Genetic diversity and SNP’s from the chloroplast coding regions of virus-infected cassava

Bruno Rossitto De Marchi Corresp., 1, Tonny Kinene 2, Renate Krause-Sakate 3, Laura M Boykin Corresp., 2, Joseph Ndunguru 4, Monica Kehoe 5, Elijah Ateka 6, Fred Tairo 4, Jamisse Amisse 7, Peter Sseruwagi 4

1 Gulf Coast Research and Education Center, University of Florida, Wimauma, Florida, United States
2 School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, Western Australia, Australia
3 Plant Protection, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil
4 Mikocheni Agricultural Research Institute, Dar es Salaam, Tanzania
5 Department of Primary Industries and Regional Development Diagnostic Laboratory Service, South Perth, Australia
6 Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
7 Mozambique Agricultural Research Institute, Nampula, Mozambique

Corresponding Authors: Bruno Rossitto De Marchi, Laura M Boykin
Email address: bruno.rossittode@ufl.edu, lboykin@mac.com

Cassava is a staple food crop in sub-Saharan Africa, it is a rich source of carbohydrates and proteins which currently supports livelihoods of more than 800 million people worldwide. However, its continued production is at stake due to vector-transmitted diseases such as Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD). Currently, the management and control of viral diseases in cassava relies mainly on virus-resistant cultivars of cassava. Thus, the discovery of new target genes for plant virus resistance is essential for the development of more cassava varieties by conventional breeding or genetic engineering. The chloroplast is a common target for plant viruses propagation and is also a potential source for discovering new resistant genes for plant breeding. Non-infected and infected cassava leaf samples were obtained from different locations of East Africa in Tanzania, Kenya, and Mozambique. RNA extraction followed by cDNA library preparation and Illumina sequencing was performed. Assembling and mapping of the reads were carried out and 33 partial chloroplast genomes were obtained. Bayesian phylogenetic analysis from 55 chloroplast protein-coding genes of a dataset with 39 taxa was performed and the SNPs for the chloroplast dataset were identified. Phylogenetic analysis revealed considerable genetic diversity present in chloroplast partial genome among cultivated cassava of East Africa. The results obtained may supplement data of previously selected resistant materials and aid breeding programs to find diversity and achieve resistance for new cassava varieties.
Genetic diversity and SNP’s from the chloroplast coding regions of virus-infected cassava

Brook Rossitto De Marchi1, Tonny Kinene2, Renate Krause-Sakate3, Laura M. Boykin2, Joseph Ndunguru4, Monica Kehoe5, Elijah Ateka6, Fred Tairo4, Jamisse Amisse7 and Peter Sseruwagi4

1 Gulf Coast Research and Education Center, University of Florida, Wimauma, FL 33598, USA
2 School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, WA 6009, Australia
3 Plant Protection, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil
4 Mikocheni Agricultural Research Institute, Sam Nujoma Road, Box 6226, Dar es Salaam, Tanzania
5 Department of Primary Industries and Regional Development Diagnostic Laboratory Service, South Perth, Australia
6 Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
7 Mozambique Agricultural Research Institute, Nampula, Mozambique

Corresponding author:
Laura Boykin
E-mail address: lboykin@mac.com

Keywords: Manihot esculenta, Cassava mosaic disease, Cassava brown streak disease, East Africa, Tanzania, Mozambique, Kenya

Abstract
Cassava is a staple food crop in sub-Saharan Africa, it is a rich source of carbohydrates and proteins which currently supports livelihoods of more than 800 million people worldwide. However, its continued production is at stake due to vector-transmitted diseases such as Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD). Currently, the management and control of viral diseases in cassava relies mainly on virus-resistant cultivars of cassava. Thus, the discovery of new target genes for plant virus resistance is essential for the development of more cassava varieties by conventional breeding or genetic engineering. The chloroplast is a common target for plant viruses propagation and is also a potential source for discovering new resistant genes for plant breeding. Non-infected and infected cassava leaf samples were obtained from different locations of East Africa in Tanzania, Kenya, and Mozambique. RNA extraction followed by cDNA library preparation and Illumina sequencing was performed. Assembling and mapping of the reads were carried out and 33 partial chloroplast genomes were obtained. Bayesian phylogenetic analysis from 55 chloroplast protein-coding genes of a dataset with 39 taxa
was performed and the SNPs for the chloroplast dataset were identified. Phylogenetic analysis revealed considerable genetic diversity present in chloroplast partial genome among cultivated cassava of East Africa. The results obtained may supplement data of previously selected resistant materials and aid breeding programs to find diversity and achieve resistance for new cassava varieties.

