Identification of maize lethal necrosis disease causal viruses in maize and suspected alternative hosts through small RNA profiling

Ronald Tonui1 | Joel Masanga2 | Remmy Kasili3 | Steven Runo2 | Amos Alakonya4

1Institute of Science and Innovation, Pan African University, Nairobi, Kenya
2Department of Biochemistry, Microbiology and Biotechnology, Kenyatta University, Nairobi, Kenya
3Department of Biological Sciences, Louisiana State University Baton Rouge, Baton Rouge, LA, USA
4International Maize and Wheat Improvement Centre (CIMMYT), Texcoco, Mexico

Correspondence
Amos Alakonya, International Maize and Wheat Improvement Centre (CIMMYT), Carretera Mexico-Veracruz KM. 45, El Batan, Texcoco, Mexico.
Email: a.alakonya@cgiar.org

Abstract
Maize lethal necrosis disease (MLND) is a devastating viral disease of maize caused by double infection with Maize chlorotic mottle virus (MCMV) and any one of the Potyviridae family members. Management of MLND requires effective resistance screening and surveillance tools. In this study, we report the use of small RNA (sRNA) profiling to detect MLND causal viruses and further the development of alternative detection markers for use in routine surveillance of the disease-causing viruses. Small RNAs (sRNAs) originating from five viruses namely MCMV, Sugarcane mosaic virus (SCMV), Maize streak virus (MSV), Maize-associated totivirus (MATV) and Maize yellow mosaic virus (MYMV) were assembled from infected maize samples collected from MLND hot spots in Kenya. The expression of the identified viral domains was further validated using quantitative real-time PCR. New markers for the detection of some of the MLND causal viruses were also developed from the highly expressed domains and used to detect the MLND-causative viruses in maize and alternative hosts. These findings further demonstrate the potential of using sRNAs especially from highly expressed viral motifs in the detection of MLND causal viruses. We report the validation of new sets of primers for use in detection of the most common MLND causal viruses MCMV and SCMV in East Africa.

KEYWORDS
maize chlorotic mottle virus, maize lethal necrosis disease, microRNA, small interfering RNA, small RNA sequencing, sugarcane mosaic virus

1 | INTRODUCTION

Maize is an important cereal crop whose global production was estimated at 1.134 tonnes per annum in 2017 (FAO, 2019). In Kenya, the average maize production was estimated to be 3.6 million tonnes in 2017, which is a drop from 4.25 million tonnes, the average for the five earlier years (FAO, 2018). The reduced production has been attributed to various biotic and abiotic challenges. Maize lethal necrosis disease (MLND), a viral disease first reported in Kenya in 2011, is one of the biotic factors associated with declining maize yields in Kenya (Fatma, Tileye, & Patrick, 2016; Mahuku et al., 2015).

In Kenya, yield losses worth USD 53.2, 180 and 198 million were reported in 2012, 2013 and 2014, respectively (De Groote, Oloo, Tongoinswattana, & Das, 2016; Marenya et al., 2018). The disease...
is caused by a synergistic response as a result of a double viral infection of the host plant by Maize chlorotic mottle virus (MCMV) and any member of Potyviridae family like Wheat streak mosaic virus (WSMV), Sugarcane mosaic virus (SCMV) or Maize dwarf mosaic virus (MDMV) (Kusia et al., 2015; Stewart, Quality, States, & Agricultural, 2017; Wangai et al., 2012). In Kenya, the disease was first reported in Bomet county and it has since spread to other regions of the country as well as neighbouring countries including Tanzania, Rwanda, Burundi, Ethiopia and Uganda (Mahuku et al., 2015; Wamaitha et al., 2018).

