

Accepted Manuscript

Title: Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* Crantz) infected with cassava brown streak viruses

Author: Emmanuel Ogwok Muhammad Ilyas Titus Alicai
Marie E.C. Rey Nigel J. Taylor



PII: S0168-1702(16)30055-7
DOI: <http://dx.doi.org/doi:10.1016/j.virusres.2016.01.015>
Reference: VIRUS 96814

To appear in: *Virus Research*

Received date: 5-10-2015
Revised date: 18-1-2016
Accepted date: 21-1-2016

Please cite this article as: Ogwok, Emmanuel, Ilyas, Muhammad, Alicai, Titus, Rey, Marie E.C., Taylor, Nigel J., Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* Crantz) infected with cassava brown streak viruses. *Virus Research* <http://dx.doi.org/10.1016/j.virusres.2016.01.015>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* Crantz) infected with cassava brown streak viruses

Emmanuel Ogwok^{a, b, c}, Muhammad Ilyas^b, Titus Alicai^a, Marie E. C. Rey^c, and Nigel J. Taylor^{b*}

^aNational Crops Resources Research Institute, Namulonge, P.O Box 7084, Kampala, Uganda;

^bInstitute for International Crop Improvement, Donald Danforth Plant Science Center, St. Louis, MO 63132, USA; ^cSchool of Molecular and Cell Biology, University of the Witwatersrand, Private Bag 3, P.O Wits 2050, Johannesburg, South Africa

*Corresponding Author: Nigel J. Taylor, Donald Danforth Plant Science Center, 975 N. Warson Road, St. Louis, MO 63132, USA; Tel: +1-314-587-1257; Fax: 314-587-1357; e-mail: ntaylor@danforthcenter.org

Highlights

- The 21-nt virus-derived small RNAs were predominant, followed by the 22-nt class
- Susceptible cassava genotypes accumulated higher CBSV- than UCBSV-derived small RNA
- Tolerant cassava genotype accumulated high CBSV- and low UCBSV-derived small RNAs
- AGO2, DCL2 and DCL4 were differentially regulated in CBSV/UCBSV-infected plants
- CBSV and UCBSV interact differently in the same host genetic background

Abstract

Infection of plant cells by viral pathogens triggers RNA silencing, an innate antiviral defense mechanism. In response to infection, small RNAs (sRNAs) are produced that associate with Argonaute (AGO)-containing silencing complexes which act to inactivate viral genomes by posttranscriptional gene silencing (PTGS). Deep sequencing was used to compare virus-derived small RNAs (vsRNAs) in cassava genotypes NASE 3, TME 204 and 60444 infected with the positive sense single-stranded RNA (+ssRNA) viruses *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), the causal agents of cassava brown streak disease (CBSD). An abundance of 21-24 nt vsRNAs was detected and mapped, covering the entire CBSV and UCBSV genomes. The 21 nt vsRNAs were most predominant, followed by the 22 nt class with a slight bias toward sense compared to antisense polarity, and a bias for adenine and uracil bases present at the 5'-terminus. Distribution and frequency of vsRNAs differed between cassava genotypes and viral genomes. In susceptible genotypes TME 204 and 60444, CBSV-derived sRNAs were seen in greater abundance than UCBSV-derived sRNAs. NASE 3,

known to be resistant to UCBSV, accumulated negligible UCBSV-derived sRNAs but high populations of CBSV-derived sRNAs. Transcript levels of cassava homologues of AGO2, DCL2 and DCL4, which are central to the gene-silencing complex, were found to be differentially regulated in CBSV- and UCBSV-infected plants across genotypes, suggesting these proteins play a role in antiviral defense. Irrespective of genotype or viral pathogen, maximum populations of vsRNAs mapped to the cytoplasmic inclusion, P1 and P3 protein-encoding regions. Our results indicate disparity between CBSV and UCBSV host-virus interaction mechanisms, and provide insight into the role of virus-induced gene silencing as a mechanism of resistance to CBSD.

Keywords: Deep sequencing, cassava brown streak disease, cassava, virus-derived small RNA

1.1 Introduction

Cassava brown streak disease (CBSD) has become an economically important disease of cassava in East and Central Africa, causing significant losses in both yield and quality of cassava storage roots (Legg et al., 2011; Maruthi et al., 2014; Patil et al., 2015). CBSD is caused by the two distinct, but closely related, positive sense single-stranded RNA (+ssRNA) virus species, *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV). Both CBSV and UCBSV belong to the family *Potyviridae*, genus *Ipomovirus* (Mbanzibwa et al., 2009b; Monger et al., 2001). The viral pathogens are transmitted by whiteflies, *Bemisia tabaci*, Gennadius (Maruthi et al., 2005), and mechanically through grafting (Wagaba et al., 2013; Yadav et al., 2011) and sap inoculation of herbaceous plant species (Lister, 1959; Ogwok et al., 2010). The viruses systemically infect cassava in the field (Kaweesi et al., 2014; Ogwok et al., 2015) causing similar CBSD symptoms on affected cassava plants. CBSV is more aggressive,

inducing more rapid and more severe CBSD symptoms than those seen on plants infected with UCBSV alone. Recent studies have shown some cassava varieties, for example NASE 3 (Ogwok et al., 2015) and Kaleso (Maruthi et al., 2014), to be highly resistant to infection by UCBSV (Kabanyolo isolate) but less so to CBSV. The difference in response of cassava varieties to infection with CBSV and UCBSV implies a difference in the mechanism of host-virus interaction between the two virus species.

Plant-infecting viruses are known to trigger RNA interference (RNAi) (Ding, 2010; Llave, 2010), an innate antiviral defense mechanism. This involves critical steps including production of double-stranded RNA (dsRNA), processing of the dsRNA into small interfering RNA (siRNA), followed by sequence-specific targeting and silencing of viral messenger RNA (mRNA) by siRNA-incorporated effector complexes (Seo et al., 2013). The siRNA molecules associate with distinct Argonaute-containing silencing complexes to target and mediate diverse silencing effects on both host and viral genomes (Ding and Voinnet, 2007; Llave, 2010; Xia et al., 2014). During virus-host interactions, virus-specific dsRNAs are generated in a variety of ways, such as activity of virus-encoded RNA polymerase during genomic replication and transcription, or imperfect folding of self-complementary single stranded viral genomic RNA (Ding and Voinnet, 2007). Host-encoded RNA-dependent RNA polymerase (RDR) plays a crucial role in the maintenance of virus-induced gene silencing by converting ssRNAs to dsRNAs for secondary siRNA synthesis (Molnár et al., 2005; Seo et al., 2013). The dsRNAs serve as substrates for Dicer-like (DCL) ribonucleases which cleave the dsRNAs to produce 21 to 24 nucleotide (nt) virus-derived small RNA (vsRNA) duplexes (Brodersen and Voinnet, 2006; Moissiard and Voinnet, 2006; Zamore et al., 2000). The vsRNAs are then loaded onto Argonaute (AGO) proteins, a component of RNA-induced silencing complexes (RISC), to facilitate

cleavage of complementary viral mRNA, consequently silencing the viral genome as a self-defense response of the plant (Ding and Voinnet, 2007; Smith et al., 2011).

Studies in *Arabidopsis thaliana* and other plant species have revealed that the core components involved in plant small RNA biogenesis and the silencing pathways are encoded by multi-protein families (Seo et al., 2013). These are diverse and exhibit functional redundancy (Llave, 2010; Silva et al., 2011). For instance, the *A. thaliana* and rice genomes each have four and eight DCLs, six and five RDRs, and ten and nineteen AGO proteins respectively, while *Populus* has five DCLs. *A. thaliana* possesses six silencing pathways (Brodersen and Voinnet, 2006; Margis et al., 2006). It has also been shown that in virus-infected *A. thaliana*, DCL4, DCL2 and DCL3 catalyze the formation of 21, 22 and 24 nt vsRNAs, respectively, to induce antiviral responses (Blevins et al., 2006; Ruiz-Ferrer and Voinnet, 2009). The DCL4 generated 21 nt and DCL2 generated 22 nt vsRNAs confer efficient antiviral defense in plants. Nevertheless, DCL4 and DCL2 exhibit functional redundancy or cooperative interaction since formation of the 22 nt vsRNAs mainly happens in the absence of DCL4 (Deleris et al., 2006; Donaire et al., 2008; Donaire et al., 2009; Ruiz-Ferrer and Voinnet, 2009). The 21 and 22 nt vsRNAs are the most predominant class amongst virus-infected host plants, with a few exceptions (Deleris et al., 2006; Ding, 2010; Garcia-Ruiz et al., 2010). Multiple AGO proteins are also involved in antiviral defense, including AGO2 and AGO5, which have been shown to bind *Cucumber mosaic virus* (CMV)-derived small RNAs (Takeda et al., 2008). Similarly, RDR1, RDR2 and RDR6 have been implicated in vsRNA biogenesis and antiviral defense in plants (Ding, 2010; Ding and Voinnet, 2007; Donaire et al., 2008; Donaire et al., 2009; Garcia-Ruiz et al., 2010; Qi et al., 2009).

The biogenesis, composition, and abundance of vsRNAs have been characterized from several host-virus pathosystems (Donaire et al., 2008; Donaire et al., 2009; Lin et al., 2010; Miozzi et al., 2013; Naveed et al., 2014; Prabha et al., 2013; Silva et al., 2011; Visser et al., 2014; Xia et al., 2014). Small RNAs are known to play regulatory roles in defense responses against pathogens in plants (Ruiz-Ferrer and Voinnet, 2009; Seo et al., 2013). Different host plants have been shown to accumulate varying amounts of vsRNAs in response to infection by different viruses (Donaire et al., 2009; Lin et al., 2010). Through species diversity and epidemiology studies, CBSV and UCBSV have been shown to be closely related (Mbanzibwa et al., 2011; Mbanzibwa et al., 2009b). This has stimulated efforts to develop successful and durable CBSD resistance by simultaneous RNAi targeting of both viruses (Ogwok et al., 2012; Patil et al., 2011). However, no information is available concerning the biogenesis and composition of vsRNAs in CBSV- and UCBSV-infected cassava, nor their important, potential contribution to host plant resistance or susceptibility. The molecular mechanisms linking CBSV- and UCBSV-derived small RNAs to RNAi and CBSD symptom expression on different cassava genotypes is unclear. Alongside the premise that CBSV and UCBSV interact differently in cassava plants of the same genetic background (Ogwok et al., 2015), examining the CBSV- and UCBSV-cassava pathosystem offers a good platform to study molecular interactions between closely related RNA viruses in their natural host.