**Introduction**

Cassava, *Manihot esculenta* Crantz subsp. *esculenta* (Euphorbiaceae) is a perennial woody shrub also known as manioc or tapioca. Cassava is globally cultivated for its starchy storage roots, providing staple food for over 800 million people throughout Africa, Asia and South America (http://faostat.fao.org). Africa is currently the biggest cassava producer in the world (Hillocks and Thresh 2002). Cassava is highly resistant to drought and its roots can be maintained in the soil for a couple of years until harvesting, which makes this crop an important reserve of carbohydrates to relieve global famine (Raheem and Chukwuma 2001). The cassava propagation process from season to season is mostly done by the use of stem cuttings. The vegetative propagation exposes the crop to the proliferation of several pathogens, including virus diseases such as Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD). These viral diseases are threatening cassava’s continued production in all major growing areas of Africa. The production losses caused by CMD and CBSD together in Africa were reported to be more than US$3 billion every year (Hillocks and Jennings 2003; Legg et al. 2006). Cassava mosaic disease is caused by several different geminivirus species including *African cassava mosaic virus*, *East African cassava mosaic virus*, and *East African cassava mosaic virus* – (Uganda) (Legg et al. 2006). Cassava brown streak disease was primarily thought to be caused by two different Ipomovirus species, the *Cassava brown streak virus* (CBSV) and the *Ugandan cassava brown streak virus* (UCBSV). However, more recent studies have found that up to four different cassava brown streak virus (CBSVs) species might be associated with CBSD (Ndunguru et al. 2015). Cassava brown streak disease incidences are higher near to the coast of Tanzania and Mozambique (Hillocks and Jennings 2003). These viruses are transmitted by the whitefly, *Bemisia tabaci* (Gennadius) which is one of the most serious pests of agriculture worldwide (Lowe et al. 2000). In East Africa, high populations of whiteflies are visible in cassava crops and are causing tremendous yield losses (Hillocks and Jennings 2003). However, a recent study suggests that aphids may also play a role in the transmission of CBSV, due to the occurrence of a conserved DAG motif that is associated with the aphid vector (Ateka et al. 2017). Currently, the management and control of viral diseases in cassava relies mainly on tolerant cultivars. The main goal of the breeders is to obtain cultivars tolerant to both diseases. Although cassava with high tolerance for CMD has been reported, the sources of resistance to CBSD are very restricted (Tumwegamire et al. 2018). Thus, new target genes for plant virus resistance are essential for the development of virus-resistant cassava. The chloroplast components are a common target for plant virus propagation and can also play a role in plant defense against viruses (de Torres Zabala et al. 2015). One of the defence lines developed by plants against pathogens is based on plant resistance proteins which detect pathogens effectors activating an effector-triggered
immunity (ETI) system (Dangl et al., 2013, Stuart et al., 2013). A very well characterized ETI is the hypersensitive response (HR), a localized form of programmed cell death (PCD) in the surrounding area of the infection site that relies on a burst of reactive oxygen species (ROS) in chloroplasts (Zurbriggen et al., 2010). In addition, the chloroplast is associated with the synthesis of key plant defense mediators, such as the hormones salicylic acid (SA), abscisic acid (ABA) and jasmonic acid (JA) (Serrano et al., 2016; Nambara and Marion-Poll, 2005). There are several chloroplast factors reported to be interacting with virus nucleic acids or proteins (Zhao et al., 2016). Therefore, this organelle is a potential source for discovering new resistant genes for plant breeding. A series of typical changes followed by chlorotic symptoms imply the occurrence of chloroplast-virus interactions. (Zhao et al. 2016; Dardick 2007; Mochizuki et al. 2014). Many plant viruses have as a signature infection pattern the attachment to the chloroplast membrane, which affects large numbers of chloroplast and photosynthesis-related genes (Wei et al. 2010; Bhattacharyya and Chakraborty 2018). In other words, damage to the chloroplast is a fundamental step in successful infection for plant viruses. Certainly, several chloroplast factors have been identified to interact with viral components that could lead a virus-tolerance to the host plant (Qiao et al. 2009; Lim et al. 2010). These factors are involved in virus replication, movement, symptoms or plant defense, suggesting that viruses have evolved to interact with chloroplast (Zhao et al. 2016). Currently, more than 800 complete chloroplast genome sequences are available in the National Center for Biotechnology Information (NCBI) organelle genome database, including the complete cassava chloroplast genome. The cassava chloroplast genome is 161,453 bp in length and includes a pair of inverted repeats (IR) of 26,954 bp. It consists of 128 genes; 96 are single copy and 16 are duplicated in the IR. There are four rRNA genes and 30 distinct tRNAs, seven of which are duplicated in the IR (Daniell et al. 2008). Complete chloroplast genome sequences, as well as other regions of the chloroplast genome, have been proved to be useful tools for improving our understanding of the phylogenetic relationships between closely related taxa and the evolution of plant species (Daniell et al. 2016; Xu et al., 2017). In addition, they are used for the accurate identification of commercial cultivars and determination of their purity (Daniell et al. 2016). The goal of this study was to evaluate the genetic diversity of cassava chloroplast coding regions from plants naturally infected with viruses from different varieties collected in the field in East Africa. Other sequences from GenBank, including cassava wild relatives, were added to the phylogenetic analysis. The results will help to correctly identify varieties commonly used in East Africa and might be used as references to select materials and supplement data for breeding programs.

Methods

The dataset consisted of 33 cassava plants including Tanzanian, Kenyan and Mozambican samples. In addition, six chloroplast genomes from cassava and cassava relatives were retrieved from GenBank and added to the final dataset as references (EU117376, SRR2847471, SRR2847474, SRR2847450, SRR2847419 and SRR2847403) (Daniell et al. 2008; Wang et al. 2014). Further details about the samples used in the analysis are found in Table 1.
Field samples collection were performed differently for each batch of samples and the methods have been described in previous studies for Tanzanian (Ndunguru et al. 2015), Kenyan (Ateka et al. 2017) and Mozambican (Amisse et al. 2019) samples.