Currently, MLND management entails good agronomic practices such as weed and pest management, crop rotation, clean seed and the use of tolerant germplasm (Fatma et al., 2016; Redinbaugh & Zambrano, 2014). Breeding for tolerance through development of inbred lines is an ongoing strategy by the International Maize and Wheat Improvement Centre (CIMMYT) and other independent research teams (Beyene et al., 2017). This strategy is supported by studies involving genetic analysis of quantitative trait loci for resistance as well as genome-wide association studies aimed at identifying tolerant germplasm for breeding against MLND (Awata, Beyene, et al., 2019; Gowda et al., 2015; Gowda et al., 2018; Nyaga et al., 2019). Through these studies, various SNPs and loci markers for MLND tolerance have been validated through germplasm screening at a quarantine MLN screening facility in KALRO Naivasha, Kenya. Despite the existing strategies in MLND control, the disease is still a major threat to maize-growing areas in sub-Saharan Africa, where susceptible germplasm is still widely cultivated. Effective detection tools against MLND can form part of an integrated management package against MLND, especially during implementation of surveillance and quarantine measures. Several tools applied in MLND surveillance include double-antigen sandwich–enzyme-linked immunosorbent assay (DAS-ELISA) and real-time polymerase chain reaction (Fatma et al., 2016), but these could have limited success due to the viral divergence across regions (Braidwood, Müller, & Baulcombe, 2019) as well as low sensitivity in maize seeds due to low viral titres (Quito-Avila, Alvarez, & Mendoza, 2016). The use of next-generation sequencing (NGS) technologies in detection and characterization of MLND viruses has led to better understanding of these viruses (Braidwood et al., 2018; Xia et al., 2014, 2016, 2018). Recently, metagenomic analysis identified viruses that had never been associated with MLND which include mastrevirus, totiviruses and poleroviruses (Wamaitha et al., 2018). The sRNA sequencing technology is a novel high-throughput next-generation approach that has enabled the unravelling of the synergistic interactions between MCMV and Potyviridae members involved in MLND development (Mbega et al., 2016; Xia et al., 2014, 2016). Furthermore, virus-sourced small RNAs (vsiRNAs) derived from RNA interference (RNAi) mechanism exist in high levels in plants and can therefore be assembled into viral genomes hence providing insight into the viral agents infecting plant systems including the genetic variability (Braidwood et al., 2019; Burgyan & Havelda, 2011; Xia et al., 2018; Younis, Siddique, Kim, & Lim, 2014).

In this study, we report the use of sRNA sequencing in detection of viruses involved in the development of MLND in maize using samples from three MLND endemic areas in Kenya. Further, we used sRNA analysis pipelines to identify alternative markers for detection and characterization of MLND causal viruses and validated these using quantitative real-time PCR. The developed markers should be part of the more efficient tool kit for detection, monitoring and management of MLND-causing viruses in maize agroecosystems in Kenya and probably the wider east Africa region that seem to share the same strains of MLN causal viruses (Braidwood et al., 2019).

2 | MATERIALS AND METHODS

2.1 | Sample collection and preparation

Leaf samples from maize plants showing typical MLND symptoms (Wangai et al., 2012) were collected from Kericho, Bomet and Nyamira counties of Kenya according to sampling procedures described by Mahuku et al. (2015). Leaf samples were collected from twelve plants from each of the six different farms sampled per county. The leaf subsamples from every county were separately macerated and then pooled to form a county sample for further processing. The sample was then placed in RNase-free Eppendorf tubes containing RNA Shield reagent (Zymo Research). The samples were taken to the laboratory and kept at −80°C to await RNA extraction. The same procedure was adopted for the second round of sample collection for validation of markers using qPCR.

2.2 | RNA extraction and sequencing

Total RNA was extracted using Direct-zol™ RNA extraction kit (Zymo Research). Briefly, 50mg of leaf tissues was frozen in liquid nitrogen and ground into fine powder by vortexing with steel beads and the next steps of RNA extraction were done according to the manufacturers’ instructions (Zymo Research). RNA quality and concentration were determined using a NanoDrop spectrometer (Maestrogen Inc). RNA samples were shipped on dry ice to BGI Hong Kong (http://en.genomics.cn/) where library construction and small RNA sequencing were performed. Sequencing of libraries was done on BGISEQ-500 platform, a high-throughput sequencing technique based on combinatorial probe-anchor synthesis (cPAS) and DNA Nanoballs (DNB) technology (Huang et al., 2017).