The present study aimed to characterize vsRNAs generated during separate CBSV and UCBSV infection in three cassava genotypes; NASE 3 (highly resistant to UCBSV but highly susceptible to CBSV), TME 204 (susceptible to both CBSV and UCBSV), and cultivar 60444 (highly susceptible to both CBSV and UCBSV). NASE 3 and TME 204 are farmer-preferred cassava varieties in Uganda (Taylor et al., 2012b), while 60444 is a model cassava genotype with well

established tissue culture and transformation protocols, used to study gene expression including virus-derived transgenes tailored for resistance to CBSD viral pathogens (Bull et al., 2009; Vanderschuren et al., 2012; Yadav et al., 2011). The information generated is an important contribution for understanding the molecular mechanisms underlying the pathogenicity of CBSV and UCBSV and the response of some cassava genotypes to CBSD.

1.2 Materials and methods

1.2.1 Plant materials and virus inoculations

Micropropagated, tissue culture-derived cassava plants of NASE 3, TME 204 and 60444 (N = 20) were confirmed to be free of CBSD viral pathogens by reverse transcription polymerase chain reaction (RT-PCR) (Ogwok et al., 2012) and established in soilless compost as previously described (Taylor et al., 2012a). For each genotype, eight plants were bud-graft challenged at 10 weeks after planting (WAP) with CBSV Naliendele isolate (CBSV-[TZ:Nal3-1:07]) or UCBSV Kabanyoro isolate (UCBSV-[UG:T04-42:04]) as described by Wagaba et al. (2013). The grafted plants were assessed a week later for graft union formation and monitored visually thereafter for CBSD symptom development. CBSD symptom severity on grafted plants was scored using a scale of 1–5 as previously described (Yadav et al., 2011; Ogwok et al., 2012).

1.2.2 Sample collection, RNA extraction and detection of CBSV and UCBSV

Systemically infected leaves of cultivar 60444 and TME 204 showing clear CBSD symptoms, and stem sections of NASE 3 showing lesions typical of CBSD infection were collected six weeks after grafting (WAG). Similar tissues were sampled for use as controls from asymptomatic, non-inoculated plants of each genotype. Presence of CBSV and UCBSV was

detected by RT-PCR and virus levels estimated by quantitative RT-PCR (qRT-PCR) as described previously (Ogwok et al., 2015). The plant materials were obtained from three independent biological replicates, wrapped with aluminum foil, immediately frozen in liquid nitrogen and stored at -80°C . Total RNA was extracted from the resulting powder as previously described (Ogwok et al., 2012). The RNA was treated with DNase I (Ambion Inc. Austin, TX) according to manufacturer recommendations to eliminate genomic DNA contamination. RNA was quantified using a Nanodrop (Model 2000C, Thermo Scientific, Waltham, MA). Integrity of RNA with $\text{OD}_{260/280}$ ratios from 1.8 to 2.1 was visually checked on 1.5% (w/v) agarose gel. Reverse transcription of RNA to cDNA was performed from 2 μg of total RNA using Superscript IIITM First-Strand Synthesis System (Invitrogen, Carlsbad, CA) following manufacturer recommendations. An aliquot of the RNA was used to prepare small RNA libraries.

1.2.3 Small RNA library preparation

Small RNA libraries were prepared using the NEBNext[®] Multiplex Small RNA Library Prep Set for Illumina[®] (Sets 1 and 2) (New England Biolabs, Inc. Ipswich, MA) following manufacturer recommendations. Briefly, 3' adapters were ligated to total RNA (6 μg) followed by hybridization of the reverse transcriptase primer to prevent adapter-dimer formation. Subsequently, 5' adapters were ligated and the RNA subjected to reverse transcription and PCR amplification. The PCR products were purified using a QIAQuick PCR Purification Kit (Qiagen Inc., Valencia, CA) according to manufacturer recommendations and quality-checked by electrophoresis on a 6% polyacrylamide gel. Bands of ~ 140 bp corresponding to the adapter-

ligated small RNA fragments were isolated, crushed in elution buffer and re-precipitated using a mixture of Acrylamide, sodium acetate, and absolute ethanol. The small RNA libraries (cDNA pellet) were re-suspended in TE buffer (10 mM Tris-HCl containing 1 mM EDTA, pH 8) and re-checked for size and concentration by electrophoresis in 6% polyacrylamide gel.

1.2.4 Sequencing and bioinformatics analysis of small RNA sequences

The small RNA libraries were sent to the Genome Technology Access Center (GTAC), Washington University in St. Louis, Missouri, USA for further size, quality, and integrity checking by an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc. Santa Clara, CA). Sequencing was performed by Illumina HiSeq 2,500 using 1x50 single-end read protocol. Raw sequence data received from GTAC was de-multiplexed by QIIME (Caporaso et al., 2010). Sequence reads with quality score below 19 were discarded. The bioinformatics tool Cutadapt (Martin, 2011) was used to remove adapter sequences. Small RNA sequences in the size range of 21 to 24 nt were selected for downstream analysis. The redundant as well as unique sRNA reads were mapped to the CBSV and UCBSV reference genome using Bowtie software (Langmead, 2010). Mapped reads data were converted to statistical data by BEDTools (Quinlan and Hall, 2010) and all outputs were graphically presented by Shell scripts provided by the Bioinformatics Core Facility at Donald Danforth Plant Science Center.

1.2.5 Quantification of DCL and AGO2 proteins by qRT-PCR

Quantitative RT-PCR (qRT-PCR) was performed to measure the expression levels of DCLs and AGO2 in CBSV- and UCBSV-infected cassava using cytochrome c oxidase (COX) mRNA as the reference gene for normalization as previously described (Ogwok et al., 2015). The primers

used to amplify the DCLs and AGO2 are listed in Table 1. Briefly, cDNA was diluted ten-fold and subjected to qRT-PCR using BioRad CFX96 Connect instrument (BioRad Laboratories Inc, Hercules, CA). Reaction mixtures contained 5 μ l SsoFast Advanced SYBR Green I SuperMix, 1 μ l of each primer (0.5 μ M final concentration) and 3 μ l of diluted cDNA template in a total reaction volume of 10 μ l. The qRT-PCR thermal cycles used were as follows: initial denaturation step for 3 min at 95°C, followed by 40 cycles of 10 s at 95°C and 30 s at 58°C. Data were collected at the end of each 58°C cycle. Each qRT-PCR run included in each plate RNA that went through cDNA synthesis process with no reverse transcriptase enzyme added (NRT) and a no-template control (NTC) containing water instead of cDNA. The qRT-PCR was performed using three biological replicates for each sample and three technical replicates of each biological replicate. The mean quantification cycle (C_q) value of each triplicate reaction was used for further calculations by the $2^{-\Delta\Delta CT}$ method for relative normalized expression analysis using COX mRNA as reference gene, and CBSV or UCBSV positive samples of predetermined virus concentration as calibrator, respectively. The relative expression levels were obtained by comparing CBSV and UCBSV titers in virus-infected plants with uninfected plants.

1.3 Results

1.3.1 Response of cassava genotypes to infection by cassava brown streak viruses

The three cassava genotypes responded differently to graft-challenge with CBSV and UCBSV. All plants (N = 8) of genotypes TME 204 and 60444 grafted with CBSV-infected scions started showing typical CBSD symptoms two weeks after grafting, a week earlier than plants grafted

with UCBSV-infected scions (Wagaba et al., 2013). NASE 3 plants grafted with UCBSV-infected buds obtained from either cultivars TME 204 or 60444 developed no symptoms, while plants of this cultivar grafted with CBSV-infected buds obtained from the same cultivars showed stem lesions around the graft union four WAG. Across all three genotypes, plants grafted with CBSV-infected buds showed more severe foliar CBSD symptoms compared to plants grafted with UCBSV-infected buds. CBSV-challenged plants of 60444, TME 204 and NASE 3 showed maximum mean severities of 4.3, 3.3 and 5.0, respectively, whereas UCBSV-challenged plants showed lower CBSD severities reaching maximum values of 3.0 and 2.0 for plants of 60444 and TME 204, respectively on a scale of 1-5 as previously described (Ogwok et al., 2012). As stated above, NASE 3 remained CBSD symptom-free after challenge with UCBSV (Table 2).

RT-PCR analysis revealed presence of CBSV in all CBSV-challenged plants showing CBSD symptoms across the three genotypes. qRT-PCR analysis showed that NASE 3 accumulated 2.7- and 1.4-fold higher CBSV RNA than 60444 and TME 204, respectively. Similarly, UCBSV was detected in all UCBSV-challenged plants of 60444 and TME 204, with 60444 found to have accumulated 1.4-fold higher UCBSV RNA than TME 204 (Supplementary Fig. S1). UCBSV was not detected by RT-PCR in plants of NASE 3 challenged with UCBSV. These results correlated with the visually assessed disease symptoms and indicated that CBSV was able to infect and accumulate viral RNA in all plants across genotypes, whereas UCBSV infected and accumulated viral RNA in TME 204 and 60444 but not in NASE 3 plants.

1.3.2 Deep sequencing of small RNAs in virus-challenged cassava genotypes

Small RNA libraries for CBSV- and UCBSV-challenged and unchallenged cassava genotypes were prepared on Illumina platform. Sequence data showing total small RNAs obtained from

leaves of uninfected and UCBSV- and CBSV-infected plants of TME 204, 60444 and NASE 3 are summarized in Table 2. The data presented is an average of three biological replicates per virus challenge for each genotype. The small RNA clean sequence reads (21 to 24 nucleotide adapter-trimmed sequence reads with Phred Quality score 20 and above) ranged from *c.* 10,000 to 1.3 million across libraries. The majority of small RNA reads were host derived. CBSV-derived small RNAs from cultivars TME 204, 60444 and NASE 3 challenged with CBSV accounted for up to 7%, 13% and 10%, respectively, of the total clean small RNA reads. Similarly, UCBSV-derived small RNAs in TME 204 and 60444 graft-inoculated with UCBSV accounted for 0.4% and 2.6%, respectively (Table 2). This contrasted with insignificant levels of UCBSV-derived small RNAs (0.002%) and UCBSV viral RNA accumulation obtained from UCBSV-challenged NASE 3 plants. The population of small RNAs in unchallenged control plants that mapped to viral genomes was minimal, accounting for 0.015%, 0.117% and 0.005% of total clean reads in TME 204, 60444 and NASE 3, respectively (Table 2; Fig. 1A).

Total small RNA reads identified from virus-infected plants were sorted according to their numbers and length. In all three cultivars, uninfected or infected with the virus, 21 nt total small RNAs were most abundant followed by 22 nt class. The 21 nt class was more abundant in virus-infected plants, representing 45-60% of small RNAs compared to 30-32% in uninfected plants (Fig. 1B). A similar trend was observed for 22 nt siRNAs. The proportion of 24 nt siRNAs was high in uninfected plants, accounting for 20-30% of siRNAs clean reads compared to <10% in virus-infected plants across genotypes. UCBSV-challenged and unchallenged plants of NASE 3 however accumulated similar levels (20-30%) of 24 nt siRNAs (Fig. 1B).

