DEVELOPMENT OF METHODS IN SCREENING FOR TOLERANCE AGAINST MAIZE LETHAL NECROSIS DISEASE (MNLD) AND DETECTION OF MLND CAUSAL VIRUSES

RONALD TONUI

MASTER OF SCIENCE

(Molecular Biology and Biotechnology)

PAN AFRICAN UNIVERSITY
INSTITUTE FOR BASIC SCIENCES, TECHNOLOGY AND INNOVATION

2018

DEVELOPMENT OF METHODS IN SCREENING FOR TOLERANCE AGAINST MAIZE LETHAL NECROSIS DISEASE (MNLD) AND DETECTION OF MLND CAUSAL VIRUSES

RONALD TONUI

A Thesis submitted to the Pan African University Institute of Science, Technology and Innovation in partial fulfillment of the requirements for the award of the degree of Master of Science in Molecular Biology and Biotechnology of the Pan African University

2018

DECLARATION

I, the undersigned, declare that this is my original work and has not been submitted to any other college, institution or university for academic credit.

Signed: .... Date: ....

Tonui Ronald

Reg. No. MB300-0008/17

This thesis has been submitted with our approval as the University supervisors.

Signed: .... Date: ....

Dr. Remmy Kasili

Institute for Biotechnology Research-Jomo Kenyatta University of Agriculture and Technology

Signed: .... Date: ....

Dr. Steven Runo

Biochemistry department -Kenyatta university

DEDICATION

To my dear Mum, Mrs. Emily C. Koskei. Your support, encouragement and love have invaluably guided me throughout my academic endeavors. Thank you, Dear Mum.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF PLATES

ABBREVIATIONS AND ACRONYMS

cDNA   Complementary deoxyribonucleic acid

DAS-ELISA   Double antibody sandwich- enzyme linked immunosorbent

                          assay

DCL   Dicer-Like Proteins

FAO   Food and Agriculture Organization

MCMV   Maize Chlorortic Mottle Viruses

MLN   Maize Lethal Necrosis

MLNV   Maize lethal necrosis viruses

MLND   Maize lethal necrosis disease

MSV   Maize streak virus

PBST   Phosphate buffered saline with Tween 20

PTGS   Post transcriptional gene silencing

Rgs-CaM-   calmodulin-like protein, regulator of gene silencing

SCMV   Sugarcane Mosaic Virus MiRNA-MicroRNA

siRNA   small interfering RNA

VsiRNA   Viral suppressor proteins

VSR   Viral suppressor protein

ABSTRACT

Maize lethal necrosis disease (MLND), a viral disease resulting from double infection by Maize Chlorotic Mottle Virus (MCMV) and Sugarcane Mosaic Virus (SCMV) is the most devastating viral disease of maize which has impacted negatively to maize production in Kenya. This study aimed to test the prospect of using small RNA transcriptome profiling in detection of MLND causal viruses and subsequent development of markers for future use in detecting the viruses. Next generation sequencing of small RNAs (sRNASeq) from leaf samples of symptomatic plants collected from Bomet, Nyamira and Kericho counties was done using BGIseq500 sequencing platform and mapping and differential expression analysis of the resulting reads done using available pipelines with reference to viral and host genomes. Five viruses including maize chlorotic mottle virus, sugarcane mosaic virus, maize streak virus, maize associated totivirus and maize yellow mosaic virus were identified through sRNAseq. Capsid protein and P7b movement protein domains of MCMV and P3 domain in SCMV were identified as the highly expressed domains and were used subsequently in design of primers for development of a diagnostic kit and in identification of the alternate host. These markers were used to identify non-maize hosts for the viruses from plants collected in fields infested with the disease. Here, sugar cane (*Saccharum officinarum*), Napier grass (*Pennisetum purpureum*) Proso millet (*Panicum miliaceum*), Sorghum (*Sorghum versicolor*), Finger millet (*Eleusine coracana*) and Wandering Jew (*Commelina benghalensis*) all tested positive for SCMV and MCMV. Further, the study aimed to test the applicability of detached leaf assay (DLA) as a pre-screening technique for tolerance against MLND. Three MLND-tolerant maize genotypes (CHMLND0093, CKIR11027, CKIR12032) from International Maize and Wheat Improvement Center (CIMMYT) and susceptible genotypes Namba nane and Hybrid 144 from Kenya Agricultural and Livestock Research Organization (KALRO) were used for the study. Agar media supplemented with 3% sucrose, 20mg/l gibberellic acid and 10mg/l kinetin was determined as the best media regime in retaining the green color and was subsequently used in assaying for the development of MLND symptoms on detached and artificially infected maize leaves. Detached leaf discs at 28, 42 and 63-day after planting