2.3 | Sequence analysis and sRNA profiling

Small RNA sequence analysis was done using RNA-Seq standard pipelines. Raw sequence reads were retrieved from BGI server and filtered to remove low-quality reads at 5’ and 3’ ends, and reads without the insert tag were eliminated using cut-adapt (Martin, 2011). Reads shorter than 15 nucleotides were also discarded. The data criteria were set at 10% adapter and null rate, q20 of 90% and above,
and small RNA tag rate of less than 20%. Clean reads were then exported to the University of East Anglia (UEA) sRNA Workbench pipeline (Stocks et al., 2012) and host-derived miRNAs filtered through subtractive mapping to Zea mays miRBase and MirGeneDB databases. Further mapping of host-filtered reads to Zea mays genome was done using Bowtie 2 tool (Langmead, Trapnell, Pop, & Salzberg, 2009) incorporated in Geneious version 11.5 (Kearse et al., 2012).

Genome assembly of host-filtered sRNAs was performed using VirusDetect pipeline (Zheng et al., 2017) and contigs with a coverage >75% reported. The detection of viruses was done based on contigs generated by pooling those from reference-guided mapping of sRNA reads and those from de novo assembly. For each identified virus, the longest continuous contig was selected for phylogenetic analysis. These sequences were deposited at NCBI GenBank and were assigned accession numbers MK481075, MK481076 and MK481077 for SCMV, and MK491604, MK491605 and MK491606 for MCMV. Genome sequences from viral isolates reported from various parts of the world were retrieved from NCBI and used to generate the phylogenetic tree and to infer ancestry of the reported isolates using the Maximum Parsimony method in MEGA 7. For target prediction analysis, plant sRNA target analysis server (psRNATarget) (Dai, Zhuang, & Zhao, 2018) was used with default parameters.

Complete Refseq genomes for MCMV (NC_003627.1) and SCMV (NC_003398.1) were retrieved from National Centre for Biotechnology Information (NCBI) and used for mapping, annotation and identification of highly expressed domains of the viruses based on host-filtered sRNAs. Mapping of sRNA reads was done using Bowtie 2 within the Geneious version 11.5 followed by calculation of expression levels based on normalized reads per kilobase million (RPKM) with reference to mature peptide domains and CDS of SCMV and MCMV, respectively. This was done to avoid bias based on the length disparity with reference to the length of mature peptide domains (mtr-PD) or the CDS. For MCMV, seven CDS were targeted including those spanning ORFs P31, P50, P111, P7a, P7b, replicase-associated protein and P32. For SCMV, eight well-characterized mtr-PD including Nla Vpg, Nlb replicase, Nla-pro, HC-Pro, 6k2, P3, CI and 6k1 protein were targeted. Sequences for the highly expressed regions were retrieved and primers designed using Primer3Plus software (Untergasser et al., 2007).

### 2.4 Validation of alternative putative markers using real-time PCR

Total RNA from three representative leaf samples collected earlier in each region for sRNA sequencing was also used here for cDNA synthesis and validation of the identified markers using real-time PCR. Briefly, DNase-treated RNA was first quantified and 5 μg used for cDNA synthesis using FIREPol cDNA Synthesis Kit (Solis BioDyne,) according to the manufacturer’s instructions. Quantitative real-time PCR was carried out using the 5X HOT FIREPol EvaGreen® qPCR Mix Plus Kit in a reaction mix comprising 5X HOT FIREPol mix (1 μl), 0.5 μl of 0.25 μM for each primer (forward and reverse) and 1 μl of cDNA. The volume was topped up to 20 μl with RNase-free water. The reaction was carried out in a Roche LightCycler (https://lifescience.roche.com/) and included p7b, CI and p31 for MCMV and P3, CI peptide, Nla-pro and 6kI markers for SCMV. The qPCR conditions were an initial denaturation of 95ºC for 15 min, denaturation at 95ºC for 15 s, annealing in the range of 60–65ºC for 20 s (depending on the primers) and elongation step at 72ºC for 60 s. Nlb replicase and replicase CDS genes targeting genomic regions for SCMV and MCMV, respectively, were used as references for calculation of relative viral expression following Pfaffl (2004) 2dCT method. The real-time PCRs were done in triplicates for each sample. Standard curves were generated for each gene to confirm the efficiencies of PCRs.

