead. Leaves are used as vegetable.
Tree Cassava 2	Appears to be an interspecific hybrid between *M. glaziovii* and *M. esculenta*.	SRI, Kibaha, Tanzania	Unknown	Unknown	Unknown	Grown around homestead. Leaves are used as vegetable.
Homestead cassava 1	Appears to be an interspecific hybrid between *M. glaziovii* and *M. esculenta*.	Tanzania	Unknown	Unknown	Unknown	Grown around homestead. Leaves are used as vegetable.
Homestead cassava 2	Appears to be an interspecific hybrid between *M. glaziovii* and *M. esculenta*.	Tanzania	Unknown	Unknown	Unknown	Grown around homestead. Leaves are used as vegetable. Has tough fibrous storage root. Not starchy.
CMC40-2	*M. esculenta* landrace	Colombia	Unknown	Unknown	Susceptible	
FLA21-5	*M. esculenta* ssp. *flabellifolia*	Brazil	Unknown	Unknown	Resistant	This line has also been reported to be resistant to casaba bacterial blight (CBB).
ECU72-2	*M. esculenta* landrace	Ecuador	Unknown	Unknown	Resistant	Parental from CIAT's whitefly segregating population.
MCOL 2246-2	*M. esculenta* landrace	Colombia	Unknown	Unknown	Susceptible	Parental from CIAT's whitefly segregating population.
MPER417-3-1	*M. esculenta* ssp. *peruviana*	Peru	Unknown	Unknown	Resistant	

Unincorporated BigDye® terminators were removed using the BigDye® XTerminator Purification Kit (Applied Biosystems, Carlsbad, CA). The reactions were sequenced on the Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems, Carlsbad, CA).

Cloning of putatively positively selected candidate genes

PCR fragments were cloned using TOPO® TA Cloning® Kit for Sequencing (Invitrogen, Carlsbad, CA) using 2 µL of PCR product according to the manufacturer's protocol. The ligated vector was transformed into Top10 competent cells (Invitrogen, Carlsbad, CA) using the Chemical Transformation protocol. The transformed *Escherichia coli* cells were plated onto LB agar plates containing 30 µg/mL kanamycin and 40 µg/mL X-gal. The plates were cultured overnight at 37°C.

Colony PCR

Colonies containing recombinant plasmids were screened using blue and white color selection followed by PCR using M13 forward and reverse primers. PCR reactions were performed in 15 µL containing 60 mM Tris-SO4 (pH 8.9), 18 mM Ammonium Sulfate, 2.0 mM Magnesium Sulfate, 0.2 mM each dNTP, 0.2 µM each Forward and Reverse primer, 0.3 Units Platinum® *Taq* Hi Fidelity (Invitrogen, Carlsbad, CA) in a Applied Biosystems Veriti thermocycler (Applied Biosystems, Carlsbad, CA). White colonies were picked and inoculated into 50 µL of distilled water which was then heated to 95°C for 10 min. 2 µL of inoculate was used as template in the PCR reaction. The PCR conditions were 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 50°C for 30 s, 68°C for 1 min and a final 68°C extension step for 10 min. PCR products were then loaded onto 0.75 - 1.5% agarose gels (Figure 1) and visualized using Gel Red Nucleic Acid Stain (Biotium, Hayward, CA). Colony PCR reactions producing products of the correct size were sequenced as above.

Sequence alignment

Sequence files from the ABI 3130 Genetic Analyzer were imported to Sequencher v4.8 Build 3767 (Gene Codes, Ann Arbor, MI). Vector sequence was trimmed using the Trim Vector tool. Sequences were then automatically aligned and manually edited to identify and remove sequencing artifacts.

Data analysis

Molecular-level adaptive evolution is indicated when comparisons of homologous protein coding sequences from closely related species show that the number of amino acid differences fixed due to selection exceeds what can be expected by neutral evolution. Molecular-level positive selection can be detected in protein-coding genes by pairwise comparisons of the ratios of nonsynonymous nucleotide substitutions per nonsynonymous site (Ka) to synonymous substitutions per synonymous site (Ks) (Li et al., 1985; Li, 1993; Kreitman and Akashi, 1995). The algorithm, by comparing substitutions per site, takes into account in rigorous fashion the effect of bias and degeneracy in the genetic code, and also compensates for the effects of multiple hits at the same site. Ka/Ks ratios significantly greater than one strongly suggest that positive selection has fixed greater numbers of amino acid replacements than can be expected as a result of chance alone (Nei and Gojobori, 1986; Stewart, 1997). Detection of positively-selected DNA sequences using Evolutionary Genomics Inc.'s bioinformatics software pipeline was performed as described (Messier and Sikela, 2001). Briefly, the software prepares cDNA sequences generated by high-throughput sequencing systems and transfers them for BLAST alignment against a database of choice, followed by Ka/Ks analysis of pairs of aligned orthologous sequences. The vast majority of the orthologous sequence pairs have Ka/Ks ratios ≤ 1, indicating they are conserved, and as such are automatically archived, leaving a small number of sequence pairs with Ka/Ks ratios > 1. The few coding sequences with Ka/Ks ratios >1 are re-sequenced and further analyzed. This provides a powerful filter, so that research efforts can rapidly be focused on a small subset of genes enriched for those likely to be associated with the adapted trait of interest (Messier and Sikela, 2001).