RNA extraction was performed from approximately 100mg of cassava leaf as described by the CTAB (cetyltrimethyl ammonium bromide) protocol (Lodhi et al. 1994). Total RNA quality check was performed followed by cDNA libraries preparation using the Illumina TruSeq Stranded Total RNA Sample Preparation kit (Illumina, San Diego, California). Paired end reads were generated using the Illumina MiSeq System. The methods used for Next Generation Sequencing were previously described (Ndunguru et al. 2015; Ateka et al. 2017).

The raw transcriptome data was trimmed and de novo assembled using the software CLC Genomics Workbench v9.5.2, as previously described (Ateka et al. 2017; Ndunguru et al. 2015). Assembled contigs were mapped to an existent cassava chloroplast genome from GenBank (EU117376) (Daniell et al. 2008) cultivar TME3 and merged to produce a single draft genome using Geneious v9.1.3 (Kearse et al. 2012). Assemblies were refined by repeatedly mapping trimmed raw reads to the draft sequencing and adjusting as necessary. Mapping was performed with the following settings in Geneious software; minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. A consensus between the mapped trimmed reads and the reference was used to form the new draft chloroplast genomes.

Draft chloroplast genomes were annotated using CpGAVAS (Liu et al. 2012). Subsequently, pairwise chloroplast alignment among the 39 cassava chloroplast genomes was performed using MAFFT v7.222 (Katoh et al. 2002) within the Geneious. Further annotations adjustment was performed by direct comparison with the reference genome (EU117376). Finally, the alignment was refined removing regions and genes with gaps or artifacts for further phylogeny reconstruction.

Phylogenetic Bayesian analyses were performed on DNA sequences of 55 protein-coding genes for a dataset with 39 taxa using ExaBayes version 1.4.1; (Aberer, Kobert, and Stamatakis 2014) and were run in parallel across 384 nodes on the Magnus supercomputer (located at the Pawsey Centre, Western Australia). Analyses were run for 1 million generations with sampling every 500 generations. Each analysis consisted of four independent runs, each utilizing four coupled Markov chains. The run convergence was monitored by finding the plateau in the likelihood scores (standard deviation of split frequencies < 0.0015). The first 25% of each run was discarded as burn-in for the estimation of a majority rule consensus topology and posterior probability for each node. Bayesian run files are available in Supplemental Files.

For virus identification on each sample, assembled contigs were sorted by length and the longest subjected to a BLAST search (blastn and blastx) (Altschul et al. 1990). In addition, reads were also imported into Geneious v9.1.3 (Kearse et al. 2012) and provided with reference sequences obtained from Genbank (NC012698 for CBSV, GQ329864 for CBSV-T and NC014791 for UCBSV). Mapping was performed with minimum overlap 10%, minimum overlap identity 80%, allow gaps 10% and fine tuning set to iterate up to 10 times.
Results

Illumina sequencing of libraries prepared from total DNA produced between 2,071,164 and 23,427,360 paired-end reads with a maximum sequence length of 100 nucleotides and minimum of 30 nucleotides for Kenyan samples, maximum sequence length of 300 nucleotides and minimum of 100 nucleotide sequence length for Tanzanian samples, maximum of 300 and minimum of 100 nucleotide sequence length for the Mozambican samples.

The phylogeny reconstruction consisted of an 35,439 bp alignment and included 55 protein-coding genes for a dataset with 39 taxa. The single-copy genes analyzed (53 genes) were psbA, atpA, atpF, atpH, atpI, rps2, rpoC1, psbM, psbD, psbC, psbZ, rps14, psaB, psaA, ycf3, rps4, ndhJ, ndhK, ndhC, atpE, atpB, rbcL, psaI, ycf4, cemA, petA, psbJ, psbL, psbF, psbE, petG, psaJ, rpl33, rpl20, clpP, psbB, psbT, psbN, psbH, petB, rps11, rpl36, rps8, rpl14, rpl16, rps3, rpl22, psaC, ndhE, ndhG, ndhI, ndhA and ndhH. The genes duplicated in the IR analyzed (2 genes) were rps7 and rpl23. All the GenBank Accessions from the nucleotide sequences obtained in this study are available at Table S1.

Single nucleotide polymorphisms (SNP’s) were evaluated by the number of differences between each of the samples, calculated by Geneious v9.1.3 (Fig. S1).

We carried out a polymorphism analysis focusing on the comparison between the cultivated cassava and the two Manihot glaziovii samples, which are cassava wild relative species. Manihot glaziovii (also named as India rubber tree species) was domesticated in South America and imported to East Africa in the early twentieth century. Since then, it has been used in cassava breeding programs to select the natural resistance of M. glaziovii to cassava pathogens, such as plant viruses (Nichols 1947; Bredeson et al. 2016).

It was found a total of 31 unique SNPs that were present only in the M. glaziovii chloroplast coding regions and absent in all the cultivated cassava samples.

Among all the coding-regions, the rbcL, which encodes the large subunits of RuBisCO, was the most variable, containing four unique SNP’s in the M. glaziovii samples. Followed by the psbC, clpP, petB and psbB with three unique SNP’s each; cemA with two SNP’s and an one nucleotide deletion; rps3, ycf4, rpl20 with two unique SNP’s each; atpB with one SNP and a six nucleotide deletion; and psaB, psbA, atpA, atpF, rpoC1, rps4, psbH, rps11, rps8, rpl22, rpl23, ndhH and rpl23 with only one SNP each.