1.3.3 Characterization of CBSV- and UCBSV-derived small RNAs

To enable comparison of data across libraries, siRNA reads were normalized as reads per one hundred thousand of the total clean reads of corresponding samples. Sequences showing no mismatches were regarded as CBSV- or UCBSV-derived small RNAs, and each could be unambiguously assigned to one unique genome position. Small RNAs in the range of 21-24 nucleotides in size were included in this study. Analyses showed that the 21 nt size class was predominant (61-79%) followed by 22 nt size class (19-37%) for both viruses across the three cassava genotypes (Fig. 1C). Proportions of 23 and 24 nt classes were very low (1-2%; Fig. 1C). The vsRNAs of all size classes were mapped throughout the CBSV and UCBSV genomes in both sense and antisense orientations. A slight bias towards the sense polarity was observed, with 56-57% of the total CBSV-derived small RNAs and 60-63% of the total UCBSV-derived small RNAs of positive polarity. Similar results were obtained for unique vsRNAs (Fig. 1D), although the proportion of sense-unique UCBSV-derived small RNAs dropped to 53-54%.

Different proportions of CBSV- and UCBSV-derived small RNAs were also reflected in line graphs obtained from normalized genome coverage data for each virus (Fig. 2A-F). The major classes of vsRNAs (21 and 22 nt sizes) were produced from all parts of the viral genomes across cassava genotypes. The distribution along the genomes were however non-homogenous. Some parts of the genomes were seen to be expressing more small RNAs than others. There was no correlation between percent of GC content (GC%) in different open reading frames (ORFs) of CBSV or UCBSV with vsRNAs. Examination of the percent vsRNAs for each genomic regions showed that in CBSV-infected NASE 3 and CBSV- and UCBSV-infected 60444 and TME 204, the CI genomic region produced the highest population of vsRNAs, followed by genomic regions

that encode NIb, P1, and P3 proteins, which produced moderate population of vsRNAs across viral genomes (Fig. 2A-E; 3A-F). The 6K1 and 6K2 genomic regions produced the lowest population of vsRNAs across viral genomes (Fig. 3A-F). Remarkably, profiles of CBSV- and UCBSV-derived sRNAs were different in the same cassava genotype, but similar for each virus across genotypes, except for UCBSV-challenged NASE 3 (Fig. 2F).

In *A. thaliana*, sorting of small RNAs into Argonaute complexes is directed by the first 5' nucleotide (Mi et al., 2008). To predict selective interaction of vsRNAs with specific AGOs, the relative abundance of total vsRNAs was determined based on the first 5'-terminal nucleotides. Small RNAs with G as the first nucleotide at 5' end were the least abundant (8-11%) for both viruses across all three genotypes (Fig. 4A). In UCBSV-infected TME 204 and 60444, the majority of UCBSV-derived small RNAs had A (40-45%) followed by U (30-34%) as the most abundant first nucleotide at 5' end. In CBSV-challenged TME 204, 60444 and NASE 3, CBSV-derived small RNAs starting with A (32-34%), U (30-35%) and C (20-30%) at 5' end were all in similar proportions.

1.3.4 Distribution and frequency of vsRNAs along CBSV and UCBSV genomes

Small RNA mapping analysis identified four major hotspots rich in 21 and 22 nt vsRNAs in each of the CBSV and UCBSV genomes (Fig. 2A-E). In the CBSV genome, the first hotspot was found within the cylindrical inclusion protein (CI) region from 2,216 to 4,105 nt position that correlated with 32-33% of all the CBSV-derived small RNAs produced from this region in TME 204 and 60444 (Fig. 3A & C). In NASE 3, only about 24% of CBSV-derived small RNAs were produced from this region (Fig. 3E). A hotspot was also found in the same region (from 2,209 to 4,092 nt) in the UCBSV genome, and was responsible for 17-20% of UCBSV-derived small

RNAs (Fig. 3A & C). A second hotspot was located in the large nuclear inclusion protein (NIb) region from 5,522 to 7,026 nt in the CBSV, and 5,506 to 7,011 nt in the UCBSV genome. These hotspots produced 11-13% CBSV-derived small RNAs and 16-19% UCBSV-derived small RNAs in the CBSV and UCBSV genomes, respectively (Fig. 3A, C & E). The third and fourth hotspots were found located in the P1 and P3 genomic regions. The P1 region in the CBSV genome (104–1,177 nt) yielded 16-20% CBSV-derived small RNAs, whereas the P3 region (1,178-2,059 nt) produced 8-12% CBSV-derived small RNAs. In the UCBSV genome, the P1 region (85-1,170 nt) produced 14-18% of UCBSV-derived small RNAs whereas the P3 region (1,171-2,052 nt) produced 19-20% of UCBSV-derived small RNAs (Fig. 3A & C). Similar trends were obtained for unique vsRNAs across genotypes (Fig. 3B, D & F).

1.3.5 Expression of DCL and AGO2 proteins in CBSV- and UCBSV-infected cassava

Dicer-like proteins play a major role in the production of small RNAs. To understand the effect of CBSV and UCBSV infection on RNA silencing pathways, qRT-PCR analysis was performed to determine transcript levels of cassava homologues of DCL (*Manihot esculenta* DCL) proteins and their correlation to the abundance of 21 and 22 nt vsRNAs in CBSV- and UCBSV-infected plants. Differing patterns of expression for these genes were observed across the three cassava genotypes when they were infected with the two viral pathogens. In CBSV-infected TME 204, there was down-regulation of *MeDCL2* mRNA and 1.2-fold up-regulation of *MeDCL4* mRNA compared to the uninfected control. In UCBSV-infected TME 204, both *MeDCL2* and *MeDCL4* transcript levels were downregulated. In CBSV-infected 60444, both *MeDCL2* and *MeDCL4* transcript levels were downregulated, whereas in UCBSV-infected 60444 plants both *MeDCL2*

and *MeDCL4* mRNA levels were upregulated (Table 2, Fig. 4B). In NASE 3, *MeDCL2* mRNA level was downregulated in both CBSV- and UCBSV-challenged plants compared to unchallenged controls. In contrast, *MeDCL4* mRNA was upregulated in both CBSV- and UCBSV-challenged plants compared to unchallenged controls (Table 2; Fig. 4B).

Multiple AGO-bound small RNAs guide effector complexes to the target viral mRNAs in a sequence-specific manner leading to either translational repression or mRNA cleavage (Ding, 2010; Ding and Voinnet, 2007; Llave, 2010). The vsRNAs, dictated by their first 5'-end nucleotides, are preferentially sorted and loaded into multiple AGO complexes (Mi et al., 2008; Takeda et al., 2008). Specifically, AGO1 has been shown to have preference for U, AGO2 and AGO4 have preference for A or U, while AGO5 prefers C at the first 5'-end of the siRNA (Donaire et al., 2009; Mi et al., 2008; Qi et al., 2009; Takeda et al., 2008). Since in this study a major proportion of vsRNAs have A or U at the first 5' end, accumulation of cassava homologue of AGO2 (*MeAGO2*) mRNAs was determined by qRT-PCR. The results showed that *MeAGO2* mRNA was upregulated in all three UCBSV-infected cassava genotypes and CBSV-infected TME 204 and NASE 3 plants. However, *MeAGO2* expression remained unchanged in CBSV-infected 60444 (Table 2; Fig. 4B).

1.4 Discussion

To decipher the molecular mechanism and RNAi components involved in CBSV and UCBSV infection, three cassava genotypes TME 204 (CBSD susceptible), 60444 (CBSD susceptible) and NASE 3 (CBSD tolerant) (Abaca et al., 2013) were challenged with CBSV and UCBSV, and the

population and characteristics of sRNAs studied by next generation sequencing. Sequence analysis showed that populations of pathogen-derived siRNA varied across genotypes. Maximum populations of small RNAs that mapped to the virus genome were found in CBSV-infected plants of 60444 (12.7%) followed by NASE (9.9%) and TME 204 (7.3%), respectively. In UCBSV-infected plants, maximum populations of UCBSV-derived sRNAs were found in 60444 (2.6%) whereas 0.4% of UCBSV-derived sRNAs accumulated in TME 204. In NASE 3, UCBSV-derived sRNAs were insignificant (0.002).

Quantitative RT-PCR analysis also showed that NASE 3 accumulated 2.7- and 1.4-fold higher CBSV RNA than CBSV-infected plants of 60444 and TME 204, respectively (Fig. S1). Similarly, UCBSV-infected 60444 accumulated 1.4-fold more UCBSV RNA than UCBSV-infected TME 204 (Fig. S1). Comparable qRT-PCR data was obtained in a previous study using field and glasshouse samples from the same cassava genotypes (Ogwok et al., 2015). Data presented here for the lack of visual symptom development and non-detectable presence of the virus confirms that observation that UCBSV does not infect NASE 3 plants (Table 2). The variation in cultivar responses could be due to their ability to differentially recognize CBSV and UCBSV and accordingly trigger antiviral defense mechanisms, or it could be due to inherent differences in the infection cycle of the two viruses including viral replication and accumulation of vsRNAs (Naveed et al., 2014). Hitherto, there was a positive correlation between virus titer and the levels of vsRNAs across genotypes. Susceptible cassava genotypes (TME 204, 60444 and NASE 3 for CBSV, and TME 204 and 60444 for UCBSV) accumulated high virus titer and high levels of vsRNAs. We suggest that the high accumulation of vsRNAs observed in susceptible genotypes may be due to failure to target virus genome resulting in continued virus replication and symptoms persistence. In contrast, NASE 3, resistant to UCBSV, accumulated

negligible UCBSV mRNA and insignificant levels of UCBSV-derived sRNAs. This may be due to efficient PTGS of viral mRNA, leading to a depletion of vsRNA populations.

CBSV or UCBSV infection was shown here to alter host small RNA profiles across the three cassava genotypes studied. The level of 23 and 24 nt size classes were high in uninfected libraries but were less abundant in virus-infected libraries across all genotypes. The 21 and 22 nt siRNAs were abundant in uninfected plants but were even higher in virus-infected plants across genotypes. Similar studies in uninfected *A. thaliana* showed that the 24 nt siRNAs were most abundant (35%), followed by 21 nt siRNAs (28%), but upon infection with *Cabbage leaf curl virus* (CaLCuV), the host siRNA profile was altered such that the 21 nt class became more abundant (32%) followed by the 24 nt class (28%) (Aregger et al., 2012). Also, in *Sugarcane mosaic virus* (SCMV)-infected maize, there was increased level of 21 and 22 nt size classes, whereas that of 24 nt size class decreased (Xia et al., 2014). Contrastingly, *Cauliflower mosaic virus* (CaMV)-infected *A. thaliana* resulted in overaccumulation of the 24 nt siRNAs (Blevins et al., 2011).