RESULTS

Transcriptomes of 13 cassava varieties (Table 1) were sequenced using the Roche 454 platform. Full transcriptome data is available in GenBank (accessions SRX333038 and SRX333040 - SRX333051). Fifty-four potentially positively-selected candidate gene sequences were obtained. The majority of these sequences were eliminated as a result of targeted Sanger re-sequencing: nucleotide base-for-base accuracy is required for rigorous computation of Ka/Ks ratios. Re-sequencing was performed after cloning out putative positively selected candidate genes. Cloning was necessary given duplication of candidate loci. Remaining false positives were eliminated as a result of phylogenetic analysis that revealed some apparent elevated Ka/Ks ratios stemmed from comparisons of paralogous gene copies. Thus, only confirmed orthologous pair-wise comparisons were used to determine true Ka/Ks ratios. Confirmed gene candidates were analyzed for putative function through homology searches, as well as by examination of the nature of the pattern of amino acid replacements that occurred in the 18 different cassava varieties described in Table 1. Variation in the predicted proteins was examined in relation to known response to disease, and whether the sample was derived from a cassava cultivar producing storage roots, or a wild species. Three candidate genes were identified that analysis confirmed had undergone significant positive selection. These positively selected genes have been designated EG2771, EG964 and EG5651; our confirmation of positive selection in these genes allowed us to consider these genes as leads, or candidate genes for possible CBSD-resistance in cassava. Sequence data for the three candidate genes identified in this project as available in GenBank (accessions KF025572 - KF025645). We did targeted Sanger sequencing of EG2771, EG964 and EG5651 in an additional 5 cassava varieties that were not initially deep-sequenced (Table 1). We provide here predicted protein sequences (Table 2) for each of the three candidate genes from 18 different cassava varieties in order to permit a broad assessment of how these candidate genes and their encoded proteins vary between cassava species/varieties.