The nucleotide phylogenetic analysis clustered samples in three main clades, herein named as clades A, B and C (Fig. 1).

Discussion

We obtained the draft chloroplast genomes from 33 cultivated cassava and wild relative species belonging to three different African countries. Most of the plants analyzed were naturally infected by the two main viral diseases affecting cassava production in Africa: CBSD and CMD. We found considerable genetic diversity in the chloroplast coding regions analyzed among cultivated cassava. Additionally, no direct relationship was observed between chloroplast coding regions and virus species, indicating the need of enhancing the diversity of germplasm available.

The polymorphism analysis revealed rbcL as the most variable chloroplast coding-region in the
wild relative species samples. The \textit{rbcL} encodes the large subunits of RuBisCO and have been
identified before to interact with the coat protein of the \textit{Potato virus Y} (PVY) (Feki et al., 2005) and it may be involved in the production of mosaic and chlorosis symptoms. In addition, \textit{rbcL} was reported to interact with the P3 protein of other Potyviruses such as \textit{Onion yellow dwarf virus} (OYDV), \textit{Soybean mosaic virus} (SMV), \textit{ Shallot yellow stripe virus} (SYSV) and \textit{Turnip mosaic virus} (TuMV) (Lin et al. 2011). It is possible that the potyvirus P3 protein affects the normal functions of RubisCO, contributing to symptom development (Lin et al. 2011). Curiously, Potyviruses and Ipomoviruses (causal agents of CBSD) are members of the same family, \textit{Potyviridae}, indicating that the \textit{rbcL} coding-region could potentially be a source of resistance or tolerance to viruses in the wild cassava \textit{M. glaziovii}.

The clade “A” of the phylogenetic tree consisted from the wild relative species \textit{M. glaziovii} from Tanzania and Mozambique. The clade “B” of the Phylogenetic Tree (Fig 1) was composed by two sub-clades. The upper sub-clade containing “Brazil FLA 444-1” which is a wild progenitor species of cassava, \textit{M. esculenta} ssp. \textit{flabellifolia} from Brazil; and the bottom sub-clade composed by the varieties “Namikonga”, “TME3” and three cassava plants from Kenya. The SNP’s analysis based on the chloroplast coding-regions revealed that “Namikonga”, “Kenya 02 Megana” and “Kenya 04 Megana” are 100% identical among them. The variety “Namikonga” is known to be CBSD-resistant but CMD-susceptible and has emerged by the introgression of the wild \textit{M. glaziovii} into \textit{M. esculenta}, (Bredeson et al. 2016). Curiously, both “Megana” cassava plants, which are 100% identical to Namikonga were naturally infected by CBSD, indicating that the single copy genes analysed may not be related to the CBSD resistance found in Namikonga. The clade “C” consisted mainly of cassava samples, excepted for “Tanzania Tree Cassava” (SRR2847471) which is supposed to be an \textit{M. glaziovii}–\textit{M. esculenta} hybrid (Bredeson et al. 2016). Other three samples (“Kenya 11 Kibandameno”, “Unknown Healthy” and “Australia TMS 150395”) showed 100% identity with “Tanzania Tree Cassava” on the SNP’s analysis (Fig. S1). It is known that DNA barcodes derived from the chloroplast genome can be used to identify varieties (Daniell et al. 2016). The 100% identity among chloroplast regions from these four cassava plants indicates that they are all related to \textit{M. glaziovii}–\textit{M. esculenta} hybrids. Curiously, “Kenya 11 Kibandameno” was naturally infected with EACMZV, EACMV-K, UCBSV and CBSV. It suggests that this variety, which is related to the \textit{M. glaziovii}–\textit{M. esculenta} hybrid, is not resistant to EACMZV, EACMV-K, UCBSV and CBSV isolates present in the field in Kenya.

The variety Kibandameno seems to either present high variability or to be misidentified in Kenya and Tanzania. The cassava plant identified as “Kenya 11 Kibandameno” was placed distantly from other plants also classified as Kibandameno and showed high numbers of SNP’s (46) compared with “Kenya 15 Kibandameno” (Fig. S1), suggesting that some varieties might be losing their purity. This also happened to the variety Megana in Kenya. The samples “Kenya02 Megana” and “Kenya 04 Megana” had no SNP’s between them. However, 20 SNP’s were present between “Kenya02 Megana” and “Kenya 05 Megana”.
In general, our phylogenetic analysis suggests that chloroplast coding-regions from East-African cassava varieties present considerably genetic diversity (Fig. 1). Cassava plants from Tanzania were more genetically similar to each other compared to the other countries. Kenya is the country where cassava presented the most genetic diversity, which increases the chances of finding natural resistance associate to plant pathogens. Additionally, the data also revealed that cassava plants collected in Kenya are genetically closer to cassava plants from Mozambique. Curiously, Kenya and Mozambique are geographically more distant from Tanzania.

**Conclusions**

Although a considerably diversity in the chloroplast of cultivated cassava has been found, there still a need for the introgression of new wild relative plants to increase the genetic diversity in East Africa and find new target genes resistant to CBSD and CMD. The polymorphism analysis revealed *rbcL* as the most variable chloroplast coding-region in the cassava wild relative species and could potentially be a source of resistance to pathogens. In addition, the data obtained in this study combined with more phenotypical data, such as resistance and cultivars traits of previously characterized genotypes, may help breeding programs to achieve resistance for new cassava varieties.