Reports in *A. thaliana* have shown that DCL4, DCL2 and DCL3 generate the 21, 22, and 24 nt vsRNAs (Blevins et al., 2006; Donaire et al., 2008; Ruiz-Ferrer and Voinnet, 2009). In addition, studies in *A. thaliana* infected with *Turnip mosaic virus* (TuMV), crucifer-infecting strain of *Tobacco mosaic virus* (TMV-Cg), *Cucumber mosaic virus* (CMV), and *Tobacco rattle virus* (TRV) (Donaire et al., 2008; Donaire et al., 2009; Lin et al., 2010; Qi et al., 2009) and *Nicotiana benthamiana* infected with *Potato virus X* (PVX), *Bamboo mosaic virus* (BMV), and *Pepper mild mottle virus* (PMMoV) (Donaire et al., 2009; Lin et al., 2010); potato infected with three different strains of *Potato virus Y* (PVY-O, PVY-N, and PVY-NTN) (Naveed et al., 2014);

grapevines infected with *Grapevine fleck virus* (GFkV) and *Grapevine rupestris stem-pitting associated virus* (GRSPaV) (Pantaleo et al., 2010); cotton infected with *Cotton leaf roll dwarf virus* (CLRDV) (Silva et al., 2011); and maize infected with SCMV (Xia et al., 2014), among others, revealed that the majority of vsRNAs in infected plants belonged to the 21 and 22 nt size classes, which are associated with activated PTGS (Blevins et al., 2006; Deleris et al., 2006; Donaire et al., 2008). Similarly, the 21 and 22 nt CBSV- and UCBSV-derived sRNAs were shown here to be the most predominant, which suggests activity of homologues of DCL4 and DCL2 in response to virus infection in cassava (Fig. 4B). The low abundance of the 23 and 24 nt vsRNAs in both CBSV- and UCBSV-infected plants across cassava genotypes suggests marginal activity of DCL3 in cassava against these viruses (Qi et al., 2009).

DCL2 and DCL4 proteins are key component of RNAi pathways and both are required for optimal resistance against viruses. Despite differences in the silencing activity of their small RNA products, DCL2 and DCL4 mostly act redundantly yet hierarchically when present simultaneously (Parent et al., 2015). Quantitative PCR analysis of DCL2 and DCL4 expression in this study showed that both proteins have different expression patterns in response to CBSV and UCBSV infection. Particularly, in NASE 3 challenged with either CBSV or UCBSV and TME 204 challenged with CBSV, DCL4 was up-regulated whereas DCL2 was down-regulated. A similar observation was reported in CLRDV-infected cotton (Silva et al., 2011). Presumably an up-regulation of DCL4 is sufficient to suppress UCBSV infection in NASE 3 plants. A detailed study of both DCL2 and DCL4 in a background of two different host genotypes would be of great interest.

Processing of dsRNAs by Dicer proteins would ideally yield an equal amount of sense and antisense vsRNAs, the life span of which depends on selective incorporation into specific AGO proteins (Qi et al., 2009). Analysis of the CBSV- and UCBSV-derived small RNA polarity showed a bias towards the sense polarity compared to antisense polarity irrespective of size class across libraries, and was consistent with previous findings in other virus-infected plants (Donaire et al., 2008; Donaire et al., 2009; Naveed et al., 2014; Qi et al., 2009). Strand biases are usually attributed to preferential processing of highly structured single-stranded genomic viral RNAs by Dicer proteins (Ding and Voinnet, 2007; Donaire et al., 2009; Silva et al., 2011), and different viruses have been shown to produce, in the same host plant, virus-derived small RNAs with different ratios of sense to antisense polarity (Donaire et al., 2009; Pantaleo et al., 2010). However, a correlation between vsRNA hotspots and structured regions of genomic viral RNAs remains unclear (Donaire et al., 2009).

The preferential use of vsRNAs with A or U residues as compared to C and G residues as the first 5'-end nucleotide has been reported in *A. thaliana* infected with TuMV and PVX, *Cucumis melo* infected with WMV (Donaire et al., 2009), and in potato plants infected with three strains of PVY (Naveed et al., 2014). In contrast, a few cases of preferential use of C as the first 5'-terminal nucleotide has been reported in grapevines infected with GFkV and GRSPaV (Pantaleo et al., 2010), and tomato plants infected with *Tomato yellow leaf curl Sardinia virus* (TYLCSV) (Miozzi et al., 2013). However, most previous studies reported a tendency to avoid vsRNAs with G residues at the first 5'-end (Donaire et al., 2009; Mi et al., 2008; Pantaleo et al., 2010), probably due to absence of AGO proteins with known preference for G residues at the first 5'-end (Mi et al., 2008). Our results indicate that A and U were the most abundant nucleotides at the first 5'-end (~ 80%), while G was the least abundant at the first 5'-end (Fig. 4A). In addition,

there was up-regulation of AGO2 mRNA in all three UCBSV-infected cassava genotypes and CBSV-infected TME 204 and NASE 3 (Fig. S1). Similar studies in *Turnip crinkle virus* (TCV)- and CMV-infected AGO2 mutant *A. thaliana* compared to wild type plants revealed the mutants were highly susceptible to TCV and CMV infection (Harvey et al., 2011). Besides, the induction of susceptibility involved activity of viral silencing suppressor proteins (Endres et al., 2010; Lewsey et al., 2010). Our results signify increased activity of AGO2 homologue in CBSV- and UCBSV-infected cassava. However, abundance of A, U and C at the first 5' terminus in similar proportions in CBSV-infected plants imply the involvement of multiple AGOs in sorting out the vsRNAs.

Host RDRs use viral ssRNA to synthesize dsRNAs, which serve as substrates for DCL-dependent formation of secondary vsRNAs to maintain systemic silencing throughout the plant. In addition, DCL4 and RDR1 were reported as major contributors to the abundant pool of 21 nt TuMV-derived siRNAs in *A. thaliana* (Garcia-Ruiz et al., 2010). However, the response of RDRs in CBSV- and UCBSV-infected cassava was not assessed in this study. Since ipomoviruses and potyviruses (TuMV) share many characteristics, such a study is worth considering.

Furthermore, viral suppressor proteins have been shown to employ a multitude of mechanisms (Csorba et al., 2015; Ding and Voinnet, 2007). For example, the P0 protein of TCV has been shown to target AGO1, leading to its degradation though it does not interfere with siRNA-RISC assembly. Similarly, potyviral HC-Pro and ipomoviral P1b protein of *Cucumber vein yellowing virus* (CVYV) have been shown to suppress plants silencing machinery through siRNA sequestration thereby interfering with viral RNA degradation (Valli et al., 2007). In addition,

although the ipomoviruses, *Sweet potato mild mottle virus* (SPMMV) and *Tomato mild mottle virus* (TomMMV), both encode HC-Pro, which is related to potyviral HC-Pro, the HC-Pro of SPMMV lacks RNA-silencing suppressor activity. Instead, the P1 protein act as the RNA-silencing suppressor by binding with AGO1 thereby interrupting RISC assembly (Dombrovsky et al., 2014). The function of the CBSV or UCBSV P1 protein, a putative RNA silencing suppressor (Mbanzibwa et al., 2009a) remains a subject for further study. This could provide insights into the observed disparity between CBSV and UCBSV pathogenicity.

To our knowledge, this study provides the first high-resolution genome map of vsRNAs for an ipomovirus in the family *Potyviridae*. The use of deep sequencing in this study has provided an insight into the molecular interaction between CBSV causing viruses and cassava. The population of vsRNAs were abundant, diverse and revealed widespread targeting of viral genomes by machinery of the gene silencing pathway. The overall composition of sRNAs in virus-infected cassava unveiled the action of different Dicer proteins in different cassava genotypes. The findings also provided an insight into the differential susceptibility of host plants to the same virus, which is reflected in the severity of symptoms they induce. Finally the results indicate that CBSV and UCBSV interact differently in the same host genetic background.

1.6 Conflict of interest: The authors have no conflict of interest to declare.

1.5 Acknowledgements

The work reported was supported with funds from the Bill and Melinda Gates Foundation, the United States Agency for International Development (USAID) from the American people and the Monsanto Fund. The authors thank Steven Hill and Doug Bryant for help with bioinformatics analysis.

1.7 References

Abaca, A., Kawuki, R., Tukamuhabwa, P., Baguma, Y., Pariyo, A., Alicai, T., Abidrabo, P., Katono, K., Bua, A., 2013. Genetic relationships of cassava genotypes that are susceptible or tolerant to cassava brown streak disease in Uganda. *J. Agric. Sci.* 5(7), p107.

Aregger, M., Borah, B.K., Seguin, J., Rajeswaran, R., Gubaeva, E.G., Zvereva, A.S., Windels, D., Vazquez, F., Blevins, T., Farinelli, L., 2012. Primary and secondary siRNAs in geminivirus-induced gene silencing. *PLoS Pathog.* 8(9), e1002941.

Blevins, T., Rajeswaran, R., Aregger, M., Borah, B.K., Schepetilnikov, M., Baerlocher, L., Farinelli, L., Meins, F., Hohn, T., Pooggin, M.M., 2011. Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Res.* 39(12), 5003-5014.

- Blevins, T., Rajeswaran, R., Shivaprasad, P.V., Beknazariants, D., Si-Ammour, A., Park, H.-S., Vazquez, F., Robertson, D., Meins, F., Hohn, T., 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* 34(21), 6233-6246.
- Brodersen, P., Voinnet, O., 2006. The diversity of RNA silencing pathways in plants. *Trends Genet.* 22(5), 268-280.
- Bull, S., Owiti, J., Niklaus, M., Beeching, J.R., Gruijssem, W., Vanderschuren, H., 2009. Agrobacterium-mediated transformation of friable embryogenic calli and regeneration of transgenic cassava. *Nat. Protoc.* 4(12), 1845-1854.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* 7(5), 335-336.
- Csorba, T., Kontra, L., Burgyán, J., 2015. Viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology* 479, 85-103.
- Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K.D., Carrington, J.C., Voinnet, O., 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* 313(5783), 68-71.
- Ding, S.-W., 2010. RNA-based antiviral immunity. *Nat. Rev. Immunol.* 10(9), 632-644.
- Ding, S.-W., Voinnet, O., 2007. Antiviral immunity directed by small RNAs. *Cell* 130(3), 413-426.
- Dombrovsky, A., Reingold, V., Antignus, Y., 2014. *Ipomovirus* - an atypical genus in the family Potyviridae transmitted by whiteflies. *Pest Manag. Sci.* 70(10), 1553-1567.
- Donaire, L., Barajas, D., Martínez-García, B., Martínez-Priego, L., Pagán, I., Llave, C., 2008. Structural and genetic requirements for the biogenesis of tobacco rattle virus-derived small interfering RNAs. *J. Virol.* 82(11), 5167-5177.

Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K.F., Aranda, M.A., Llave, C., 2009. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* 392(2), 203-214.

Endres, M.W., Gregory, B.D., Gao, Z., Foreman, A.W., Mlotshwa, S., Ge, X., Pruss, G.J., Ecker, J.R., Bowman, L.H., Vance, V., 2010. Two plant viral suppressors of silencing require the ethylene-inducible host transcription factor RAV2 to block RNA silencing. *PLoS Pathog.* 6(1), e1000729.

Garcia-Ruiz, H., Takeda, A., Chapman, E.J., Sullivan, C.M., Fahlgren, N., Brempelis, K.J., Carrington, J.C., 2010. Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *Plant Cell* 22(2), 481-496.

Harvey, J., Lewsey, M.G., Patel, K., Westwood, J., Heimstadt, S., Carr, J.P., Baulcombe, D.C., 2011. An antiviral defense role of AGO2 in plants. *PLoS One* 6(1), e14639-e14639.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G., Ferguson, M.E., 2014. Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virol. J.* 11(1), 216.

Langmead, B., 2010. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinform.* 32, 11.17. 11-11.17. 14.

Legg, J.P., Jeremiah, S.C., Obiero, H.M., Maruthi, M.N., Ndyetabula, I., Okao-Okuja, G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 159, 161-170.

Lewsey, M.G., Murphy, A.M., MacLean, D., Dalchau, N., Westwood, J.H., Macaulay, K., Bennett, M.H., Moulin, M., Hanke, D.E., Powell, G., 2010. Disruption of two defensive signaling pathways by a viral RNA silencing suppressor. *Mol. Plant-Microbe Interact.* 23(7), 835-845.

- Lin, K.-Y., Cheng, C.-P., Chang, B.C.-H., Wang, W.-C., Huang, Y.-W., Lee, Y.-S., Huang, H.-D., Hsu, Y.-H., Lin, N.-S., 2010. Global analyses of small interfering RNAs derived from Bamboo mosaic virus and its associated satellite RNAs in different plants. *PLoS One* 5(8), e11928.
- Lister, R., 1959. Mechanical transmission of cassava brown streak virus. *Nature* 183, 1588 - 1589.
- Llave, C., 2010. Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends Plant Sci.* 15(12), 701-707.
- Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J., Waterhouse, P.M., 2006. The evolution and diversification of Dicers in plants. *FEBS Lett.* 580(10), 2442-2450.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17(1), 10-12.
- Maruthi, M., Bouvaine, S., Tufan, H.A., Mohammed, I.U., Hillocks, R.J., 2014. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. *PLoS One* 9(5), e96642.
- Maruthi, M.N., Hillocks, R.J., Mtunda, K., Raya, M.D., Muhanna, M., Kiozia, H., Rekha, A.R., Colvin, J., Thresh, J.M., 2005. Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *J. Phytopathol.* 153(5), 307-312.
- Mbanzibwa, D., Tian, Y., Tugume, A., Patil, B., Yadav, J., Bagewadi, B., Abarshi, M., Alicai, T., Changadeya, W., Mkumbira, J., 2011. Evolution of cassava brown streak disease-associated viruses. *J. Gen. Virol.* 92(4), 974-987.
- Mbanzibwa, D.R., Tian, Y., Mukasa, S.B., Valkonen, J.P., 2009a. Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *J. Virol.* 83(13), 6934-6940.

- Mbanzibwa, D.R., Tian, Y.P., Tugume, A.K., Mukasa, S.B., Tairo, F., Kyamanywa, S., Kullaya, A., Valkonen, J.P., 2009b. 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.
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., 2008. Sorting of small RNAs into *Arabidopsis* Argonaute complexes is directed by the 5' terminal nucleotide. *Cell* 133(1), 116-127.
- Miozzi, L., Pantaleo, V., Burgyán, J., Accotto, G.P., Noris, E., 2013. Analysis of small RNAs derived from tomato yellow leaf curl Sardinia virus reveals a cross reaction between the major viral hotspot and the plant host genome. *Virus Res.* 178(2), 287-296.
- Moissiard, G., Voinnet, O., 2006. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four *Arabidopsis* Dicer-like proteins. *Proc. Natl. Acad. Sci.* 103(51), 19593-19598.
- Molnár, A., Csorba, T., Lakatos, L., Várallyay, É., Lacomme, C., Burgyán, J., 2005. Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. *J. Virol.* 79(12), 7812-7818.
- Monger, W., Seal, S., Cotton, S., Foster, G., 2001. Identification of different isolates of Cassava brown streak virus and development of a diagnostic test. *Plant Pathol.* 50(6), 768-775.
- Naveed, K., Mitter, N., Harper, A., Dhingra, A., Pappu, H.R., 2014. Comparative analysis of virus-specific small RNA profiles of three biologically distinct strains of Potato virus Y in infected potato (*Solanum tuberosum*) cv. Russet Burbank. *Virus Res.* 191, 153-160.
- Ogwok, E., Alicai, T., Rey, M., Beyene, G., Taylor, N., 2015. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta* Crantz) plants. *Plant Pathol.* 64(5), 1235-1246.
- Ogwok, E., Odipio, J., Halsey, M., Gaitan-Solis, E., Bua, A., Taylor, N.J., Fauquet, C.M., Alicai, T., 2012. Transgenic RNA interference (RNAi)-derived field resistance to cassava brown streak disease. *Mol. Plant Pathol.* 13(9), 1019-1031.

- Ogwok, E., Patil, B.L., Alicai, T., Fauquet, C.M., 2010. Transmission studies with Cassava brown streak Uganda virus (Potyviridae, Ipomovirus) and its interaction with abiotic and biotic factors in *Nicotiana benthamiana*. *J. Virol. Methods* 169(2), 296-304.
- Pantaleo, V., Saldarelli, P., Miozzi, L., Giampetruzzi, A., Gisel, A., Moxon, S., Dalmay, T., Bisztray, G.r., Burgyan, J., 2010. Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. *Virology* 408(1), 49-56.
- Parent, J.-S., Bouteiller, N., Elmayan, T., Vaucheret, H., 2015. Respective contributions of *Arabidopsis* DCL2 and DCL4 to RNA silencing. *Plant J.* 81(2), 223-232.
- Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: a threat to food security in Africa. *J. Gen. Virol.* 96(Pt 5), 956-968.
- Patil, B.L., Ogwok, E., Wagaba, H., Mohammed, I.U., Yadav, J.S., Bagewadi, B., Taylor, N.J., Kreuze, J.F., Maruthi, M., Alicai, T., 2011. RNAi-mediated resistance to diverse isolates belonging to two virus species involved in Cassava brown streak disease. *Mol. Plant Pathol.* 12(1), 31-41.
- Prabha, K., Baranwal, V., Jain, R., 2013. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian J. Virol.* 24(2), 157-165.
- Qi, X., Bao, F.S., Xie, Z., 2009. Small RNA deep sequencing reveals role for *Arabidopsis thaliana* RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One* 4(3), e4971.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), 841-842.
- Ruiz-Ferrer, V., Voinnet, O., 2009. Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.* 60, 485-510.
- Seo, J.-K., Wu, J., Lii, Y., Li, Y., Jin, H., 2013. Contribution of small RNA pathway components in plant immunity. *Mol. Plant-Microbe Interact.* 26(6), 617-625.

Silva, T.F., Romanel, E.A., Andrade, R.R., Farinelli, L., Østerås, M., Deluen, C., Corrêa, R.L., Schrago, C.E., Vaslin, M.F., 2011. Profile of small interfering RNAs from cotton plants infected with the polerovirus Cotton leafroll dwarf virus. *BMC Mol. Biol.* 12(1), 40.

Smith, N.A., Eamens, A.L., Wang, M.-B., 2011. Viral small interfering RNAs target host genes to mediate disease symptoms in plants. *PLoS Pathog.* 7(5), e1002022.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., Watanabe, Y., 2008. The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol.* 49(4), 493-500.

Taylor, N., Gaitan-Solis, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., Trembley, C., Abernathy, V., Corbin, D., Fauquet, C.M., 2012a. A High-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Trop. Plant Biol.* 5(1), 127-139.

Taylor, N.J., Halsey, M., Gaitan-Solis, E., Anderson, P., Gichuki, S., Miano, D., Bua, A., Alicai, T., Fauquet, C.M., 2012b. The VIRCA project: virus resistant cassava for Africa. *GM Crops Food* 3(2), 93-103.

Valli, A., Lòpez-Moya, J.J., Garcia, J.A., 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* 88(3), 1016-1028.

Vanderschuren, H., Moreno, I., Anjanappa, R.B., Zainuddin, I.M., Gruissem, W., 2012. Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PLoS One* 7(9), e45277.

Visser, M., Maree, H.J., Rees, D.J., Burger, J.T., 2014. High-throughput sequencing reveals small RNAs involved in ASGV infection. *BMC Genomics* 15(1), 568.

Wagaba, H., Beyene, G., Trembley, C., Alicai, T., Fauquet, C.M., Taylor, N.J., 2013. Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. *BMC Res. Notes* 6(1), 516.

Xia, Z., Peng, J., Li, Y., Chen, L., Li, S., Zhou, T., Fan, Z., 2014. Characterization of small interfering RNAs derived from Sugarcane mosaic virus in infected maize plants by deep sequencing. *PLoS One* 9(5), e97013.

Yadav, J.S., Ogwok, E., Wagaba, H., Patil, B.L., Bagewadi, B., Alicai, T., Gaitan-Solis, E., Taylor, N.J., Fauquet, C.M., 2011. RNAi-mediated resistance to Cassava brown streak Uganda virus in transgenic cassava. *Mol. Plant Pathol.* 12(7), 677-687.

Zamore, P.D., Tuschl, T., Sharp, P.A., Bartel, D.P., 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101(1), 25-33.

Figure Legends

Fig. 1: Comparison of CBSV- and UCBSV-derived small RNA populations recovered from infected and uninfected plants of TME 204, 60444 and NASE 3 (n = 3). The histograms represent (A) sRNA mapped to CBSV and UCBSV genomes in sense and antisense orientations, (B) Total sRNA sorted by percent nucleotide sequence lengths, (C) sRNA mapped to CBSV and UCBSV genomes, sorted by sequence lengths and orientations, and (D) unique (non-redundant) sRNA mapped to CBSV and UCBSV genomes, sorted by sequence lengths and orientations. Error bars indicate standard error of the mean (SEM).