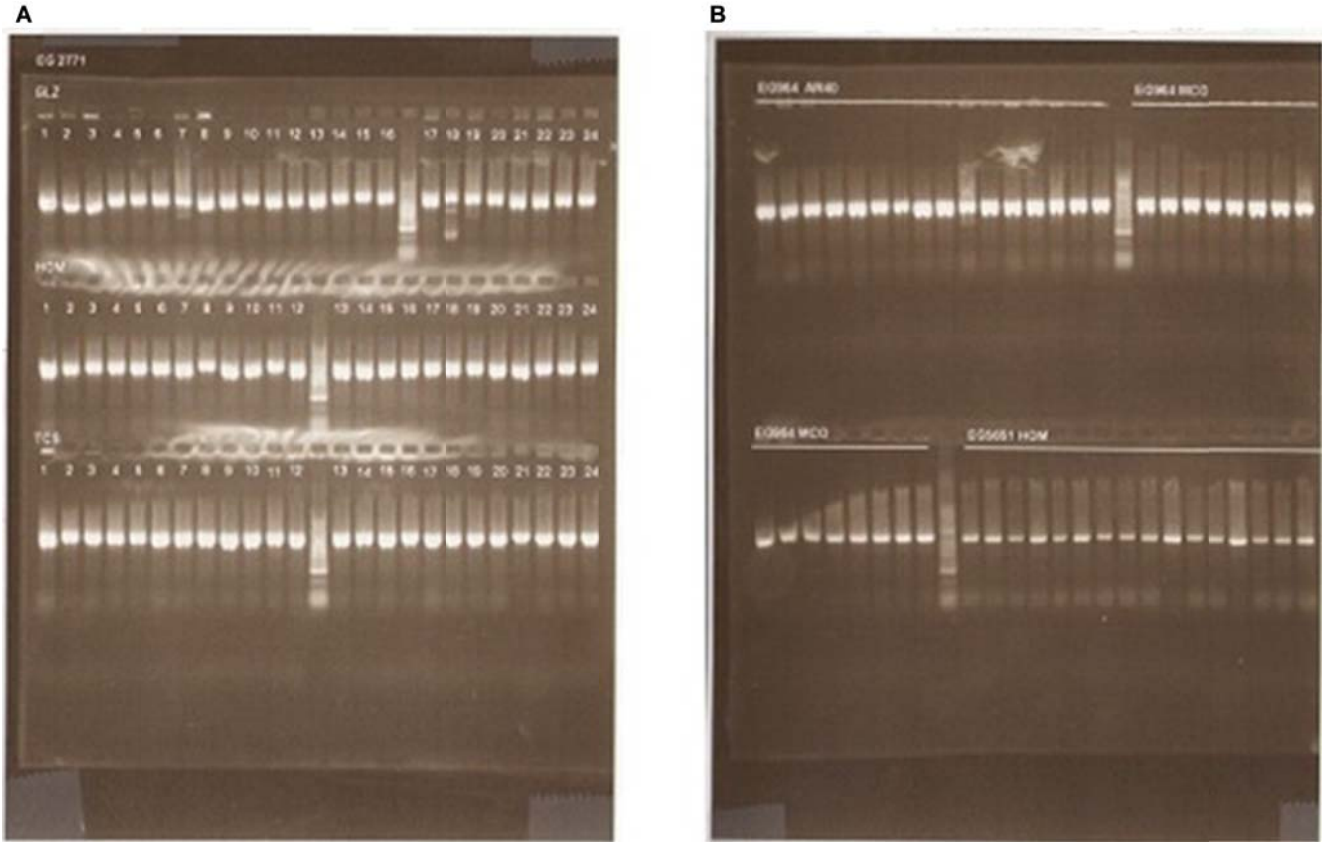

Figure 1. A. Colony PCR. Gene candidate EG2771 from *Manihot glaziovii* (GLZ), "homestead cassava 1" (HOM), and "tree cassava 1 (TCS). Table 1 show descriptions of these germplasm accessions. **B.** Gene candidate EG964 from *accessions* MCOL 2246-2 (MCOL), and AR40-6 (AR40) and gene candidate EG5651 from "homestead cassava 1" (HOM): Table 1 show descriptions of these germplasm accessions.

EG2771

This gene is positively selected (Ka/Ks =2.1), with strong statistical support. This comparison comprises the full-length coding region. The cassava sequence is homologous to genes that code for metal ion binding proteins, (for example, Gen Bank Acc. XM_002517346, from castor bean, *Ricinus communis*). The two alleles of EG2771 identified in Namikonga differ in length by 3 bp within the coding sequence that is, by the deletion of a single codon in allele A, as judged relative to all the other EG2771 accession coding sequences. As a result of shared alleles between 18 cassava varieties, the coding sequences for EG2771 give rise to 10 predicted protein sequences which are listed in relation to the response of each genotype to CBSD infection, response to whiteflies and storage root formation (Table 2). Predicted protein 3 occurs in 'wild' cassava genotypes, not classified as *M. esculenta*. Predicted proteins 4, 5, 9 and 10 occur exclusively in CBSD resistant/tolerant genotypes. The alleles coding for these proteins are thus candidates for CBSD resistance. Predicted protein 1 appears in both CBSD susceptible and tolerant genotypes, as well as non-*M. esculenta* genotypes, and is therefore probably not of particular interest in terms of CBSD resistance. Predicted protein 7 occurs in representatives of each CBSD response category, and predicted protein 8 occurs only in 'Mkombozi', previously thought to be resistant, but which has recently succumbed to CBSD; thus predicted proteins 1 and 7 are probably not of interest in terms of CBSD resistance. Predicted protein 6 only occurs in a genotype from Peru with resistance to whitefly, but unknown response to CBSD, so may be of interest in terms of whitefly resistance.

EG964

Full-length coding region of EG964 indicated that this gene was strongly positively selected with a Ka/Ks ratio of 1.44, with good statistical support. Homology searches suggest that EG964 codes for a member of the sterol-carrier protein-2 family. These lipid-transfer proteins influence inter-membrane lipid transfer (Zheng et al., 2008); they are believed to inhibit plant pathogens (Yeats and Rose, 2008; García-Olmedo et al., 1995). The coding

Table 2. Classification of predicted proteins of gene EG964 according to response to CBSD infection.