**Acknowledgements**

Computational resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia supported this work.

**References**

Aberer, Andre J., Kassian Kobert, and Alexandros Stamatakis. 2014. “ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era.” *Molecular Biology and Evolution* 31 (10): 2553–56. https://doi.org/10.1093/molbev/msu236.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amisse, J. J. G., J. Ndunguru, F. Tairo, E. Ateka, L. M. Boykin, M. A. Kehoe, N. Cossa, C. Rey, and P. Sseruwagi. 2019. “Analyses of Seven New Whole Genome Sequences of Cassava Brown Streak Viruses in Mozambique Reveals Two Distinct Clades: Evidence for New Species.” *Plant Pathology* 68 (5): 1007–18. https://doi.org/10.1111/ppa.13001.

Ateka, Elijah, Titus Alicai, Joseph Ndunguru, Fred Tairo, Peter Sseruwagi, Samuel Kiarie, Timothy Makori, Monica A. Kehoe, and Laura M. Boykin. 2017. “Unusual Occurrence of a DAG Motif in the Ipomovirus Cassava Brown Streak Virus and Implications for Its Vector Transmission.” Edited by Hanu R. Pappu. *PLOS ONE* 12 (11): e0187883. https://doi.org/10.1371/journal.pone.0187883.

Bhattacharyya, Dhriti, and Supriya Chakraborty. 2018. “Chloroplast: The Trojan Horse in Plant-Virus Interaction.” *Molecular Plant Pathology* 19 (2): 504–18. https://doi.org/10.1111/mpp.12533.

Bredeson, Jessen V, Jessica B Lyons, Simon E Prochnik, G Albert Wu, Cindy M Ha, Eric Edsinger-Gonzales, Jane Grimwood, Jeremy Schmutz, Ismail Y Rabbi, Chiedozie Egesi, Poasa Nauluvula, Vincent Lebot, Joseph Ndunguru, Geoffrey Mkamilo, Rebecca S Bart, Tim L Setter, Roslyn M Gleadow, Peter Kulakow, Morag E Ferguson, Steve Rounsley & Daniel S Rokhsar. 2016. “Sequencing Wild and Cultivated Cassava and Related Species...
Reveals Extensive Interspecific Hybridization and Genetic Diversity.” *Nature Biotechnology* 34 (5): 562–70. https://doi.org/10.1038/nbt.3535.

Daniell, Henry, Choun-Sea Lin, Ming Yu, and Wan-Jung Chang. 2016. “Chloroplast Genomes: Diversity, Evolution, and Applications in Genetic Engineering.” *Genome Biology* 17 (1): 134. https://doi.org/10.1186/s13059-016-1004-2.

Daniell, Henry, Kenneth J. Wurdack, Anderson Kanagaraj, Seung-Bum Bum Lee, Christopher Saski, and Robert K. Jansen. 2008. “The Complete Nucleotide Sequence of the Cassava (Manihot Esculenta) Chloroplast Genome and the Evolution of AtpF in Malpighiales: RNA Editing and Multiple Losses of a Group II Intron.” *Theoretical and Applied Genetics* 116 (5): 723–37. https://doi.org/10.1007/s00122-007-0706-y.

Dangl J L, Horvath D M, Staskawicz B J. 2013. Pivoting the plant immune system from dissection to deployment. *Science* 341, 746–751.

Dardick, Christopher. 2007. “Comparative Expression Profiling of Nicotiana Benthamiana Leaves Systemically Infected with Three Fruit Tree Viruses.” *Molecular Plant-Microbe Interactions* MPMI 20 (8): 1004–17. https://doi.org/10.1094/MPMI-20-8-1004.

Feki, S., Lotikili, M.J., Triki-Marrakchi, R., Karimova, G., Ounouna, H., Nato, A., Nato, F., Guesdon, J.L., Lafaye, P., Ben, A. and Elgaied, A. 2005 Interaction between tobacco Ribulose-1,5-bisphosphate Carboxylase/Oxygenase large sub- unit (RubisCO-LSU) and the PVY Coat Protein (PVY-CP). *European Journal of Plant Pathology* 112, 221–234.

Hillocks, R. J., and J. M. Thresh, eds. 2002. *Cassava: Biology, Production and Utilization*. Wallingford: CABI. https://doi.org/10.1079/9780851995243.0000.

Hillocks, RJ, and DL Jennings. 2003. “Cassava Brown Streak Disease: A Review of Present Knowledge and Research Needs.” *International Journal of Pest Management* 49 (3): 225–34. https://doi.org/10.1080/0967087031000101061.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66. http://www.ncbi.nlm.nih.gov/pubmed/12136088.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond. 2012. “Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data.” *Bioinformatics* 28 (12): 1647–49. https://doi.org/10.1093/bioinformatics/bts199.

Legg, J P, B Owor, P Sseruwangi, and J B T - Advances in Virus Research Ndunguru. 2006. “Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of A Regional Pandemic.” In *Plant Virus Epidemiology*, 67:355–418. Academic Press. https://doi.org/10.1016/S0065-3527(06)67010-3.