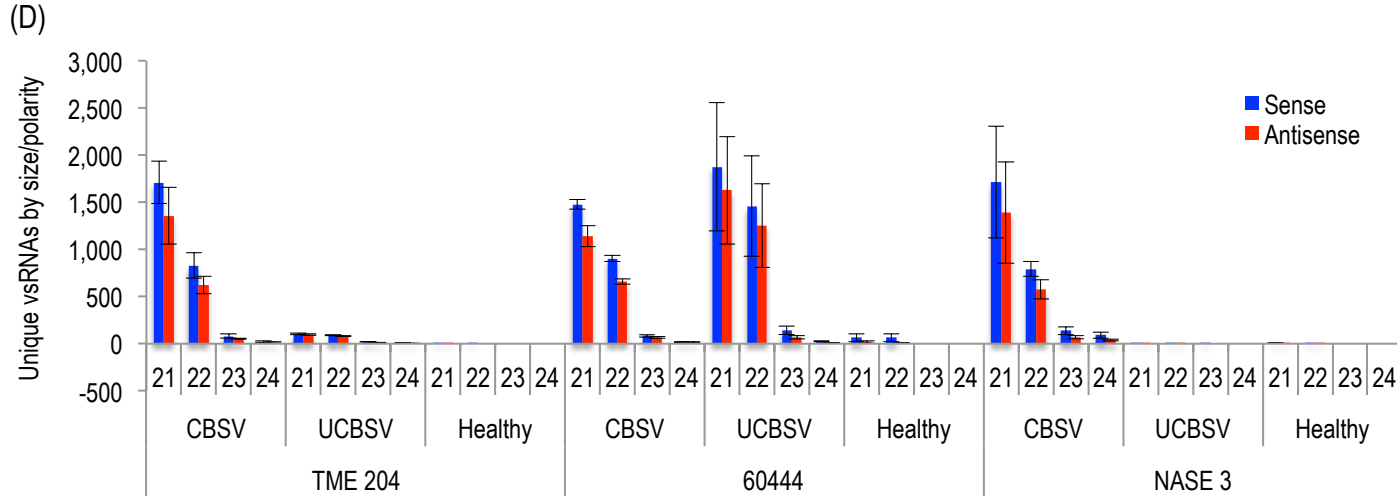
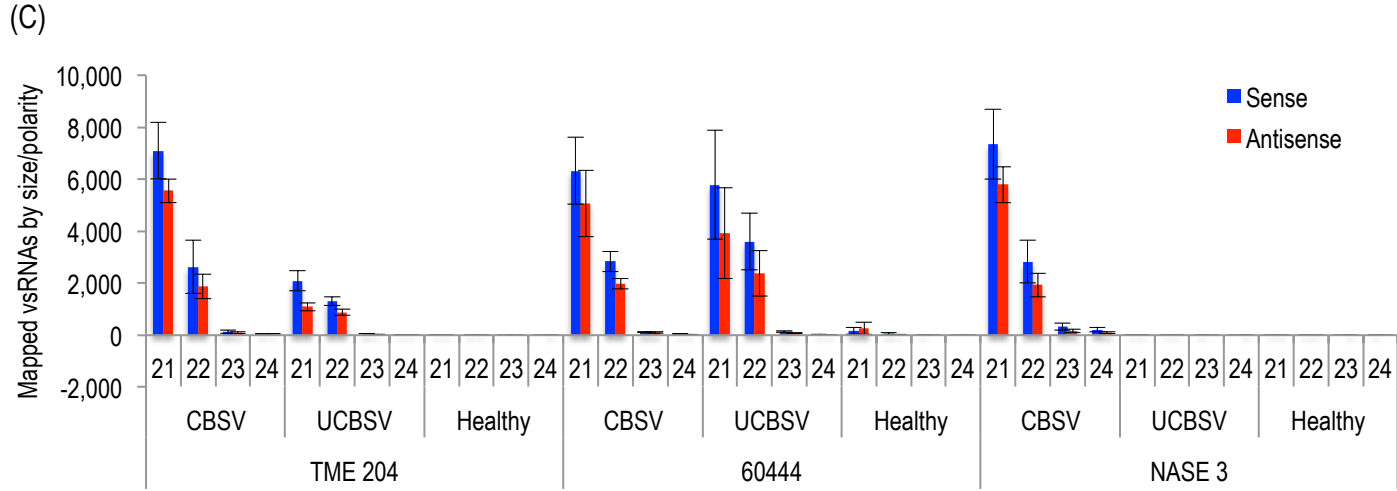
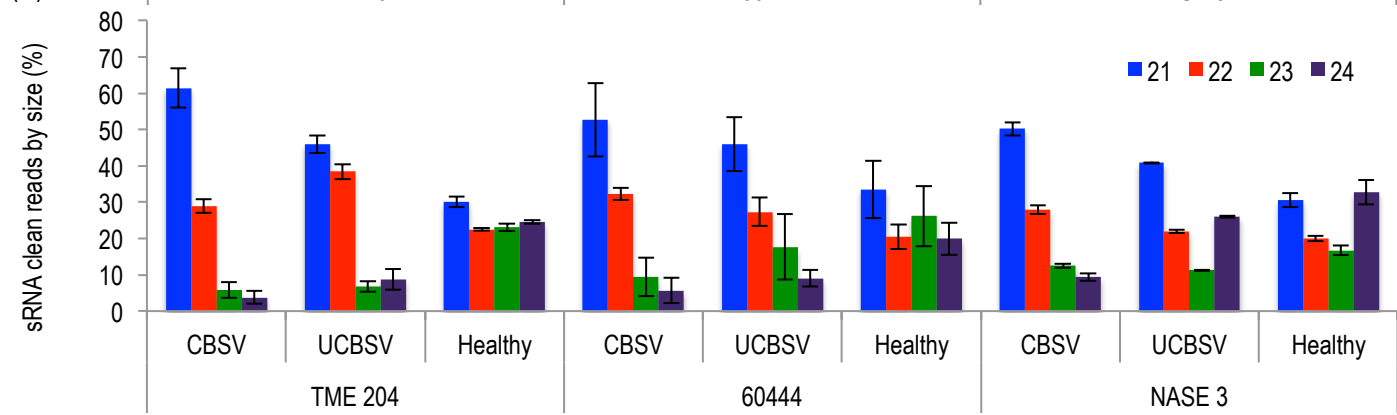
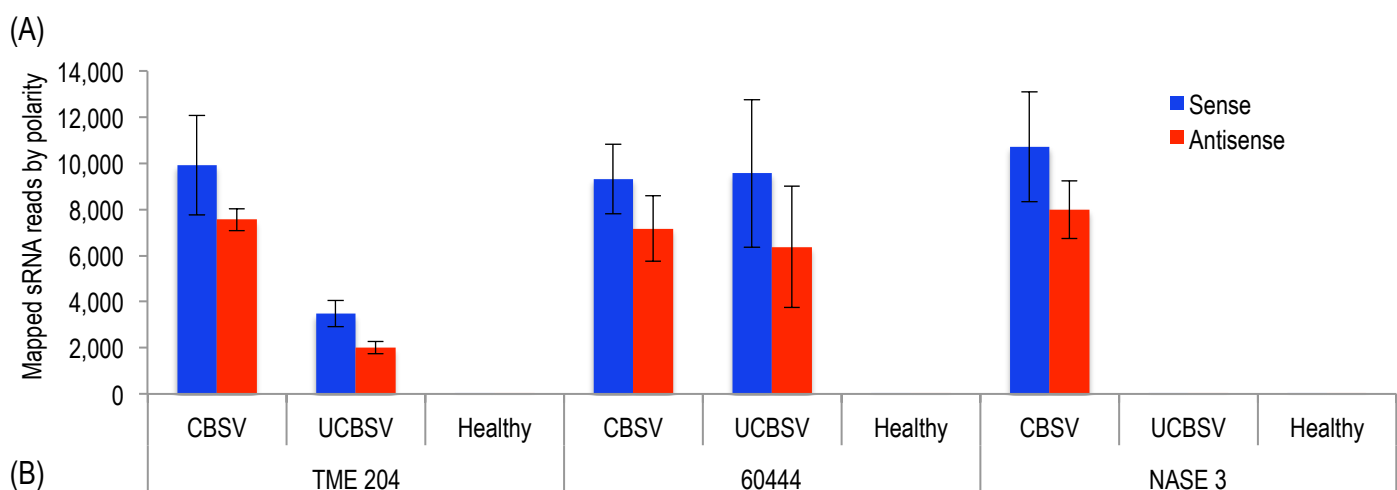


Fig. 2: Genomic map of CBSV- and UCBSV-inoculated libraries. The graphs plot the number of 21–24 nucleotide virus-derived small RNAs at each position along the CBSV and UCBSV genomes recovered from (A) CBSV- and (B) UCBSV-inoculated TME 204; (C) CBSV- and (D) UCBSV-inoculated 60444; and (E) CBSV-inoculated NASE 3. Bars above and below the x-axis represent sense and antisense virus-derived small RNA reads, respectively.