were infected with an inocula consisting of MCMV and SCMV and maintained on this media and scored for disease severity for 19 consecutive days using CIMMYT MLND scoring scale. Detached leaf assay from plants at 28th day of growth was able to distinct different genotypes as either tolerant or susceptible with tolerance increasing with plant growth age. Overall, the results of this study showed the potential of small RNAs for use in MLND diagnosis as well as in development of markers for future detection and further revealed that MLND causal viruses are co-hosted by majority of grass family members. The potential of DLA as a pre-screening technique was demonstrated and could accelerate the identification of promising germplasm for use as candidates in development and screening for tolerant germplasm in MLND breeding programs prior to extensive field testing.

CHAPTER ONE

INTRODUCTION

1.1. Background to the study

Cereals including maize, sorghum, millet, rice and wheat are the major contributors to food security globally. In Kenya, maize, a member of Poaceaee family is the most cultivated cereal with its production averaged at 3.604 million tonnes for 2016 compared to 0.316 million tonnes for wheat which is the second cereal after rice in terms of global production (FAO, 2017). Globally, maize represents over one-third of world cereal output (Zhang *et al.,* 2012). The plant is relied upon in several parts of the world as a source of food, industrial raw material for manufacture of various industrial metabolites such as alcohol and as feed for animals. In Kenya, maize is a staple food crop and is produced by more than 98% of 3.5 million small- scale farmers (Kirimi *et al*. 2011).

Despite the immense importance as a staple food, maize production in Kenya faces many challenges ranging from environmental factors such as drought and salinity to parasitic weeds like striga as well as diseases (Andersson and Halvarsson, 2011; FAO, 2017). This is evident by decline in production with aggregate production in 2017 estimated by FAO to be 3.6 million tonnes, an 8% drop from the previous year and 18% below the average of the previous five years (FAO, 2017).

Maize diseases cut across vast pathogenic families including fungi, bacteria and viruses. The common pathogenic fungi of maize includes *Ustilago zeae* (smut) and *Puccinia polysora* (maize rust) (Pataky and Snetselaar 2006) as well as bacterial diseases such as

1

bacterial wilt caused by *Pantoea* sp. (Degani *et al*. 2014). Pests ranging from animals to insects have also contributed to the losses in maize production with fall armyworms being the current greatest threat in Kenya (FAO, 2018b; Kumela *et al*. 2018). These pests coupled with viral diseases and unfavorable climatic conditions greatly contribute to the decline in maize production.

Bomet, Kericho and Nyamira counties are among the arable regions in Kenya with reliable rainfall and loamy soil and are among the hotbeds for maize production. The three regions are currently facing the most devastating effects of maize lethal necrosis disease (MLND). The disease was first reported in 2011 in Bomet region but has since spread to several other regions in Kenya and to neighboring countries including Tanzania, Burundi, Rwanda and Uganda (Mahuku *et al*., 2015; Wangai *et al*., 2012). Maize Chlorotic Mottle Virus (MCMV) is the main virus in MLND development with the symptoms amplified in the presence of any member of the *Potyviridae* family. The members of *Potyviridae* include Maize Dwarf Mosaic Virus (MDMV), Wheat Streak Mosaic Virus (WSMV) and Sugarcane Mosaic Virus (SCMV) which is the predominant member of potyvirus in Kenya and East Africa (Mahuku *et al*., 2015).

MLND is characterized by chlorotic mottling of leaves leading to premature plant death, stunted growth, malformation in maize ears, rotting cobs and failure to tassel as well as sterility in male plants (Uyemoto *et al*., 1980). Depending on the host susceptibility and the age, the symptoms progress until the plant dies off (Kusia *et al*., 2015). Losses associated with the disease range from 50-70% in mild cases and up to a 100% in severe cases. The causative viruses are effectively spread through seeds (0.33%), insect vectors such as aphids and through mechanical means due to plant injuries (Mahuku *et al*., 2015).

Soil has also been implicated in the spread of the disease with wet/moist conditions increasing the potential of soil to spread of these viruses (Mahuku *et al*., 2015). Maize plants are susceptible to MLND at all stages of growth with the component viruses being transferred from one plant to another and from one field to another by insect vectors (Nault *et al*., 1978).