Lim, H.-S., A. M. Vaira, H. Bae, J. N. Bragg, S. E. Ruzin, G. R. Bauchan, M. M. Dienelt, R. A. Owens, and J. Hammond. 2010. “Mutation of a Chloroplast-Targeting Signal in Alternanthera Mosaic Virus TGB3 Impairs Cell-to-Cell Movement and Eliminates Long-Distance Virus Movement.” *Journal of General Virology* 91 (8): 2102–15. https://doi.org/10.1099/vir.0.019448-0.

Lin, L., Luo, Z., Yan, F., Lu, Y., Zheng, H. and Chen, J. 2011. Interaction between potyvirus P3 and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) of host plants. *Virus Genes*, 43, 90–92.

Liu, Chang, Linchun Shi, Yingjie Zhu, Haimei Chen, Jianhui Zhang, Xiaohan Lin, and Xiaojun Guan. 2012. “CpGAVAS, an Integrated Web Server for the Annotation, Visualization,
Analysis, and GenBank Submission of Completely Sequenced Chloroplast Genome Sequences.” *BMC Genomics* 13 (1): 715. https://doi.org/10.1186/1471-2164-13-715.

Lodhi, Muhammad A., Guang-Ning Ye, Norman F. Weeden, and Bruce I. Reisch. 1994. “A Simple and Efficient Method for DNA Extraction from Grapevine Cultivars AndVitis Species.” *Plant Molecular Biology Reporter* 12 (1): 6–13. https://doi.org/10.1007/BF02668658.

Lowe S., Browne M., Boudjelas S., De Poorter M. 2000. 100 of the World’s Worst Invasive Alien Species A Selection from the Global Invasive Species Database. Auckland, New Zealand: The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN). http://s1.downloadmienphi.net/file/downloadfile2/169/1398476.pdf.

Mochizuki, Tomofumi, Yoshiyuki Ogata, Yuki Hirata, and Satoshi T. Ohki. 2014. “Quantitative Transcriptional Changes Associated with Chlorosis Severity in Mosaic Leaves of Tobacco Plants Infected with Cucumber Mosaic Virus.” *Molecular Plant Pathology* 15 (3): 242–54. https://doi.org/10.1111/mpp.12081.

Nambara E, Marion-Poll A. 2005. Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology* 56, 165–185.

Ndunguru, Joseph, Peter Sseruwagi, Fred Tairo, Francesca Stomeo, Solomon Maina, Appolinaire Djinkeng, Monica Kehoe, Laura M. Boykin, and Ulrich Melcher. 2015. “Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation.” Edited by Ulrich Melcher. *PLoS ONE* 10 (10): 1–18. https://doi.org/10.1371/journal.pone.0139321.

Qiao, Y., H. F. Li, S. M. Wong, and Z. F. Fan. 2009. “Plastocyanin Transit Peptide Interacts with *Potato Virus X* Coat Protein, While Silencing of Plastocyanin Reduces Coat Protein Accumulation in Chloroplasts and Symptom Severity in Host Plants.” *Molecular Plant-Microbe Interactions* 22 (12): 1523–34. https://doi.org/10.1094/MPMI-22-12-1523.

Raheem, Dele, and Chrysanthus Chukwuma. 2001. “Foods from Cassava and Their Relevance to Nigeria and Other African Countries.” *Agriculture and Human Values* 18 (4): 383–90. https://doi.org/10.1023/A:1015233606665.

Serrano I, Audran C, Rivas S. 2016. Chloroplasts at work during plant innate immunity, *Journal of Experimental Botany* 67 (13): 3845–3854. https://doi.org/10.1093/jxb/erw088

Stuart LM, Paquette N, Boyer L. 2013. Effector-triggered versus pattern-triggered immunity: how animals sense pathogens. *Nature Reviews Immunology* 13, 199–206.

Torres Zabala, Marta de, George Littlejohn, Siddharth Jayaraman, David Studholme, Trevor Bailey, Tracy Lawson, Michael Tillich, Dirk Licht, Bettina Bölter, Laura Delfino, William Truman, John Mansfield, Nicholas Smirnoff & Murray Grant. 2015. “Chloroplasts Play a Central Role in Plant Defence and Are Targeted by Pathogen Effectors.” *Nature Plants* 1 (6): 15074. https://doi.org/10.1038/nplants.2015.74.

Tumwegamire, Silver, Edward Kanju, James Legg, Rudolph Shirima, Salehe Kombo, Geoffrey Mkamilo, Kiddo Mtunda, Karoline Sichalwe, Heneriko Kulembeka, Innocent Ndyetabura, Haji Saleh, Robert Kawuki, Titus Alicai, Gerald Adiga, Ibrahim Benesi, Albert Mhone, Anabela Zacarias, Sofrimento Fenias Matsimbe, Theresia Munga, Elijah Ateka, Lynet Navangi, Midatharahally Narasegowda Maruthi, Francis Mwatuni, George Ngundo, Maureen Mwangangi, Edward Mbugua, Joseph Ndunguru, Cyprian Rajabu, Deogratius Mark. 2018. “Exchanging and Managing In-Vitro Elite Germplasm to Combat Cassava
Brown Streak Disease (CBSD) and Cassava Mosaic Disease (CMD) in Eastern and Southern Africa.” *Food Security* 10 (2): 351–68. https://doi.org/10.1007/s12571-018-0779-2.