### 2.5 Identification of alternate hosts for MCMV and SCMV using small RNA markers

We also investigated whether plants not previously reported to be MLND hosts and growing near or in infected maize fields could harbour MLND-causing viruses. We collected seventeen plant species (Figure S1) across different angiosperm orders and screened them via semiquantitative RT-PCR using MCMVP7b and SCVMVP3 markers identified in this study. Plant leaves were collected from fields in the aforementioned counties in August 2018, macerated and kept in Eppendorf tubes containing DNA/RNA Shield and transported to the laboratory for RNA extraction. The Global Positioning System (GPS) co-ordinates for each location where the samples were collected were captured.

RNA extraction and cDNA synthesis were done using Direct-zol kit (Zymo Research) and FIREPol cDNA Synthesis Kit (Solis BioDyne), respectively. A two-step RT-PCR was adopted for the detection of the viruses in the alternate host plants. Complementary DNA (cDNA) preparation was carried out using the FIREPol Kit (Solis BioDyne). One (1 μl) of the cDNA was used for RT-PCR in a 10 μl reaction volume consisting of 1 μl 5X FIREPol® PCR Master Mix and 0.5 μl of 0.25 μM of each primer. Primers from 2 (SCMV P3 and MCMV capsid protein) validated markers in sRNA-Seq and qPCR (Table S1) above were used alongside PPDK (FP 5’CGCGACGAATTAACAACGCT3’ and RP 5’ATCGTGTTGCTAGCGTCCAA3’) to confirm the success of library preparation. PCR conditions were an initial denaturation step of 95°C for 3 min, 35 cycles of denaturation at 95°C for 30 s, annealing at 50°C for 30 s (for both primers) and extension at 72°C for 1 min. PCR products were confirmed by gel electrophoresis.

### 3 RESULTS

#### 3.1 Subtractive mapping revealed five classes of microRNAs

Following a stringent quality check and cleaning of sequence reads, an average of 25 million clean reads per region was obtained. The reads ranged from 15 to 50 nucleotides long with a GC content of between 50% and 55%. Sequence reads ranging from 18 to 24...
nucleotides long accounted for the highest percentage of the sRNAs (Figure S2). On mapping the reads onto the *Zea mays* miRBase database, hairpin antisense, sense and mature antisense miRNAs were detected at extremely low abundance of 1.05%, 1.23% and 1.15% for Bomet, Kericho and Nyamira (Borabu), respectively. From the total mature sense miRNAs, five classes of microRNAs including zma-miR159a-3p, zma-miR168a-5p, zma-miR166a-3p, zma-miR167a-5p and zma-miR444a were highly expressed across the three regions as shown by the read per kilobase million (RPKM) values, and zma-miR159a-3p was the most frequently mapped miRNA (Table S2).
3.2 | Genome assembly and phylogenetic analysis of MLND-associated viruses

Genome assembly resulted in identification of five main families of viruses across the Bomet, Kericho and Nyamira counties. These included a Machlomovirus-MCMV, Polerovirus-MaYMV, Potyvirus-SCMV, Mastrevirus-MSV and totivirus-MATV. Full genome assembly for MCMV was successful across all the regions and effective for SCMV in Bomet and Nyamira counties (Figure 1). A complete MSV genome was also assembled from the pooled samples collected in Kericho. MaYMV and MATV genome assembly was generally poor and with no effective complete assembly in samples collected from all the three counties (Figure S3). Aligning the fully assembled genomes and a Maximum Parsimony tree generated from the alignment revealed evolutionary relationships for each virus with those reported across the world. For SCMV, the isolates in this study clustered close to those earlier reported from Elgeyo Marakwet, Embu and Bomet counties in Kenya. In general, SCMV isolates showed a high evolutionary divergence compared to those from across the globe (Figure 2). Similarly, the MCMV isolates showed a close relationship with those from across the world but also clustered together with those previously reported in Kenya.