Strategies currently employed in the management of MLND include agronomic practices such as weed and pest management, crop rotation, use of tolerant germplasm and proper screening of planting materials (Fatma *et al*., 2016; Redinbaugh and Zambrano, 2014). Identification of tolerant germplasm and breeding for tolerance and/or resistance is the most viable remedy though there has not been any immunity reported. All tropical and temperate varieties show high levels of susceptibility (Nelson *et al*., 2011). Proper screening of planting materials including seeds is also a possible remedy with several techniques ranging from double-antigen sandwich-enzyme linked immunosorbent assay (DAS-ELISA) and Reverse transcription polymerase chain reaction (RT-PCR) being employed (Fatma *et al*., 2016). As a control strategy, screening of planting materials faces challenges as the existing methods are not effective due to the low levels of viruses in seeds (Quito-Avila *et al*., 2016).

Sequencing approaches including Sanger and next generation sequencing offers more reliable and sensitive approaches in detection of plant viruses (Fatma *et al.,* 2016). The recovery of RNA molecules associated with the genomes of MLND causing viruses as well as library construction of the whole genomes is however an impediment to detection of these viruses using conventional sequencing platforms. Small RNA sequencing (sRNASeq) is the latest next generation sequencing platform which counteracts the

limitation of requirement for complete genomic library construction prior to sequencing and detection. Besides using small RNA molecules of 15-50 nucleotides base pairs long which are extremely staple, sRNASeq targets not only the products of degradome but also those of RNA-interference pathways (Illumina, 2013). This makes sRNASeq platform a more reliable detection approach especially when the causal viruses exist in low titres (Guimerase *et al.,* 2015). With the clear need to develop efficient, highly sensitive techniques for the detection of MLND causal viruses as a control strategy, this study seeks to use small RNA sequencing both in detection of the viruses and in development of small RNA markers for use in development of a diagnostic kit as well as in identification of potential alternate host for MCMV and SCMV. The study further seeks to optimize detached leaf assay technique in screening for tolerance in maize germplasm against MLND.

1.2. Statement of the problem

MLND caused maize crop loss of up to 90% in 2012 alone in Kenya. This led to an estimated grain loss of 126,000 tons valued at USD 52 million in Kenya alone that resulted in decreased food security (Mahuku *et al*., 2015). The threat increased to USD 198 million in 2014. Even though some maize varieties were reported by Nelson *et al.,* (2012) to have shown some resistance to MLND, no immunity has been observed. Additionally, almost all temperate climate inbred lines and hybrids are highly susceptible to the viruses (Nelson *et al.,* 2012). While the most feasible remedy to MLND is breeding of resistance coupled with their cultivation, resistant cultivars are yet to be identified. The screening process currently relies on artificial inoculation of viruses to

4

whole plants in the field. This strategy faces challenges as it is dependent on weather conditions, requires large fields for establishment of test plants and it is labour intensive when screening bulk germplasm.

Detection of the causative viruses currently relies on enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) and genome sequencing (Fatma *et al.,* 2016). Even though these techniques are reliable, they are limiting when it comes to the molecular interactions between the viruses and their hosts. The existence of these viruses in low titers in the seeds is also an impediment to detection by these techniques. The cost of equipment required for application of ELISA, PCR and genome sequencing for the identification of positive and negative samples is an additional challenge to the use of these techniques.

Several extenuation measures including agronomic practices such as weed management and crop rotation have been recommended and applied for the control of the disease (Mahuku *et al*., 2015). Despite this, the disease has persisted suggesting the existence of natural alternate hosts which may be highly immune to these viruses and do not show symptoms but harbor these viruses and serve as source of inoculum for subsequent maize generations.

This study analysed small RNA transcriptome profiles of maize plants presenting symptoms of MLND for identification of present viruses and develop markers for detection of SCMV and MCMV. These markers were further used in identification of potential alternate hosts for MCMV and SCMV. It was the object of the study to optimize detached leaf assay in screening for resistance to MLND by various maize germplasm.

5

1.3.   Justification and significance of the study

Detached leaf assay (DLA) entails maintenance of excised leaves on an agar plate to ensure their life sustenance while analyzing the impacts on the leave by introducing pathogens. Several studies for resistance against viruses based on DLA have been carried out (Mahuku *et al.*, 2015; Stobbs *et al.*, 2015).The approach was yet to be applied to screening for MLND tolerance in maize. The study demonstrated the possible applicability of the same test as a pre-screening technique in analysis of tolerance against MLN with optimization based on senescence delaying hormones, nutrient media and the stage of plant growth. This technique once adopted will aid in reducing the bulk of germplasm to be screened as it will allow pre-selection of promising cultivars before screen house and field analysis through whole plant inoculations.

Small Ribonucleic Acids (sRNAs) including siRNAs, vsiRNAs and viral mRNA can serve the purpose of detection of MLND causal viruses and further provide information on the potential interaction between the causal viruses and the host. Transcriptomic profiles generated from the infected plants will further aid in identification of potential markers for resistance against MLN viruses by the host plants. This is based on the finding that infected plants have been shown to express high levels of siRNA molecules to target viruses by interacting with argonaut protein of the RNA-Induced Silencing Complex (RISC) system (Gowda *et al.*, 2015).