Wang, Wenquan, Binxiao Feng, Jingfa Xiao, Zhiqiang Xia, Xincheng Zhou, Pinghua Li, Weixiong Zhang, Ying Wang, Birger Lindberg Møller, Peng Zhang, Ming-Cheng Luo, Gong Xiao, Jingxing Liu, Jun Yang, Songbi Chen, Pablo D. Rabinowicz, Xin Chen, Hong-Bin Zhang, Henan Ceballos, Qunfeng Lou, Meiling Zou, Luiz J.C.B. Carvalho, Changying Zeng, Jing Xia, Shixiang Sun, Yuhua Fu, Haiyan Wang, Cheng Lu, Mengbin Ruan, Shuigeng Zhou, Zhicheng Wu, Hui Liu, Rubini Maya Kannangara, Kirsten Jørgensen, Rebecca Louise Neale, Maya Bonde, Nanna Heinz, Wenli Zhu, Shujuan Wang, Yang Zhang, Kun Pan, Mingfu Wen, Ping-An Ma, Zhengxu Li, Meizhen Hu, Wenbin Liao, Wenbin Hu, Shengkui Zhang, Jinli Pei, Anping Guo, Jianchun Guo, Jiaming Zhang, Zhongwen Zhang, Jianqiu Ye, Wenjun Ou, Yaqin Ma, Xinyue Liu, Luke J. Tallon, Kevin Galens, Sandra Ott, Jie Huang, Jingjing Xue, Feifei An, Qingqun Yao, Xiaojing Lu, Martin Fregene, L. Augusto Becerra López-Lavalle, Jiajie Wu, Frank M. You, Meili Chen, Songnian Hu, Guojian Wu, Silin Zhong, Peng Ling, Yeyuan Chen, Qinghuang Wang, Guodao Liu, Bin Liu, Kaimian Li & Ming Peng. 2014. “Cassava Genome from a Wild Ancestor to Cultivated Varieties.” *Nature Communications* 5: 5110. https://doi.org/10.1038/ncomms6110.

Wei, T., T.-S. Huang, J. McNeil, J.-F. Laliberte, J. Hong, R. S. Nelson, and A. Wang. 2010. “Sequential Recruitment of the Endoplasmic Reticulum and Chloroplasts for Plant Potyvirus Replication.” *Journal of Virology* 84 (2): 799–809. https://doi.org/10.1128/JVI.01824-09.

Xu Chao, Wenpan Dong, Wenqing Li, Yizeng Lu, Xiaoman Xie, Xiaobai Jin, Jipu Shi, Kaibong He and Zhili Suo. 2017. Comparative Analysis of Six Lagerstroemia Complete Chloroplast Genomes. *Frontiers in Plant Science* 8:15. https://doi.org/10.3389/fpls.2017.00015

Zhao, Jinping, Xian Zhang, Yiguo Hong, and Yule Liu. 2016. “Chloroplast in Plant-Virus Interaction.” *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2016.01565.

Zurbriggen MD, Carrillo N, Hajirezaei MR. 2010. ROS signaling in the hypersensitive response: when, where and what for? *Plant Signaling and Behaviour* 5, 393–396.
Figure 1

Phylogenetic relationships among cassava chloroplast coding regions collected in East Africa.

The cassava plants were clustered in three main clades (A, B and C). The viruses detected on the samples are indicated in colourful letters inside grey squares: *Ugandan cassava brown streak virus* (UCBSV), *Cassava brown streak virus* (CBSV), *East African cassava mosaic virus* (EACMV-Ug) and *East African cassava mosaic Zanzibar virus* (EACMZV). Samples obtained from GenBank are highlighted in blue.
Table 1 (on next page)

Next generation sequencing data from cassava samples collected in East Africa and cassava samples retrieved from GenBank.
1 Table 1. Next-generation sequencing data from cassava samples collected in East Africa and cassava samples retrieved from GenBank.