Predicted protein name	Predicted protein [b]	CBSD Resistant	CBSD Susceptible	Unknown response to CBSD	Whitefly Resistant	Whitefly Susceptible	Not classified as *Manihot esculenta*
EG964-1	MAADLKSDSIFDLMGRFLETDEG NELQKRINLIYQFNIAPKKIGIDEV SYTVNLKKGKVTKGPYEGGKPD ATFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPDIFP KPSKL	AR40-6 Nachinyaya Namikonga NDL06/132	Albert TMS4 (2)1425 AR-37-80 Mkombozi [a]	CMC40-2		CMC40-2	FLA21-5
EG964-2	MAADLKSDSIFDLMGRFLETDEG NELQKRINLIYQFNIAPKKIGIDEV SYTVNLKKGKVTKGPYEGGKPD ATFSIKDEDFVKLSEGKMNPQIA FMRGALKIKGSLSAAQKFTPDIF PKPSKL	Muzege	AR37-80 Mkombozi [a]	ECU72-2	ECU72-2		
EG964-3	MAADLKSDSIFDLMGRFLETDEG NELQKRINLIYQFNIAPKKIGIDEV SYTVNLKKGEVTKGPYEGGKPD ATFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPDIFP KPSKL						Homestead cassava [c]
EG964-4	MAADLKSDSIFDLMGRFLETDEG NELQKRINLIYQFNIAPKKIGIDEV SYTVNLKKGEVTKGPYEGGKPD ATFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPDIFP KPSKL	Kiroba Namikonga		MCOL 2246-2		MCOL 2246-2	
EG964-5	MATDLKSDSIFDLMDRFLQTDEG IELQKKINLIYQFNIAPKKIGIDEVS YTVNLKKGDVTKGPYEGGKPDA TFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPGIFP KPSKL						*M. glaziovii* [c]
EG964-6	MATDLKSDSIFDLMDRFLQTDEG IELQKKINLIYQFNIAPKKIGIDEVS YTVNLKKGEVTKGPYEGGKPDA TFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPDIFP KPSKL						*M. glaziovii* [c] *Homestead cassava* [c] *Tree cassava1* [c]
EG964-7	MAADLKPDSIFDLMGRFLETDEG NELQKRINLIYQFNIAPKKIGIDEV SYTVNLKKGEVTKGPYEGGKPD ATFSIKDEDFVKLSEGKLNPQIAF MRGALKIKGSLSAAQKFTPDIFP KPSKL			MPER417 -3-1	MPER417 -3-1		
EG2771-1	MATLSSLLISTPNSLTIPSHSLSLL LPNLHFQSHGVPYSFRHLTLKNS TKSKFPFSHVTKLRSVEEETQIP EEEEQQQAQEQEEEPVQEGPE QQTVSVPVSPSDILTMFFQAEGT MNETAIPTVTSALEETDGITNLKV QVLEGIASVELTKQTTVQATGVA SSLVELIQGSGFKLQTLNLSFMD EEDVLV	AR40-6 Nachinyaya	Albert AR37-80 Mkombozi [a] TMS 4(2)1425	CMC40-2 MCOL 2246-2	FLA21-5	CMC40-2 MCOL 2246-2	FLA21-5

Table 2. Contd.

Predicted protein name	Predicted protein b	CBSD Resistant	CBSD Susceptible	Unknown response to CBSD	Whitefly Resistant	Whitefly Susceptible	Not classified as Manihot esculenta
EG2771-8	MATLSSLLISTPNSLTIPSQSLSLL LPNLHFQSHGVPYSFRHLTLKNS TKSKFPFSHVTKLRSVEEETQIP EEEEQQQTQEQEQEPVQEGPE QQTVSVPVSPSDILTMFFQAEGT MNETAIPTVTSALEETDGITNLKV QVLEGIASVELTKQTTVQATGVA SSLVELIQGSGFKLQTLNLSFMD EEDVLV		Mkombozi [a]				
EG2771-9	MATLSSLLISTPNSLTIPSHSLSLL LPNLHFQSHGVPYSFRHLTLKNS TKSKFPFSHVTKLRSVEEETQIP EEEQQQAQEQEEEPVQEGPEQ QTVSVPVSPSDILTMFFQAEGTM NETAIPTVTSALEETDGITNLKVQ VLEGIASVELTKQTTVQATGVAS SLVELIQGSGFKLQTLNLSFMDE EDVLV	Namikonga					
EG2771-10	MATLSSLLISTPNSLTIPSQSLSLL LPNLHFQSHGVPYSFRHLTLKNS TKSKFPFSHVTKLRSVEEETQIP EEEEQQQAQEQEEEPVQEGPE QQTVSVPVSPSDILTMFFQAEGT MNETAIPTVTSALEETDGITNLKV QVLEGIASVELTKQTTVQATGVA SSLVELIQGSGFKLQTLNLSFMD EEDVLV	Namikonga NDL06/132					
EG5651-1	MDAKEAHSEVSEELTRELLIGISY LLPEKVQNSDVDEVLNAGKSVS RTNSDGADKYRSELISISYCSSP DMMPSPVIGKP						Homestead cassava2 [c]
EG5651-2	MDAKEVHSEVSEELTRETLIGIS YLLPEKVQNSDVDEILNAEKSVS RTNSDGADKYRSELISISYCPSP DMMPSPVIGKP	AR40-6 Nachinyaya Kiroba Muzege Namikonga NDL06/132	Albert AR37-80 Mkombozi [a] TMS 4(2)1425	CMC40-2 MCOL 2246-2 ECU72-2 MPER417 -3-1	ECU72-2 MPER417 -3-1 FLA21-5	CMC40-2 MCOL 2246-2	FLA21-5 Homestead cassava2 [c] Tree Cassava1 [c]
EG5651-3	MDAKEAHSEVSEELTRELLIGIS HLLPEKVQNSDVDEVLNAGKSV SRTNSDGADKYRSELISISYCSS PDMMPSPVIGKP						Tree Cassava1 [c]