Despite the existing crop rotation measures in maize growing regions, maize lethal necrosis disease still persists in areas where the disease has been reported. There is a possibility that alternate hosts may exist that harbor one or both causative viruses.

There is a need to identify alternate hosts for the MLND causal viruses in Kenya and screening for the presence of traces of these viruses will aid in guiding the farmers on management of weeds and other crops grown in close proximity to maize.

1.4. Research hypotheses

1. Small RNA sequencing cannot be applied in the identification of MLND causal viruses as well as in development of markers for their detection,

2. DLA technique is not applicable in screening maize germplasm for tolerance to MLN disease.

3. Small RNA markers cannot be used in the identification of Maize lethal necrosis-causing viruses in alternate hosts.

1.5. General objective

To develop methods for screening for tolerance against maize lethal necrosis disease (MNLD) and for detection of MLND causal viruses.

1.6. Specific objectives

1. To analyze sRNA transcriptome profiles of MLND infected maize for identification of potential markers for detection of the causal viruses and develop markers for detection of MCMV and SCMV isolates in Kenya.

2. To optimize the detached leaf assay for screening maize germplasm against MLND.

3. To identify alternate hosts of MLND viruses using sRNA biomarkers.

1.7. Scope of the study

Sample collection was done in three counties in Kenya namely Bomet, Kericho and Nyamira. The study began with collection of leaves of infected plants in the mentioned regions from 15th February 2018 followed by sequencing at Beijing Genomics Institute-Hong Kong. For DLA, leaves from three lines of maize plants which are tolerant namely E1-SS-15-01/CHMLND0093, E2-SS-20-25/CKIR11027 and E3-SS-15-01/CKIR112032 and two susceptible varieties including *Namba nane* and H144 were used to test the applicability of the technique in screening germplasm for resistance to MLND. Germination of seeds began on 27th January 2018 and first round of screening was done on from 2nd March 2018 and the repeat experiment was carried out between May 21st 2018 to 2nd August 2018. Samples for alternate host identification and validation of sRNA markers were collected on 10th august 2018 from the previously mentioned

regions (Bomet, Nyamira and Kericho) with emphasis on Poaceae and leguminae family members growing in close proximity with plants infected with MLN. This study was limited geographically to Kericho, Nyamira and Bomet counties and samples from different locations within a given county were pooled for sequencing thus individual locale profiles may have been lost.

CHAPTER TWO

LITERATURE REVIEW

2.1. Maize growth and production in Kenya

Maize belongs to the grass family (Poaceae) and is among the oldest, most cultivated cereal crops that originated from the Mexican highlands (Dallwitz, 1992). It is grouped among the most important grass family members after rice and wheat; with the current world production of maize being estimated at 594 million tons from about 139 million ha (FAO, 2017). The immense popularity of maize is derived from its ability to adopt to a broad range of environmental conditions including well-drained, nutrient-rich soils with pH ranging from 5.5 to 7.0; temperature of 14 to 16°C and rainfall from as low as 250mm to 5000mm per annum (Beckingham and NSW, 2007). Depending on the variety, maize grows up to 4M high with alternate, parallel venated leaves and with prop and fibrous roots. Maize is used in diverse ways with the main production being grain crop for livestock feed, food processing such as corn chips, grits, breakfast cereals and flour, popcorn and industrial starch (Farrell and Stecyk, 2007). Green sever and ensiled maize are also commonly used as supplementary feed in dairy production and also as feed for other animals, including goats, chicken, horses and goats (Farrell and Stecyk, 2007; Roche *et al*., 1996). Maize is also a major feedstock for bioethanol (McCutcheon, 2007).

In Kenya, majority of producers are small-scale farmers cultivating the crop for food and a small percentage for commercial purposes. Its production is averaged at 3.6 million tonnes, way above other cereal crops like wheat whose production was 0.316 million

tons and that of rice which averaged at 0.164 million tons (FAO, 2018a). It is the main staple food for over 85% of the Kenyan population. The average per capita cnsumption is 98 kilograms translating to at least 2.7million metric tonnes, per year (Nyoro *et al*. 2004). Over 38% of the farmers who grow food crops in Kenya grow maize with majority (about 70%) being small scale farmers and the remaining 30% being large scale commercial producers (Export Processing Zone Authority, 2005). The small-scale farmers mainly cultivate the crop for subsistence with up to 58% being for household consumption (Sibhatu and Qaim, 2017). Maize provides carbohydrates in form of starch and is a source of vitamins such as vitamin A, B, C and E, iron, protein and fiber.