3.3 | siRNA expression profile mapping to different viral domains in maize

To identify highly expressed viral domains in the isolates, we mapped a total of 3,046,799, 6,537,375 and 4,865,520 reads from Bomet, Nyamira and Kericho, respectively, onto the MCMV Refseq genome (accession id. NC_003627.1). There was variation in how the sRNA reads mapped across the entire MCMV genome (Table 1). The viral capsid protein, p7a and p7b CDS were the most frequently mapped domains followed by the replicase CDS, while p32 CDS domain

**FIGURE 2** Evolutionary relationships of whole-genome MCMV and SCMV isolates assembled in this study (marked with *) with those from different regions of the world or from the same region in a previous study. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates) [Colour figure can be viewed at wileyonlinelibrary.com]
showed moderate mapping (Figure 3a). We determined that the lowest sRNA expression mapping was to the replicase-associated peptide and p32 domains of the MCMV. For SCMV, a total of 1,474,576, 1,174,148 and 506,721 raw reads from Bomet, Nyamira and Kericho, respectively, were mapped (Table 2) to the SCMV Refseq genome (accession id. NC_003398.1). We noted that the *Potyviridae* ORF (PIPO) peptide-coding region, a low-frequency frameshift product and the capsid protein were the most frequently mapped domains in all the three regions. The Nlb replicase peptide was moderately mapped across the three regions with the Nla and 6K2 domains being the lowest mapped (Figure 3b).

### 3.4 | Validation of viral marker expression using quantitative real-time PCR analysis

To evaluate expression of the identified viral domains, we carried out real-time PCR on cDNA from the three replicates of the same tissues used for RNA sequencing experiments. A total of eight markers for highly expressed and conserved regions of both MCMV (capsid protein, P31 and p7b) and SCMV (P3, C1, p70 and Nla-pro) (Table S1) were used in the qRT-PCR. The location of the markers in the MCMV and SCMV genomes is shown in Figure S5 and Figure S6, respectively. The efficiencies of Nlb replicate and replicate reference genes were confirmed by 6-point, 10-fold dilutions of the template (Figure S4). Primer efficiencies were first confirmed using SCMV- and SCMV-positive samples (obtained from KALRO) through serial dilution before applying them in amplification (Table S3). The qRT-PCR analysis revealed a high relative expression of MCMV domains and one SCMV domain. For MCMV, the P7b together with all the three capsid protein regions is amplified with expression of log 7-fold relative to the replicase gene (Figure 4). In SCMV, the P3 domain showed the highest relative expression of log 3.5-fold relative to Nlb replicase gene. The rest of the regions tested recorded relative expression of \(<\text{log} 1.5\text{-fold}\) (Figure 4).

### 3.5 | Application of siRNA markers in identification of alternative hosts of MLND causal viruses

A total of 17 plant species were collected and screened for MLND causal viruses. It was observed that nine species including *Commelina benghalensis*, *Sacharum officinarium*, *Pennisetum purpureum*, *Panicum miliacium*, *Sorghum versicolor*, *Eleusine coracana*, *Cyperus rotundus*, *Eleusine Africana* and *Cynodon dactylon* showed amplification of the capsid protein gene sequence, a clear indication that MCMV was present in the tissues collected. Similarly, a 164 bp product of the P3 gene was amplified in six species including *Commelina benghalensis*, *Sacharum officinarium*, *Pennisetum purpureum*, *Panicum miliacium*, *Sorghum versicolor* and *Cyperus rotundus* confirming the presence of SCMV (Figure 5). *Sacharum officinarum* tested positive for both SCMV and MCMV and displayed typical MLND symptoms; however, all the other species were symptomless even when they tested positive for any of the MLND causal viruses. One interesting observation was that *Commelina benghalensis*, a herbaceous field weed, showed positive amplification for both SCMV and MCMV but did not show any MLND symptoms. To confirm further the possibility of capsid protein and p3 marker for MCMV and SCMV that amplified a 116 and 164 bp fragment, we mapped their sequences to *Zea mays* genome through NCBI and Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/) which did not return any results. This together with the positive controls proved conclusively that the bands were indeed products of the two viral markers.

### 3.6 | In silico applicability of host-derived miRNAs in detection of SCMV and MCMV

Through an in silico approach, three main miRNAs derived from the host mature miRNA (zma-miR167b-3p, zma-miR168b-3p and zma-miR528a-3p) were identified to target MCMV using psRNATarget
All miRNAs predicted target replicase (zma-miR167b-3p and zma-miR168b-3p) and replicase-associated protein (zma-miR528a-3p) domains of MCMV. The mechanisms of inhibition identified are cleavage and translation interference. For SCMV, fourteen host-derived miRNAs targeted HC-Pro, capsid protein, CI peptide, 6K2 and P3 CDS. Both translation and cleavage inhibition were identified with low unpaired energy requirement (Table S5). Six of the miRNAs possibly targeted the HC-Pro region, a viral suppressor protein that has been recognized for its inhibitory activity on RNA interference mechanism.
4 | DISCUSSION