[a]Initially was thought to be resistant but recently has succumbed to CBSD. [b]Amino acid residues that vary between the predicted proteins are underlined. [c]No true storage roots.

sequences for EG964 in the 18 cassava varieties give rise to seven predicted protein sequences (Table 2). Some accessions yielded two alleles, others only a single allele. Predicted protein sequences are shown in relation to the response of each genotype to CBSD infection and whitefly infestation (Table 2). Predicted proteins 3, 5, 6 and 7 occur exclusively in 'wild' cassava genotypes, not classified as *M. esculenta*. Predicted proteins 1 and 2 appear in both CBSD susceptible and resistant genotypes, as well as non *M. esculenta* genotypes, and therefore are probably not of particular interest in terms of CBSD resistance.

Predicted protein 4 exists in two CBSD resistant genotypes and one genotype of unknown response to CBSD. Predicted protein 7 is from a Peruvian genotype with resistance to whitefly, but unknown response to CBSD.

EG5651

The full-length coding region of gene EG5651 was strongly positively selected with a Ka/Ks ratio of 4.1, with good statistical support. Although, BLAST homology searches reveal some similarity (circa 72%) to protein sequences from several different plant species, none of these proteins has been characterized in any way, and most are termed "hypothetical", that is, no wet lab physical evidence supports these hypothetical proteins. The coding sequences of the 18 cassava varieties for EG5651 give rise to three predicted protein sequences, which are viewed in relation to response of each genotype to CBSD infection (Table 2). Most genotypes give rise to predicted protein 2; two non-*M. esculenta* varieties possess predicted proteins 1 and 3. This suggests that this gene may be associated with domestication.

DISCUSSION

This analysis identified three genes that have been subject to positive selection in cassava. These genes have been designated EG2771, EG964, and EG5651 with Ka/Ks ratios of 1.44, 2.1 and 4.1 respectively. Homology results identify them as, respectively, a metal ion binding protein (EG2771), a member of the sterol-carrier protein-2 family; these proteins influence inter-membrane lipid transfer and have been implicated in disease resistance (EG964), and a 'hypothetical' protein (EG5651). All three genes were identified from full-length sequences. It is extremely rare that full-length comparisons reveal the signature of past selective events. In the vast majority of cases full-length comparisons do not exhibit Ka/Ks ratios >1, as the accumulation of silent mutations swamps the signal of positive selection (Messier and Stewart, 1997). Comparisons between shorter regions of the protein coding sequence of a gene are far more likely to show Ka/Ks ratios >1. These shorter regions generally correspond to functional sub-regions or domains of a protein that have evolved in an adaptive, Darwinian fashion (Nei and Gojobori, 1986). Full-length comparisons show ratios >1 only when the selective pressure was especially strong. Such situations could have occurred during plant domestication or during the first contact between a naive plant population and a potent new pathogen/predator or when the selective event was relatively recent (such that synonymous mutations have not yet accumulated) (Nei and Gojobori, 1986; Messier and Stewart, 1997; Messier and Sikela, 2001). A very close phylogenetic relationship is evident between all the cassava varieties that were examined. For each of the three positively-selected genes, some alleles have been shared between varieties. This complicates Ka/Ks analysis in that minimal phylogenetic distance tends to mask positively selected gene pairs, however, when positively selected genes are found in such close comparisons, the likelihood is that the selective events were of unusually severe magnitude (Nei and Gojobori, 1986).