2.2. Constraints to maize production in Kenya

Maize cultivation faces a lot of challenges from both abiotic and biotic factors (Figure 2.1). Climate change threats including drought, hailstorms and floods as well as soil factors such as salinity, acidity and poor fertility are among the major abiotic stresses in maize production especially in the tropics (Kiptanui 2013). Maize is classified as a salt-sensitive plant with very minimal tolerance to salt (RL *et al*., 2000). Acidity, which results from excessive application of nitrogenous and phosphate fertilizers has also been identified as a challenge in maize production (Kiptanui, 2013; Ngoune *et al*., 2018). Lack of certified germplasm or poor adoption of improved varieties, also affects maize production in terms of yield (Langyintuo *et al*., 2010).

Besides abiotic challenges, maize faces vast biotic threats which range from microbial pathogens to pests in higher animal kingdoms (Figure 2.1). Fungal diseases identified as of great economic importance include Common rust (*Puccinia sorghi*), Stalk and

11

ear rots (*Diplodia* spp., *Fusarium* spp. *and Aspergillus* spp.), Corn leaf blight (*Bipolaris maydis*), northern corn leaf blight (*Exserohilum turcicum*), Southern rust (*Puccinia polysora*) and gray leaf spot (*Cercospora* species) (Frenken, 2013; Patancheru and Bengtson, 1980; Tamra Jackson-Ziems, 2016). Some bacterial strains including those causing bacterial streak disease (*Xanthomonas oryzae* pv. *Oryzicola)* and bacterial stripe of maize *(Burkholderia andropogonis)* infect maize even though the losses caused by bacteria are not as huge as those caused by the fungi (Zhao *et al*., 2005).

Currently, viral diseases are the main threat to maize production in Kenya. Fall army worm (*Spodoptera exempta),* an insect pest of maize is reported to have caused an average loss of 6-8 million bags in 2017 (FAO, 2018b). Various strategies including the use of pesticides and tolerant varieties have been used as mitigation measures against the pests (syngenta, 2018). Viral diseases are among the major threat to maize production in Kenya with maize streak virus, sugarcane mosaic virus as well as maize dwarf mosaic virus having highest economic impacts. Latest viral disease of great economic importance to maize production is Maize Lethal Necrosis disease (MLND) caused by a co-infection of a single host by two distinct viruses namely Maize Chlorotic Mottle Virus (MCMV) and Sugarcane Mosaic Virus (SCMV). The disease is relatively new in Africa with its loss averaged at USD 53.2million in 2012 which increased to USD 198 million in 2014 (N. Li *et al.* 2015).

North America
Undernourishment <5%

Maize (372.6MM t/year)	• Gray leaf spot • Monoculture
Soybean (112.9)	• Soybean rusts • Cyst nematode • Drought
Wheat (84.4)	• Wheat stripe & stem rust • Soil degradation

Central America
Undernourishment <5–15%

Sugarcane (111.9)	• Stalk feeders • Manual harvest
Maize (27.1)	• MDMV • MRFV • Drought
Bananas (9.1)	• Tropical race 4

South America
Undernourishment <5–15%

Sugarcane (840.8)	• Giant cane borer
Soybean (156.3)	• Asiatic soybean rust • Deforestation
Maize (126.2)	• Corn stunt virus diseases • Drought

North Africa
Undernourishment <5%

Sugarcane (22.1)	• Mosaic virus • Manual harvets • Water
Wheat (19.0)	• Fungal diseases • Rust
Sugar beet (14.3)	• Leaf blotch • Limited arable land

West Africa
Undernourishment 5–25%

Cassava (29.8)	• Brown streak • Short shelf life
Bananas (5.8)	• Xanthomonas wilt
Plantains (5.5)	• Black sigatoka disease

Undernourishment <5%

Wheat (249.3)	• Stem rust
Sugar beet (191.7)	• Aphids • Yellow virus • Rhizomania
Maize (128.9)	• Fungal diseases • Corn border • Drought

South Africa
Undernourishment <5%

Sugarcane (23.3)	Thrips Stemborer
Maize (14.5)	• Armyworm • Drought
Potatoes (2.4)	• Southern wilt • Water shortage

West Asia
Undernourishment <5%

Wheat (28.9)	• Yellow & stem rust • Drought
Sugar beet (17.1)	• Curly top virus • Drought
Tomatoes (16.6)	• Yellow leaf curl virus

East Africa
Undernourishment <25%

Sugarcane (39.2)	• Mosaic virus • Water
Maize (30.7)	• MLND • Water
Cassava (28.8)	• Brown streak • Short shelf life