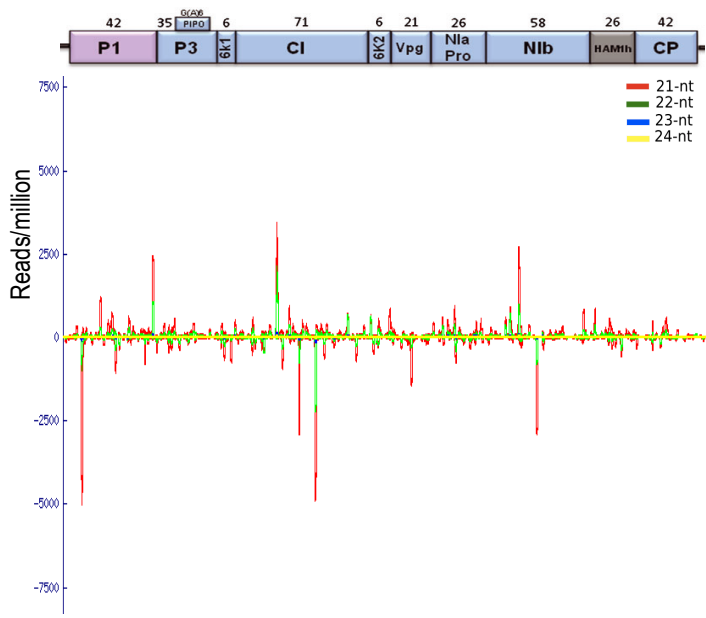
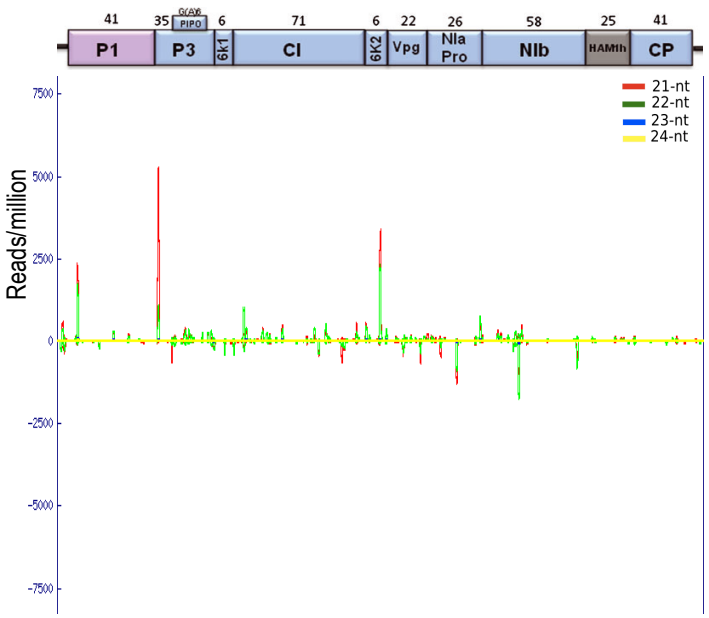
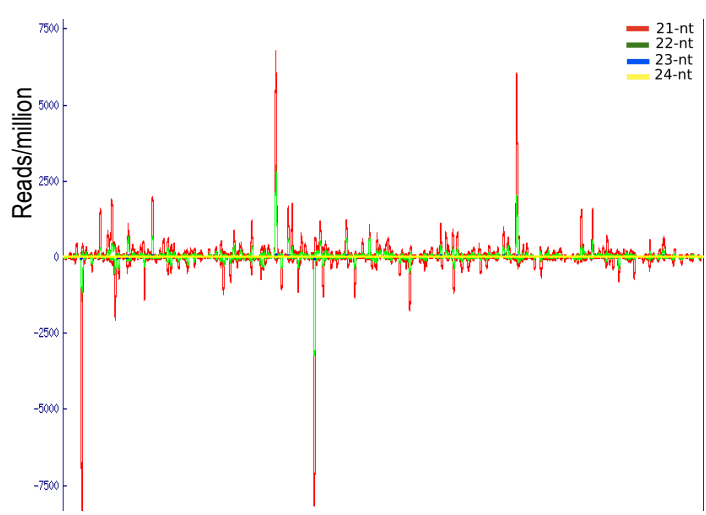
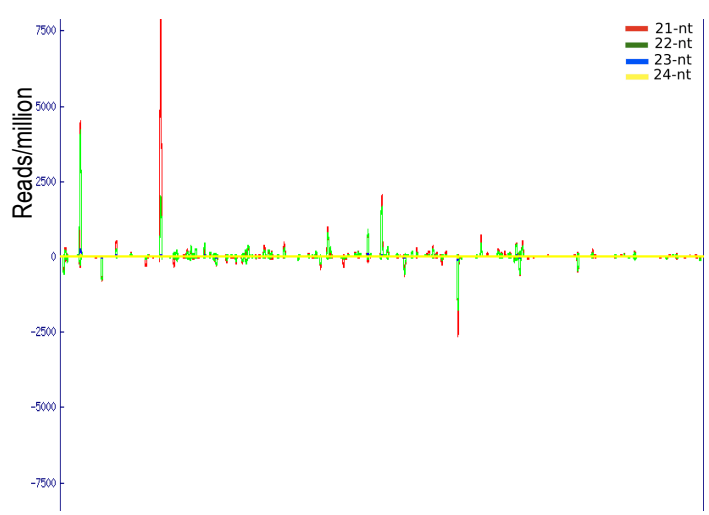
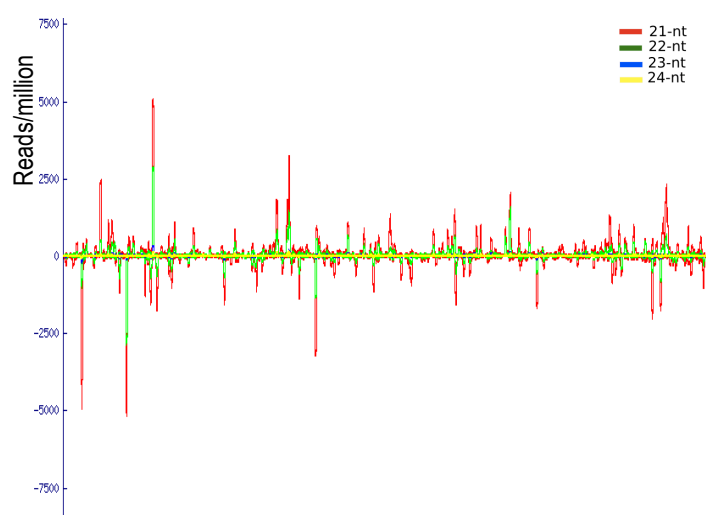
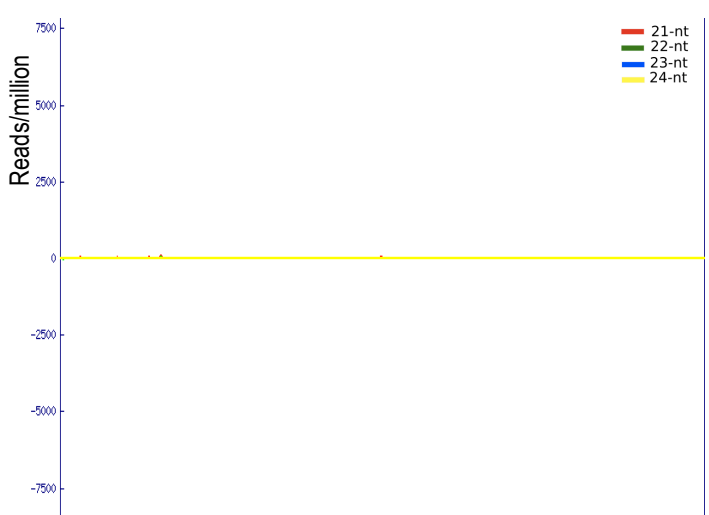
(A)**(B)****(C)****(D)****(E)****(F)**

Fig. 3: Total versus unique virus-derived small RNA population from individual genes of CBSV and UCBSV genomes in infected plants of TME 204, 60444 and NASE 3 ($n = 3$). The histograms represent (A) total mapped and (B) unique virus-derived small RNA CBSV- and UCBSV-infected libraries of TME 204; (C) total mapped and (D) unique virus-derived small RNAs in CBSV and UCBSV-infected libraries of 60444; and (E) total mapped and (F) unique virus-derived small RNAs in CBSV-infected libraries of NASE 3. Error bars indicate standard error of the mean (SEM).