In EG964 predicted proteins 3, 5, 6 and 7 occur exclusively in 'wild' cassava genotypes, not classified as *M. esculenta*. We suspect that EG964 may have undergone selection during the domestication process. Predicted protein 4 exists in two CBSD resistant genotypes and one genotype of unknown response to CBSD. This protein may thus be of interest in terms of CBSD resistance. It will be valuable to screen a larger number of CBSD susceptible genotypes to determine whether this predicted protein occurs within them. Similarly, predicted protein 7 was found only in one whitefly resistant South American wild *Manihot* genotype (MPER 417-3-1) making this protein potentially important in terms of whitefly resistance. Four alleles of EG2771 (4, 5, 9, and 10) are found only in cassava varieties known to be CBSD resistant (Namikonga, Muzege, and NDL06/132) and are thus excellent candidates for CBSD resistance and should be further investigated/validated. The cassava variety called 'Namikonga' in Tanzania and 'Kaleso' in Kenya has shown to date some of the best resistance to CBSD. It was generated under the Amani breeding program and is thought to be a third backcross from *M. glaziovii* onto *M. esculenta* (Hillocks and Jennings, 2003). It has been grown under high disease pressure, and to date has only been found to show very mild symptoms even 24 months after planting and rarely shows leaf symptoms. A similar disease response is seen in Muzege (E. Kanju, per comm.). NDL06/132 is a self of variety 'Naliendele' (NDL 90/34) which is half sib of 'Kibaha' which has *M. flabellifolia* background. It rarely shows root symptoms even 24 months after planting, but can show moderate leaf symptoms. These three varieties have been grown under high disease pressure for as long as the disease has been recognized. The cassava brown streak viruses are low titer viruses and tend to build up in the plant over successive years of clonal propagation. Thus, a variety that initially appears resistant can suddenly succumb to the virus. 'Mkombozi', a variety from the Lake Zone of Tanzania is such an example (It is of course possible that the resistant line encountered a different CBSV species or strain variant. However, this phenomenon of a variety suddenly collapsing and succumbing to the disease has been observed many times in many locations). Allele 6 of EG2771 is found only in the wild genotype MPER 417-3-1, which shows the best whitefly (*A. socialis*) and green mite (*Mononychellus tanajoa*) resistance found within the wild *Manihot* germplasm screened to date (Burbano et al., 2007). Allele 6 is thus a candidate for whitefly resistance. Very little information is available about the protein coded by EG5651 in cassava or in any other plant species. This is not unexpected as Lopez et al. (2004) were only able to assign likely biochemical function to 37% of unigenes in

their study of 5,800 cassava unigenes. Furthermore, some 16% of their sequences showed no homology to known sequences and were thus likely to be cassava-specific.

Our analysis reveals positively selected genes which may have a variety of functions in addition to, and including, CBSD and whitefly resistance because the cassava varieties examined exhibit a range of phenotypic differences. However, these positively selected genes are a potential starting point to look for disease resis-tance genes, particularly those alleles of EG2771 that were found exclusively in CBSD and whitefly-resistant varie-ties. Our candidate genes require validation for any potential role in conferring resistance to CBSD and/or to whitefly. We plan to test these genes in segregating populations derived from the resistant lines used in our analysis. We are mapping populations with the varieties used in this study as parents these could be screened with markers designed from these candidate genes to look for co-location of QTL and candidate genes (MF is now engaged in the first year of phenotyping for QTL analysis). Several points must be considered when thinking about our results.

First, the cassava brown streak virus lacks a long shared evolutionary history with *Manihot* before its introduction by Europeans in recent centuries. Since, all the *Manihot* species in question originate in South/ Central America whereas the pathogen apparently evolved in East Africa, the likelihood that the selection pressure investigated was to a pathogen other than CBSV/Ug-CBSV must be considered (The same logic applying also to the whitefly vector encountered in East Africa). However, any challenge by pathogens other than CBSV could have provided selective pressure for a broad-spectrum response that is upregulated by subsequent challenges from novel pathogens. Ongoing work by WM in soybean (unpublished) has shown that some positively selected genes that we identified using our Ka/Ks approach provide broad-spectrum resistance to a number of pathogens, even including some novel pathogens to which the plants is naïve.

Second, because the CBSD-resistance we investigate here derives from wide crosses with distantly-related *Manihot* species, there is a likelihood of large introgression blocks that could not be broken by species backcrosses to cultivated cassava. We plan to address this, by anchoring our candidate genes onto the physical map of cassava (which in the latest version (5), of the genome is anchored into chromosomes). We can then see whether these do in fact come from one introgression block. Clearly, we cannot yet claim that our positively-selected cassava genes definitely provide CBSD resis-tance; this obviously awaits completion of subsequent validation efforts now underway. At present, we offer a set of positively-selected genes from cassava. This approach to candidate gene identification should be viewed as complementary to other approaches such as RNASeq and Serial Analysis of Gene-Expression (SAGE) (Velculescu et al., 1995).

In summary, we detected three genes with strong positive selection, although we cannot yet definitely ascribe a selection pressure that was responsible for the observed positive selection. This may have been disease pressure, domestication, or some other as yet unknown selective pressure.