Undernourishment <5%

Wheat (22.6)	• Yellow rust • Stem rust • Iron deficiency
Potatoes (8.4)	• Late blight • Potato beetle • Water shortage
Seed cotton (4.8)	• Vascular wilt

South-East Asia
Undernourishment 5–25%

Oil palm (235.6)	• BSR disease • deforstation
Rice (210.0)	• Weeds • Decline of productivity
Sugarcane (190.6)	• SCGS • SCWL • Water

Australia & Oceania
Undernourishment <5%

Sugarcane (63.1)	• Sugarcane smut • Extreme climate
Wheat (51.4)	• Stem rust • Frost & salinity
Barley (19.2)	• Russian wheat aphid

Figure 2. 1:Challenges to Maize production: pest and diseases (Donateli *et al.,* 2017)

13

2.3.   Maize lethal necrosis disease (MLND)

Maize lethal necrosis disease is a viral disease of maize caused by a single infection with Maize Chlorotic Mottle Virus (MCMV) or a co-infection with any other member of *Potyviridae* family which infects cereals (Family, 2012). The disease was first reported in 1976 as corn lethal necrosis disease in North Central and South Central Kansas-USA (Bockhelman *et al*., 1982; Niblett and Claflin, 1978). It later spread to other regions including Central Nebraska (Doupnik and Wysong, 1979), Peru (Castillo, 1977; Uyemoto, 1983) and Hawaii US (Kauai) (Fatma *et al.,* 2016; Jiang *et al*. 1992). To date, several other regions have reported the disease and this include Brazil and Texas (Fatma *et al*., 2016), Argentina (Gingery *et al*. 1983), China (Xie *et al*., 2011) and East Africa (Adams *et al*., 2017; Wangai *et al*., 2012). A synergistic infection with a *Potyviridae* family, commonly Sugarcane Mosaic Virus (SCMV) or Wheat Streak Mosaic Virus (WSMV) results in a more profound necrosis effect (Figure 2.2).

In Kenya and East Africa, SCMV is the major potyvirus involved in MLND. The development of a severe systemic necrosis ranging from molecular to cytopathological effects during co-infection has been associated with the dramatic increase in the non-potyvirus components in dually infected plants (Kay Scheets, 1998). Morales *et al*. (1999) demonstrated using an electron microscope that maize leaves co-infected with the two viruses contained viroplasms in xylem and parenchyma leaf cells with viroplasms often occluding the lumen of xylem vessels and frequently leads to high disorganization of chloroplasts. Analysis of starch content of the infected leaf showed much smaller starch grains in the MCMV/SCMV infected plants suggesting that

photosynthesis in these cells was significantly impede. The low levels of starch have been attributed to low expression of Pyruvate-phosphate-dikinase (PPDK) gene in leaves co-infected with MCMV and SCMV (Bacheller, 2017). Pyruvate, orthophosphate di-kinase is a key enzyme in photosynthesis in plants that exploit the C4 photosynthetic pathway for the fixation of CO_2 (Wang' *et al*., 2017).