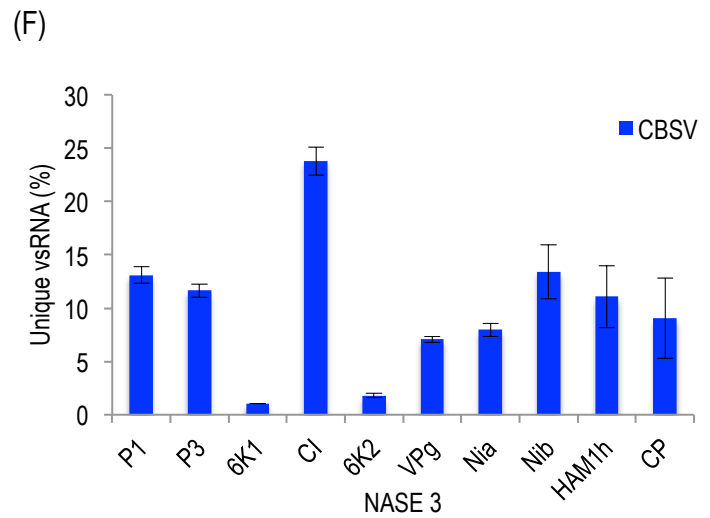
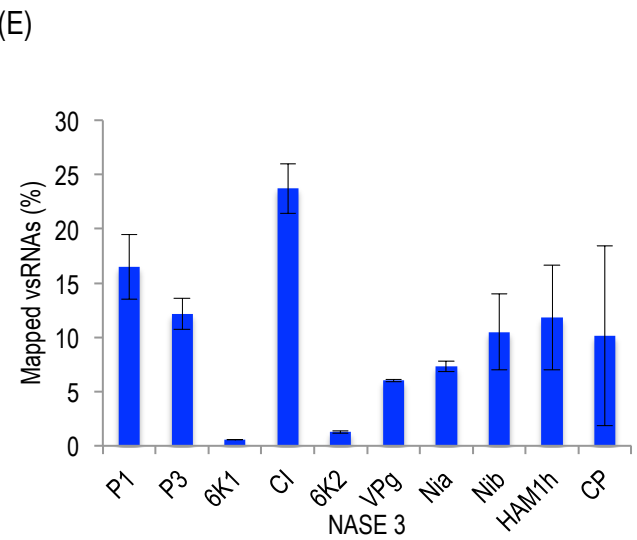
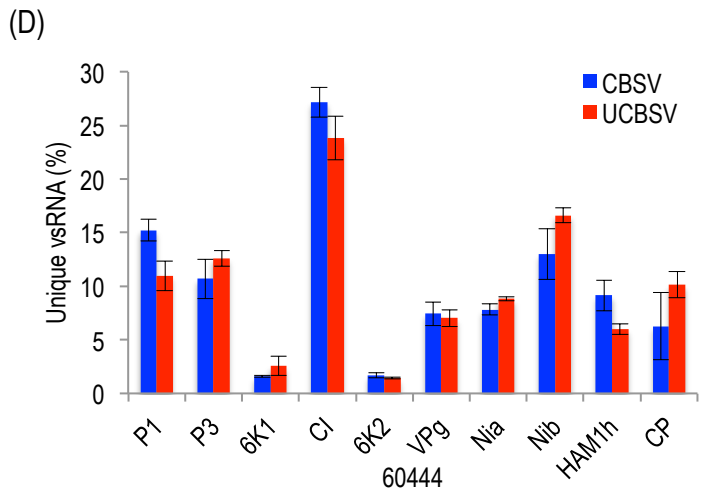
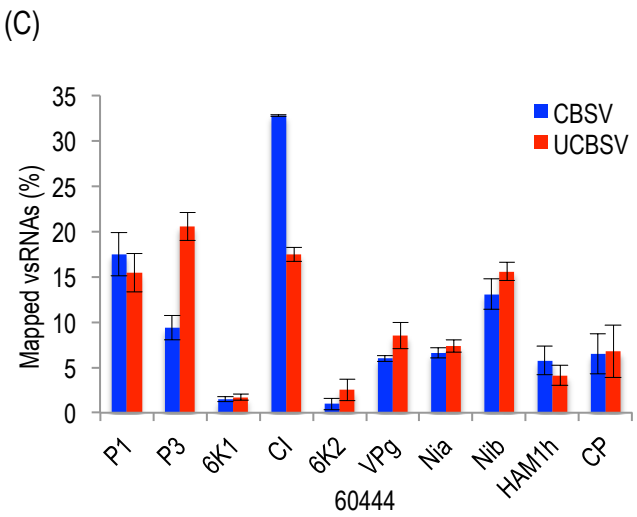
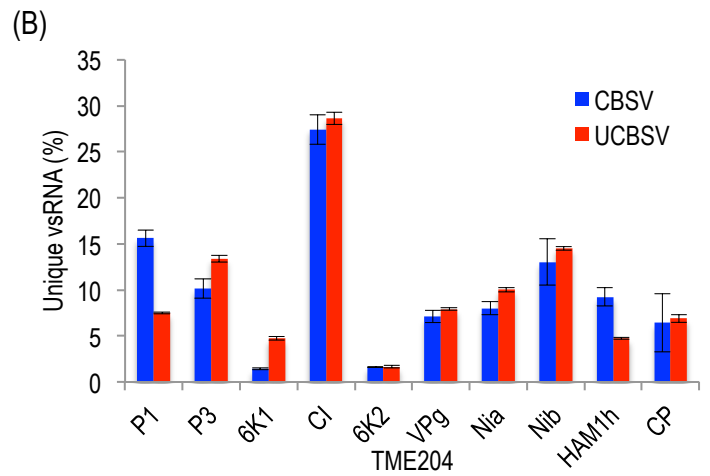
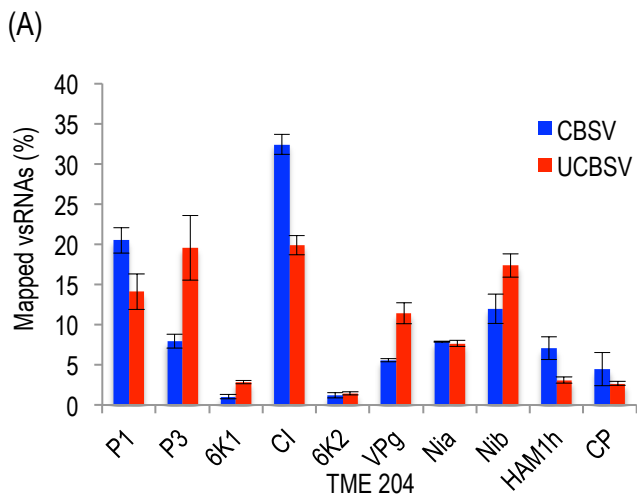
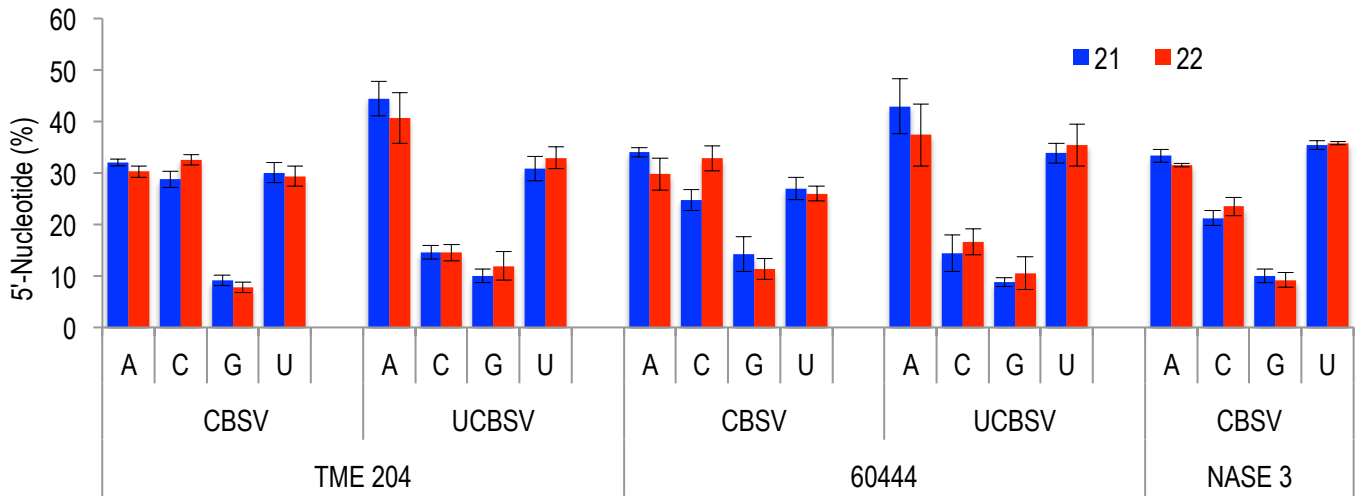


Fig. 4: Relative frequency of first 5'-end nucleotide and relative expression levels of Dicer and Argonaute proteins mRNAs. (A) The histograms represent relative abundance of the four distinct nucleotides at the 5'-end of 21 and 22 nucleotide sized virus-derived small RNAs of CBSV- and UCBSV-infected libraries of TME 204, 60444 and NASE 3 ($n = 3$). (B) Real-time PCR analysis of the expression levels of cassava homologues of AGO2, DCL2 and DCL4 in CBSV- and UCBSV-infected and uninfected plants of TME 204, 60444 and NASE 3 ($n = 3$). Error bars indicate standard error of the mean (SEM).

(A)



(B)

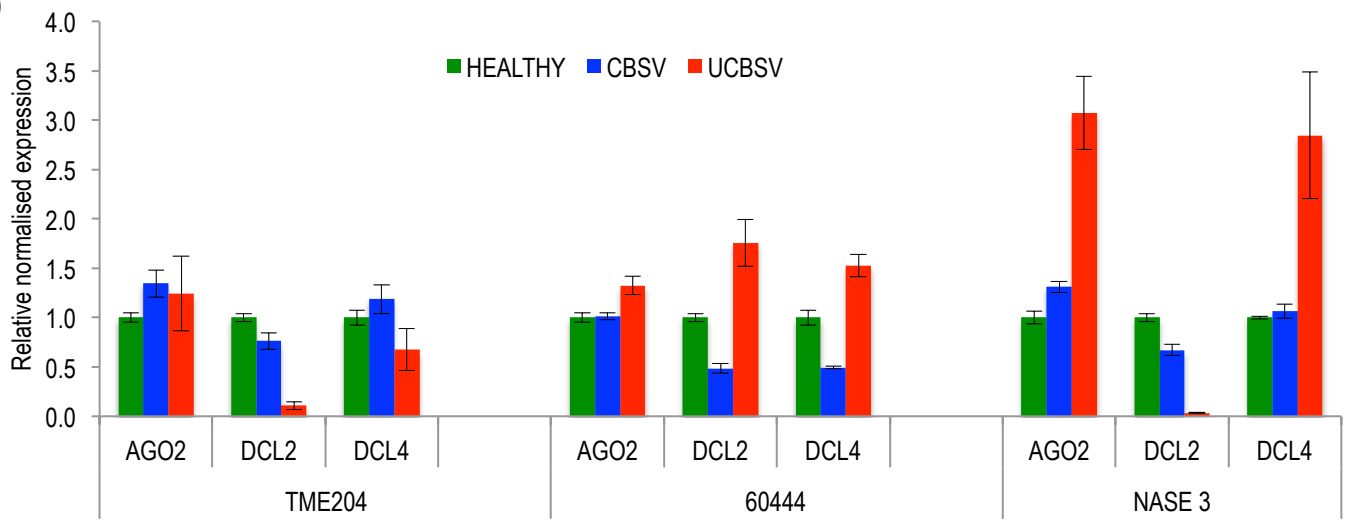


Fig. S1: Virus titer (relative to predetermined virus concentration obtained after performing standard serial dilution curves of CBSV or UCBSV) in tissues of cassava genotypes infected with CBSV and UCBSV. Relative expression levels were normalized using cytochrome c oxidase mRNA as internal control. Error bars indicate standard error of the mean (SEM).

Table 1: Primer sequences and amplicon sizes of cassava homologues of Argonaute 2 (AGO2), and Dicer-like proteins 2 and 4 (DCL2 and DCL4)

Gene	Primer code	Primer sequence	Amplicon size (bp)	Efficiency (%)	R ²	Transcript ID
AGO2	AGO2-F1	GGCAATCTCCAGCTTCAGCA	78	104.5	0.993	cassava4.1_000920m
	AGO2-R1	TCCAATGAAGCAGCCGATGA				
DCL2	DCL2-F3	ATGCACACTGACCTCGTC	110	99.1	0.986	cassava4.1_000931m
	DCL2-R3	GTCATCACAAGCACCTCA				
DCL4	DCL4-F3	TGCTACTAAAGTGGGTGAAGAAG	109	106.8	0.977	cassava4.1_001038m
	DCL4-R3	CGCACGTCCTCTAGATGGTATG				

Table 2. Summary of small RNA reads from two cassava brown streak viruses in three different cassava genotypes^a and effect of the viruses on AGO2, DCL2 and DCL4 expression

Genotype	Virus species	Mean CBSD severity 6 WAG (1-5) ^b	Total raw reads	Total clean reads	Small RNAs mapped to virus genome	Percent vsRNAs with respect to total clean reads	Effect of virus on AGO2, DCL2 and DCL4 expression		
							AGO2	DCL2	DCL4
TME204	CBSV	3.3	3,185,175	100,750	17,470	7.340	up	down	up
	UCBSV	2.0	3,994,736	1,282,370	5,488	0.428	up	down	down
	Healthy	1.0	2,740,771	86,524	13	0.015	na	na	na
60444	CBSV	4.3	3,198,984	130,400	16,496	12.650	na	down	down
	UCBSV	3.0	4,020,006	612,987	15,939	2.600	up	up	up
	Healthy	1.0	3,347,340	10,238	12	0.117	na	na	na
NASE3	CBSV	5.0	3,019,728	188,617	18,705	9.917	up	down	up
	UCBSV	1.0	2,511,404	410,641	10	0.002	up	down	up
	Healthy	1.0	2,154,804	213,163	10	0.005	na	na	na

^aData presented are an average of three biological replicates per genotype. ^bCBSD symptoms on grafted plants were scored as previously described (Ogwok et al., 2012). na = not applicable.