Conflict of Interests

The author(s) have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors thank the Bill and Melinda Gates Foundation (Seattle, WA, USA) for funding this work. We are also indebted to Mr. Haidari Nawabu, IITA, for assisting in the collection of *M. glaziovii*, tree cassava and homestead cassava. The Kibaha Sugarcane Research Institute gave permission to collect 'tree cassava' leaves from a tree cassava plant maintained by the Institute for research purposes. We thank the Sugarcane Research Institute for permission to obtain samples, and for maintaining the tree cassava plants. Permission was obtained from local land owners when ‘homestead’ cassava was collected and when *M. glaziovii* was collected from the wild population at Sirawi in Tanzania.

REFERENCES

Allem AC (1994). The origin of *Manihot esculenta* Crantz (Euphobiaceae). Genet. Resour. Crop Evol. 41: 133-150.

Allem AC (2002). The origins and taxonomy of cassava. In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Cassava: Biology, Production and Utilization. CABI Publishing, Wallingford, UK, pp. 1-16.

Bellotti AC (2002). Arthropod pests. In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Cassava: Biology, Production and Utilization. CABI Publishing, Wallingford, UK, pp. 209-235.

Bellotti AC, Arias B (2001). Host plant resistance to whiteflies with emphasis on cassava as a case study. Crop Prot. 20: 813-823.

Bouvaine S, Gowda M (2014). Comparative transcriptome analysis of cassava plants differing in their resistance level to Cassava Brown Streak Virus. In: PAG XXII conference, Plant and Animal Genome, 2014.

Burbano MM, Carabali MA, Montoya LJ, Bellotti AC (2007). Resistance of *Manihot* species to *Mononychellus tanajoa* (Acariformes), *Aleurotrachelus socialis*, and *Phenacoccus herreni* (Hemiptera). Rev. Colomb. Entomol. 33(2): 110-115.

Byrne DN, Bellows TS (1991). Whitefly biology. Annu. Rev. Entomol. 36: 431-457.

Carabalí A, Bellotti AC, Montoya-Lerma J, Fregene M (2010a). *Manihot flabellifolia* Pohl, wild source of resistance to the whitefly *Aleurotrachelus socialis* Bondar (Hemiptera: Aleyrodidae). Crop Prot. 29: 34-38.

Carabalí A, Bellotti AC, Montoya-Lerma J, Fregene M (2010b). Resistance to the whitefly, *Aleurotrachelus socialis*, in wild populations of cassava, *Manihot tristis*. J. Insect Sci. 10(170): 1-10.

Cours G (1951). Le manioc à Madagascar. Mém. Inst. Sci. Madagascar. Sér B, Biol. Vég. 3: 203-400.

Cours G, Fargette D, Otim-Nape GW, Thresh JM (1997). The epidemic of cassava mosaic virus disease in Madagascar in the 1930s-1940s: lessons for the current situation in Uganda. Trop. Sci. 37: 238-248.

FAO (2013). Welcome to FAOSTAT.