Unlike other diseases that affect maize, M L N D is the most difficult to control since the plants are susceptible to it at all stages of growth, from seedling to maturity. Insect vectors have been implicated in transmission of MLND causing viruses. These include thrips (Cabanas *et al*., 2013) and chrysomelid beetles (Nault *et al*., 1978) for MCMV and aphids for potyviruses (Brault *et al*., 2010). Maize lethal necrosis causal viruses (MLNVs) can also be transmitted through infected seeds, soils and human activities (Fatma *et al*., 2016). Since the initial report of MLND in 2011 in Bomet region of Kenya, MLND has spread to almost all maize growing regions in Kenya including Kericho, Nyamira, Kisii, Kiambu, Nakuru and Busia. Neighboring countries including Uganda, Tanzania, Rwanda, Burundi, Ethiopia, Zambia, Malawi and Mozambique have also reported the disease (Figure 2.4.) (Mahuku *et al*., 2015).   Integrated management of the disease has been met with challenges as the disease has many sources of infection. Also, the ability of MLND to attack maize at any stage of growth from seedling to maturity (Gowda *et al*., 2015) makes it difficult to use pesticides to control it as it means continuous applications from germination to harvesting which is not environmentally friendly and expensive.
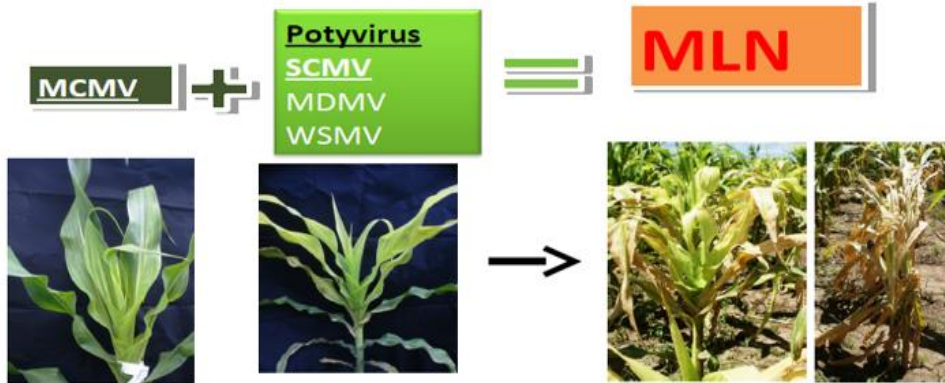
Figure 2. 2:Illustration of un- infected, singly and co-infected maize plants for comparing the synergism effects (Kay Sheets, 1998)
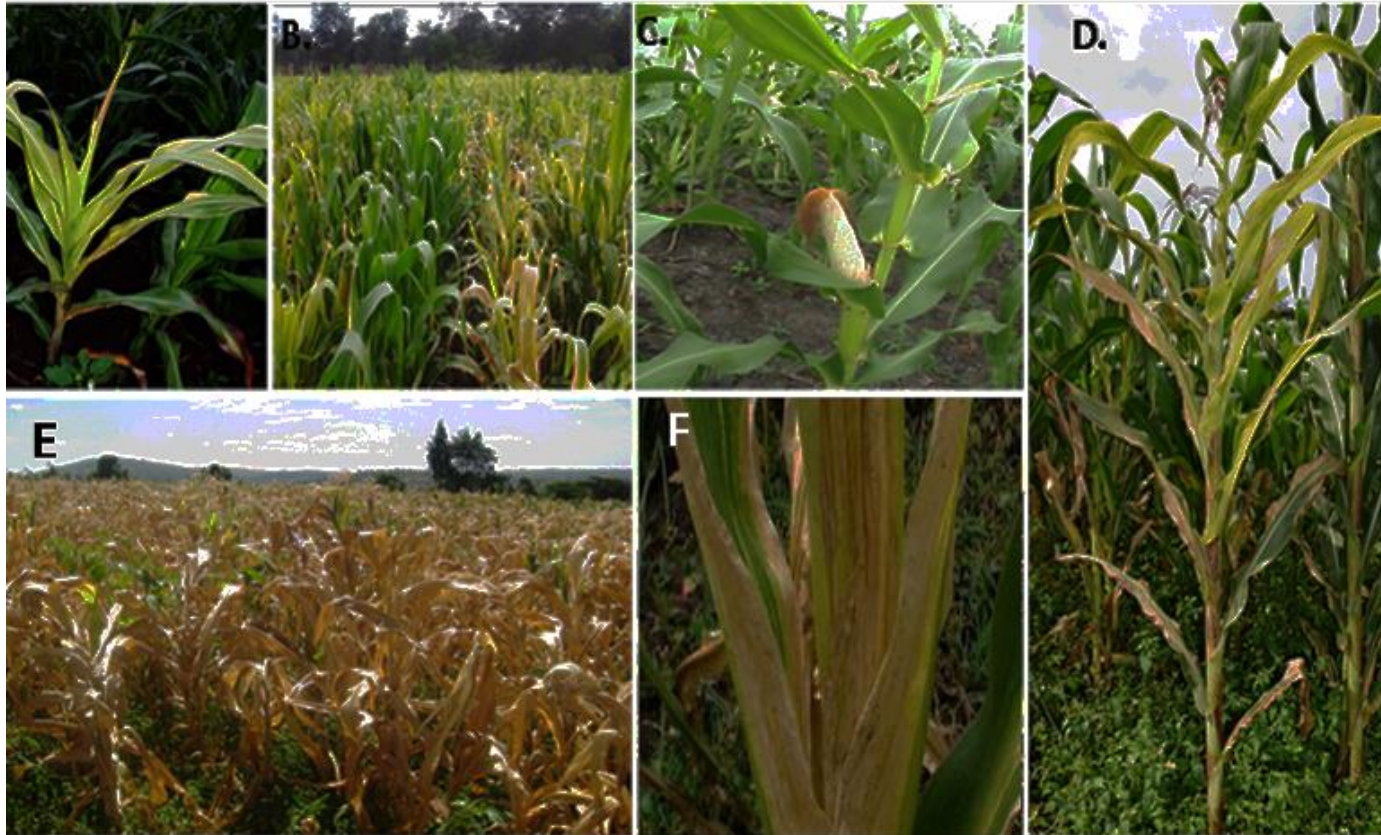
Figure 2. 3: Plants showing typical symptoms of MLND

A-Young maize plant with early symptoms of MLND, B- extensively damaged field with high number of plants infected, C-cob drying off before maturity, a sign of MLND, D-extensive necrosis on the leaves of plants at knee-height stage, E-complete necrosis in maize plantation and F-Necrosis in leaves of a maize plant at knee-height.