The present study successfully describes detection of MCMV and SCMV, the main causal viruses of MLND in samples from three highly infected regions in Kenya through sRNA sequencing. We further report the development of markers based on the identified sRNAs which we use to successfully identify the MLND causal viruses in previously unreported alternate hosts. These results buttress the key role of sRNAs in identification of these viruses. A majority of RNA molecules recovered after sequencing had lengths of 21–22 nt which represented the large body of small RNA that includes microRNAs (miRNAs), small interfering RNA (siRNAs), Piwi-interacting RNAs (piRNAs) and viral-sourced small RNAs (vsiRNAs) that originate from different plant response pathways mainly post-transcriptional gene silencing (PTGS) as well as aiding viral replication in host (Iqbal et al., 2017; Lam et al., 2015; Saito, 2013).

### TABLE 2  sRNA classification and abundance to SCMV coding regions

<table>
<thead>
<tr>
<th>Viral coding regions</th>
<th>Raw Read Count</th>
<th>Read per Million Kilobase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bomet</td>
<td>Nyamira</td>
</tr>
<tr>
<td></td>
<td>Bomet</td>
<td>Nyamira</td>
</tr>
<tr>
<td>Capsid protein peptide</td>
<td>196,634</td>
<td>122,657</td>
</tr>
<tr>
<td>NLb replicase peptide</td>
<td>128,027.2</td>
<td>85,660</td>
</tr>
<tr>
<td>Nla-Pro protein peptide</td>
<td>19,975.67</td>
<td>26,269.5</td>
</tr>
<tr>
<td>Nla-Vpg protein peptide</td>
<td>49,710.17</td>
<td>38,313.5</td>
</tr>
<tr>
<td>6K2 protein peptide</td>
<td>66.17</td>
<td>7,279.33</td>
</tr>
<tr>
<td>C1 protein peptide</td>
<td>110,334.5</td>
<td>112,840.3</td>
</tr>
<tr>
<td>PIPO CDS</td>
<td>62,353.5</td>
<td>53,605</td>
</tr>
<tr>
<td>Polyprotein CDS</td>
<td>907,474.8</td>
<td>727,524.3</td>
</tr>
</tbody>
</table>

Note: Data are from one pooled sample per county.

A sample comprised of a total of 48 plants sampled as twelve symptomatic plants from each of the six different farms per county (Kericho, Bomet and Nyamira).

**FIGURE 4** Expression profiles of viral domains used as markers targeting regions of MCMV and SCMV using qRT-PCR. MCMV markers—MCMV p7b, MCMV coat 245, MCMVCP116, MCMV coat 139 and MCMV 113—were compared across three sampled counties including Bomet, Nyamira and Kericho. Colour codes range from green-yellow-red with green representing low and red for high, log expression relative to replicase gene (reference gene). Four markers including SCMV P3, SCMV Nla79, SCMV 6KI and SCMV C1 were tested for SCMV with the colour codes assigned as in MCMV expression. The values presented are log of relative expression (Log 2^ΔCt) [Colour figure can be viewed at wileyonlinelibrary.com]
A low percentage of miRNAs was mapped to the annotated miRNA database of maize. Of the five identified miRNAs sourced from the host, *zma-miR159a-3p* was the only microRNA that has been previously identified to target *MCMV* virus. sRNA target analysis for all the identified host-derived miRNAs by plant small RNA target (psRNATarget) analysis demonstrated an excellent target site accessibility based on the observed low unpaired energy (UPE) required to open the secondary structure around the target site. For *MCMV*, only three miRNAs were identified with replicase and replicase-associated domains being the only identified targets. However, *SCMV* had 14 miRNAs all having UPE values of −1. Several regions of *SCMV* including 6K2, CI, HC-Pro, capsid protein and P3 are potential targets of the 14 host-derived miRNAs. The relative abundance of all the identified miRNAs with the exception of *zma-miR159a-3p* was very low, and detection of these species for inferring the presence of virus may not be efficient since *zma-miR159a-3p*, the only highly expressed miRNA from the host, is associated with several other physiological activities including co-modulation of *myb74* in the plant cell (Wang et al., 2019). Despite the observation by Xia et al. (2019) that *miR159*, *miR393* and *miR394* may be involved in antiviral defence mechanisms against co-infecting viruses, the same study noted that these miRNAs are down-regulated. This, coupled with the diverse roles of these miRNAs and their low detection level in this study, further ruled out the possibility of using them as markers for detecting MLND viruses. From our in silico analysis, there was an indication that these miRNAs could be playing a role in the *MCMV* and *SCMV* interaction that included possible binding to viral motifs including HC-Pro. Nevertheless, there was little or no chance that further analysis of this interaction could yield any diagnostic tool. As is the case of this study, past studies have independently either inferred based purely on in silico data especially for predictive purposes (Iqbal et al., 2017) or in some instances combined both in silico and wet-lab approaches to obtain more confirmatory results where the prediction was more definitive (Xia et al., 2016).