http://faostat3.fao.org/home/index.html#DOWNLOAD. Accessed 28 March 2013.
Fauquet CM, Bisaro DM, Briddon R, Brown JK, Harrison BD, Rybicki EP, Stenger DC, Stanley J (2003). Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. Arch. Virol. 148:405-421.
García-Olmedo F, Molina A, Segura A, Moreno M (1995). The defensive role of nonspecific lipid-transfer proteins in plants. Trends Microbiol. 3(2):72-74.
Herrera Campo B, Hyman G, Bellotti AC (2011). Threats to cassava production: known and potential geographic distribution of four key biotic constraints. Food Sec. 3:329-345.
Hillocks RJ (2003). Cassava Brown Streak Virus Disease: Summary of Present Knowledge on Distribution, Spread, Effect on Yield and Methods of Control. In: Legg JP, Hillocks RJ (eds) Cassava Brown Streak Virus Disease: Past, Present and Future. Proceedings of an International Workshop, Mombasa, Kenya. 27-30 October 2002. Natural Resources International Limited, Aylesford, UK.
Hillocks RJ, Jennings DL (2003). Cassava brown streak disease: a review of present knowledge and research needs. Int. J. Pest Manage. 49:225-234.
Hillocks RJ, Raya MD, Thresh JM (1999). Factors affecting the distribution, spread and symptom expression of CBSD in Tanzania. AJRTC 3:57-61.
Hillocks RJ, Thresh JM, Bellotti AC (eds) (2002). Cassava: Biology, Production and Utilization. CABI Publishing, Wallingford, UK.
Jennings DL (1957). Further studies in breeding cassava for virus resistance. East Afr. Agric. J. 22: 213-219.
Jennings DL, Iglesias, C (2002). Breeding for crop improvement. In: Hillocks, RJ, Thresh, JM, Bellotti, AC (eds) (2002) Cassava: Biology, Production and Utilization. CABI Publishing, Wallingford, UK.
Jones WO (1959). Manioc in Africa. Standard University Press, Stanford, California.
Kreitman M, Akashi H (1995). Annu. Rev. Ecol. Syst. 26: 403-422.
Lebot V (2009). Tropical Root and Tuber Crops: cassava, sweet potato, yams and aroids. Crop production science in horticultural series. 17 MPG Books Group, UK.
Legg JP (2008). African cassava mosaic disease. In: Mahy BWJ, Van Regenmortel MHV (eds) Encyclopaedia of Virology. Elsevier, Oxford, UK, pp. 30-36.
Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G, Mkamilo G, Alicai T, Kumar, LP (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. Virus Res. 159:161-170.
Legg JP, Ogwal S (1998). Changes in the incidence of African cassava mosaic geminivirus and the abundance of its whitefly vector along south-north transects in Uganda. J. Appl. Entomol. 122:169-178.
Legg JP, 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.
Legg JP, Raya M (1998). Survey of cassava virus diseases in Tanzania. Int. J. Pest Manage. 44(1):17-23.
Li W-H (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36: 96-99.
Li W-H, Wu C-I, Luo C-C (1985). A new method for estimating synonymous and nonsynonymus rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2:150-174.
Lokko Y, Dixon AGO, Offei SK, Danquah EY (2006). Gene complementarity of resistance to the cassava mosaic disease among African cassava accessions. Afr. Crop Sci. J. 14:207-220.
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.
Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ (2014). Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS ONE 9(5): e96642. doi:10.1371/journal.pone.0096642.
Maruthi MN, Hillocks RJ, Mtunda K, Raya MD, Muhanna M, Kiozia H, Rekha AR, Colvin J, Thresh JM (2005). Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). J. Phytopathol. 153:307-312.
Mbanzibwa DR, Tian YP, Tugume AK, Mukasa SB, Tairo F, Kyamanywa S, Kullaya A, Valkonen JPT (2009). Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch. Virol. 154(2):353-359.
Messier W, Sikela JM (2001). Methods to identify evolutionarily significant changes in polynucleotide and polypeptide sequences in domesticated plants and animals. U.S. Patent No. 6,274,319.
Messier W, Stewart C-B (1997). Episodic adaptive evolution of primate lyozymes. Nature 385:151-154.
Monger WA, Alicai T, Ndunguru J, Kinyua ZM, Potts M, Reeder RH, Miano DW, Adams IP, Boonham N, Glover RH, Smith J (2010). The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. Arch. Virol. 155:429-433.
Monger WA, Seal S, Isaac AM, Foster GD (2001). Molecular characterization of Cassava brown streak virus coat protein. Plant Pathol. 50:527-534.
Nei M, Gojobori T (1986). Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.
Nichols RFW (1950). The Brown Streak Disease of Cassava: Distribution, Climatic Effects and Diagnostic Symptoms. East Afr. Agric. J. 15(3):154-160.
Oliveira M, Henneberry T, Anderson P (2001). History, current status, and collaborative research projects for *Bemisia tabaci*. Crop Prot. 20:709-723.
Olsen KM (2004). SNPs, SSRs and inferences on cassava origin. Plant Mol. Biol. 56:517-526.
Patil BL, Fauquet CM (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. Mol. Plant Pathol. 10(5): 685-701.
Thresh JM (2003). Brief History of Cassava Brown Streak Virus Disease. In: Legg JP, Hillocks RJ (eds) Cassava Brown Streak Virus Disease: Past, Present and Future. Proceedings of an International Workshop, Mombasa, Kenya. 27-30 October 2002. Natural Resources International Limited, Aylesford, UK.
Thresh JM, Cooter RJ (2005). Strategies for controlling cassava mosaic virus disease in Africa. Plant Pathol. 54:587-614.
Vargas O, Bellotti A (1983). Daño economico causado por moscas blancas en el cultivo de la yuca. In: Yuca: control integrado de plagas. Centro Internacional de Agricultura Tropical, Cali, Colombia pp. 129-144.
Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995). Serial analysis of gene expression. Science 270: 484-487.
Winter S, Koerbler M, Stein B, Pietruszka A, Paape M, Butgereitt A (2010). Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen. Virol. 91:1365-1372.
Yeats TH, Rose JK (2008). The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). Protein Sci. 17(2):191-198.
Zheng BS, Rönnberg E, Viitanen L, Salminen TA, Lundgren K, Moritz T, Edqvis J (2008). *Arabidopsis* sterol carrier protein-2 is required for normal development of seeds and seedlings. J Exp. Bot. 59(12):3485-3499.