| Country/Sample ID | Species/Cultivar name | District/Region | Virus species | No. of reads obtained | No. of reads after trimming | GenBank Accession | Reference | Additional Information |
|-------------------|------------------------|-----------------|---------------|-----------------------|-----------------------------|-------------------|-----------|------------------------|
| Tanzania 03 (Tz: Ser-6) | Rumara | N/A | CBSV | 4,988,502 | 3,936,531 | Table S1 | This study | N/A |
| Tanzania 04 (Tz: Tan 19-1) | Kibandameno | N/A | CBSV | 2,735,840 | 2,279,998 | Table S1 | This study | CMD and CBSD Susceptible |
| Tanzania 05 (Tz: Tan 23) | Mkunungu | N/A | UCBSV | 4,978,272 | 4,115,933 | Table S1 | This study | N/A |
| Tanzania 06 (Tz: Tan 26) | Kibandameno | N/A | CBSV | 2,732,618 | 2,249,619 | Table S1 | This study | CMD and CBSD Susceptible |
| Tanzania 07 | Unknown | N/A | Not detected | 3,500,480 | 3,031,625 | Table S1 | This study | N/A |
| Tanzania 08 | Unknown | N/A | Not detected | 5,719,724 | 4,387,729 | Table S1 | This study | N/A |
| Tanzania 09 (Tz: Nya-38) | Mfangawandu | N/A | CBSV | 2,560,544 | 2,126,102 | Table S1 | This study | N/A |
| Tanzania 10 (Tz MF-49) | Kibembe | N/A | CBSV | 2,088,040 | 1,752,853 | Table S1 | This study | N/A |
| Tanzania 11 (Tz: Maf-51) | Kibembe | N/A | UCBSV | 2,071,164 | 1,852,435 | Table S1 | This study | N/A |
| Tanzania 12 (Tz: Maf-58) | Mwarabu | N/A | UCBSV | 2,548,594 | 2,030,993 | Table S1 | This study | N/A |
| Kenya 01 | Local | Bumula (Western) | CBSV & UCBSV | 15,573,022 | 15,282,670 | Table S1 | This study | N/A |
| Kenya 02 | Magana | Bumula (Western) | CBSV-Tz | 18,381,178 | 17,703,620 | Table S1 | This study | N/A |
| Kenya 03 | Tereka | Bumula (Western) | N/A | 17,711,474 | 17,024,115 | Table S1 | This study | N/A |
| Kenya 04 | Magana | Teso (Western) | UCBSV | 22,509,994 | 21,799,542 | Table S1 | This study | N/A |
| Kenya 05 | Magana | Busia (Western) | UCBSV & CBSV | 22,191,388 | 21,848,109 | Table S1 | This study | N/A |
| Kenya 07 | Local | Bondo (Nyanza) | EACMV - Ug | 21,488,968 | 21,240,329 | Table S1 | This study | N/A |
| Kenya 10 | Local | Bondo (Nyanza) | EACMV - Ug | 23,427,360 | 23,129,332 | Table S1 | This study | N/A |
| Location       | Province/Region    | Municipality       | Virus combination                                      | Infected Count 1 | Infected Count 2 | Reference | Notes                           |
|----------------|--------------------|--------------------|--------------------------------------------------------|------------------|------------------|-----------|---------------------------------|
| Kenya 11       | Kibandameno        | Malindi (Coast)    | EACMZV, EACMV-Ke, CBSV, UCBSV                          | 19,856,994       | 19,163,125       | Table S1  | This study CMD and CBSD Susceptible |
| Kenya 12       | Local              | Malindi (Coast)    | UCBSV & EACMV-Ke                                      | 19,187,524       | 18,865,588       | Table S1  | This study N/A                   |
| Kenya 13       | Kibandameno        | Msambweni (Coast)  | UCBSV & CBSV - Kilifi (Kenya)                         | 21,512,336       | 21,261,827       | Table S1  | This study CMD and CBSD Susceptible |
| Kenya 14       | Kibandameno        | Msambweni (Coast)  | CBSV & EACMV-Ke, UCBSV                                | 22,057,126       | 21,755,991       | Table S1  | This study CMD and CBSD Susceptible |
| Kenya 15       | Kibandameno        | Msambweni (Coast)  | CBSV                                                   | 22,124,706       | 21,851,119       | Table S1  | This study CMD and CBSD Susceptible |
| Tanzania       | Unknown DRJL030    | N/A                | Healthy                                                | 7,329,128        | 6,584,936        | Table S1  | This study Tall cassava - Healthy |
| Mozambique     | Ezelamalithi       | Namapa             | CBSV                                                   | 2,460,222        | 2,406,924        | Table S1  | N/A                             |
| Mozambique     | Calamidade         | Namapa             | CBSV                                                   | 3,608,398        | 3,562,380        | Table S1  | N/A                             |
| Mozambique     | Calamidade         | Nampula            | CBSV                                                   | 1,541,978        | 1,519,424        | Table S1  | N/A                             |
| Mozambique     | Buana              | Gile District      | Not detected                                          | N/A              | N/A              | Table S1  | N/A                             |
| Mozambique     | Fernando Alto      | Mocloce District   | Not detected                                          | N/A              | N/A              | Table S1  | N/A                             |
| Mozambique     | Bwana              | Alto Mocloce       | CBSV                                                   | 2,591,794        | 2,563,714        | Table S1  | N/A                             |
| Mozambique     | Mulaleia Mocuba    | District           | Not detected                                          | N/A              | N/A              | Table S1  | N/A                             |
| Mozambique     | Cadri              | Mocuba             | CBSV                                                   | 2,872,090        | 2,826,912        | Table S1  | N/A                             |
| Mozambique     | Mulaleia           | Quelimane          | CBSV                                                   | 2,635,892        | 2,619,932        | Table S1  | N/A                             |
| Mozambique     | Manihot glaziovii  | Namapa             | CBSV                                                   | N/A              | N/A              | Table S1  | CBSD and CMD resistant           |
| Unknown_T ME3 | TME3               | N/A                | N/A                                                    | N/A              | N/A              | EU117376  | Derived from the original cassava germplasm |

Note: CMD and CBSD susceptibility status is based on the study by Daniell et al. (2008).
| Country  | Variety          | Origin       | Notes                                                                 |
|----------|------------------|--------------|----------------------------------------------------------------------|
| Tanzania | Tree Cassava     | N/A          | Presumed to be an *M. glaziovii*–*M. esculenta* hybrid               |
| Australia| TMS I50395/Unk   | N/A          | Arose by introgression of *M. glaziovii* into *M. esculenta*, derived from the Amani program. It’s CBSD-resistant but CMD-susceptible |
| Unknown  | Namikonga        | N/A          | M. esculenta ssp. *flabellifolia*. Brazilian accession of the wild progenitor species of cassava |
| Tanzania | *Manihot glaziovii* | N/A          | CBSD and CMD resistant                                               |
| Brazil   | FLA444-1         | N/A          |暹羅花天冬。Brazilian accession of the wild progenitor species of cassava |