We fully assembled the *MCMV* genome but only managed to assemble 95% of *SCMV* across the three regions. This is evidence that the typical MLND symptoms observed during sample collection were due to co-infection by *MCMV* and *SCMV*. *SCMV*, the most reported potyvirus in MLND development, encodes protein HC-pro which has been previously identified as a viral suppressor of RNA silencing (VSR) in the maize immune system. The protein also plays a role in plant-to-plant transmission of potyviruses and maturation of the polyprotein (Valli, Gallo, Rodamilans, López-Moya, & García, 2018). The ability of VSR proteins to suppress the plant’s immunity could have led to upregulation of vsiRNAs associated with *MCMV* enabling complete assembly of the viral genome. Xia et al. (2016) observed that there is a synergistic interaction between the two viruses where the potyvirus-*SCMV* promotes the expression of the co-infecting *MCMV*. Similarly, the study by Awata, Ifie, et al. (2019) determined that during co-infections, there was an increase in the levels of *MCMV* while the concentration of *SCMV* remained the same as in singly infected maize plants. This led to the conclusion that *MCMV* is the main cause of the MLND. The comparatively low *SCMV* viral concentrations are attributed to a weakness in *SCMV* proteins such as P1 and VPg that are supposed to promote viral replication and movement (Xia et al., 2016).
Our analysis also identified Maize streak virus (MSV), a mastrevirus that has previously been reported in other studies. Wamaitha et al. (2018) detected the presence of this virus in MLND-positive maize samples in Kenya through a metagenomics approach. Our results and those of Wamaitha et al. (2018) together corroborate that MSV could in some way be a contributor to the development of MLND. This is because the development of MLND is associated with co-infection and VSR proteins are produced by SCMV. To date, no viral suppressors have been identified for MSV. A similar study by Xu et al. (2017) on Wheat dwarf virus (belonging to the same family as MSV) indicated that it encodes a replicase (Rep) protein that acts as a VSR of PTGS. There is a possibility therefore that MSV has a VSR-like activity that contributes to the development of MLND although this needs to be investigated further. Other viruses including MYMV were also detected although their assembly was poor. This could be attributed partly to the low depth of sequencing (Zheng et al., 2017) or the low complexity characterized by repeat sequences in the genome of these viruses which leads to generation of misassembled contigs. As reported by Claros et al. (2012), repeat sequences are difficult to assemble as high-identity reads could come from different portions of the genome, generating gaps, ambiguities and collapses in alignment and assembly. The presence of these viruses in infected plant material could, however, be contributing to development of the disease since they have been reported to code for a PO protein that inhibits both local and systemic RNA silencing (Chen et al., 2016).

Phylogenetic analysis of SCMV isolates under the current study showed a divergence with other isolates from the database across the world. It was, however, evident that our isolates clustered closely with others previously reported in Kenya and neighbouring countries. It is possible that our SCMV isolate(s) could be the same one (or a variant) of those previously reported in the region although this was not confirmed. Wamaitha et al. (2018) reported genetically distinct groups of this virus in Kenya using single nucleotide polymorphisms. Recently, findings by Braidwood et al. (2019) on recombination events across the genome of SCMV in Ethiopia and Kenya corroborate these observations but identified the P3 as being conserved domains. In summary, this study determined that due to variation across the SCMV genome, it is difficult to infer variability from one domain of the virus. Equally, Nigam, LaTourrette, Souza, and Garcia-Ruiz (2019) demonstrated that potyviruses contain fixed hypervariable areas in key parts of the genome that provide mutational sturdiness and may be involved in host adaptation. These hypervariables in the SCMV genome like many other potyviruses will continue to be a challenge in the development of robust detection tools for the virus which is a major contributor to MLN in Africa. Because the stable food crops in Africa are also hosts of MLN causal viruses and the continent is yet to be self-reliant in terms of food, there is still an imminent risk of introduction of other SCMV isolates from other regions through importation of food and feed.

In contrast, all the MCMV isolates clustered together indicating a slight evolutionary divergence. The MCMV identified in this study clustered together with those previously isolated from Bomet and KALRO demonstrating a relatively high genetic stability (Wamaitha et al., 2018). Previous studies have described the low evolution and divergence in MCMV strains. Mapping of reads onto respective viral genomes enabled identification of a number of domains and revealed their expression profiles. For MCMV, capsid protein and the two p7 domains namely p7a and p7b were the most frequently mapped. These domains are all encoded by subgenomic RNA (sgRNA) with capsid protein being expressed at the 3’-proximal end of the sgRNA (Lommel, Kendall, Xiong, & Nutter, 1991). Capsid proteins of various viruses have been associated with several functions including playing a role in translation of viral RNA, suppression of RNA silencing, virulence during viral infection and dictation of specificity of vectors for virus transmission (Bol, 2008). A well-characterized role of MCMV’s capsid protein is its terminal-encoded amino acids which allow subcellular localizations of MCMV (Zhan, Lang, Zhou, & Fan, 2016). For SCMV, the P3 protein peptide, 6K2 and Na protein regions were frequently mapped. The SCMV genome encodes a single, large polypeptide encoding ten proteins which are cleaved by self-encoded proteases (Akbar, Tahir, Wang, & Liu, 2017; Mbega et al., 2016). P3 is a movement protein that plays a role in enlargement of the pore size of plasmodesmata and translocation of the viral genetic material into the neighbouring cell thus promoting local and systemic distribution of the virus within the host (Cabanas, Watanabe, Higashi, & Bressan, 2013). The protein domain 6K2 has been associated with replication, while Nla-pro, a small nuclear inclusion protein, is the main protease involved in cleavage of the polypeptide into individual proteins (Van der Vlugt, 1994). We observed variation in mapping of abundance of siRNA across regions as well as viral genome. The variation across the different locations sampled was possibly due to the disparities in symptom stages of the MLND-infected plants that were sampled. Differences in symptoms could be an indication of varied stages of infection or possibly different host responses. Since our interest was to capture siRNAs at all stages of the disease progression, there is a high possibility that the concentration of viral-sourced miRNAs will vary at different stages of infection/disease progression in the pooled subsamples. It also provided an opportunity to capture whole host responses in one sample. Nevertheless, this variation does not compromise the results of this study since a diagnostic tool should be able to detect the pathogen at the different stages of infection. Further, the observed variations across viral genome domains may have resulted because DICER-like proteins (DCLs) may have a similar but slightly different targeting preference towards viral genome regions. DCLs are responsible for not only biogenesis of vsiRNAs but also their amplification and hence accumulation (Xia et al., 2014). Xia et al. (2016) have illustrated how viral infections differentially modified expression of the different factors involved in antiviral RNA silencing mechanisms.

Expression levels resulting from the mapping statistics were further validated via qRT-PCR using p3 for SCMV and p7b for MCMV. The two markers were selected for use in identification of these viruses in plants not previously reported to host the MLND-causing viruses. The two markers target the most conserved and expressed domains of MLND viruses as confirmed through mapping and qPCR. MCMV and SCMV have previously been reported to infect members of the grass family including millet and sorghum
Amos Alakonya

REFERENCES


CONFLICT OF INTERESTS

The authors have no conflict of interest to declare.

ORCID

Ronald Tonui https://orcid.org/0000-0001-7540-6935

Amos Alakonya https://orcid.org/0000-0001-7710-499X