Figure 2. 4: Reported occurence of maize lethal necrosis disease in Africa (CIMMYT).
Established known incidence of MLN by country (Red Markers), No MLN (Green Markers), reported but No confirmation (Yellow Markers)

2.4. Maize lethal necrosis disease causal viruses

2.4.1. *Tombusviridae* and *Potyviridae* families of viruses

Tombusviridae family consists of a group of single- stranded positive sense Ribonucleic acid (RNA) plant viruses. There are currently 71 known species in this family, divided

in 13 genera (ICTV, 2018.). All *Tombusviridae* members have a non-segmented linear genome, with the exception of *Dianthoviruses*, whose genome is bipartite (Sastry, 2013). The genome is approximately 4kb in length as shown in Figure 2.6, lacks a 5' cap and a poly(A) tail, and it encodes 4–6 open reading frames (ORFs) (Xu *et al*., 2017).The polymerase codes for an intermediate stop codon which is the position of a read-through event within open-reading frame1, yielding two products essential for replication. Among the members of *Tombusviridae* family are *Machlomovirus,* a genus of plant viruses where plants belonging to Poaceae family serve as natural hosts with one notable member being Maize Chlorotic Mottle Virus (MCMV). MCMV is the only species in this genus and is the causative agent of maize (corn) lethal necrosis disease which is responsible for significant losses in maize production worldwide (Omarov *et al*., 2006).

Potyviridae is the largest family of plant viruses comprising of *potyvirus* genera which cause significant losses in a wide range of crops (Revers and Garcia, 2015). Potyviruses are transmitted by aphids in a non-persistent manner while others are transmitted by seed (Koenig and Kühn-institut, 1981). They have a (+)-strand ssRNA genome, which codes for a polyprotein that is processed to the final gene products. Most Potyviruses have non-segmented genomes, though a number of species are bipartite (Audy *et al*., 1994). Among the members of this family are Sugarcane Mosaic Virus (SCMV) and Wheat Streak Virus (WSV) which co-infect maize plants with MCMV and results in severe necrosis.

2.4.2. Maize Chlorotic Mottle Virus

Maize chlorotic mottle virus (MCMV) belongs to tombusviridae family and is the only member of the genus *Machlomovirus* (Kay Scheets, 1998). It is related with Tombusvirus in that they possess a positive sense, single-stranded RNA that is encapsulated in isometric particles about 30nm in diameter. The entire genome of MCMV has six (Figure 2.6) identified open reading frames (ORFs) including ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (Wang *et al*., 2017). ORF1 encodes a 32-kDa hypothetic protein of unknown function. Protein-50 (p50), a 50-kDa protein is encoded by ORF2 with its N-terminus-overlapped with p111 protein produced by translational read-through of the UAG stop codon of ORF2. Movement protein p7, a 7kDa protein is encoded by ORF4. P31 protein is expressed from ORF5 when the UGA stop codon of ORF4 is suppressed and plays a role in cell-to-cell movement of MCMV (Wang *et al*., 2017; Iqbal *et al*., 2017). P50/P111 (p111 being a read-through protein) both contain a glycine- aspartate- aspartate motif (GDD) box which is also named polymerase domain, common in positive strand RNA viruses, and have been predicted to be Ribonucleic acid- dependent RNA-replicases (RdRp) (Wang' *et al*., 2017). The virion RNA is infectious and acts as both the genome and viral messenger RNA.
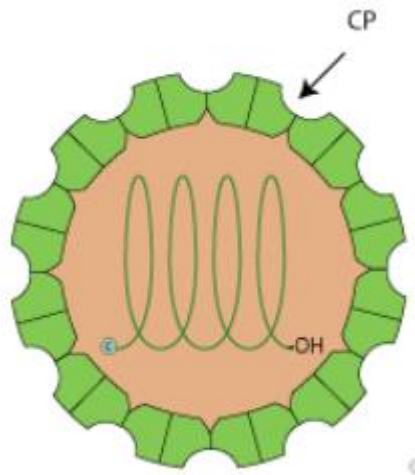
Figure 2. 5:Encapsulated structure of MCMV virion (Viralzone, 2018a)
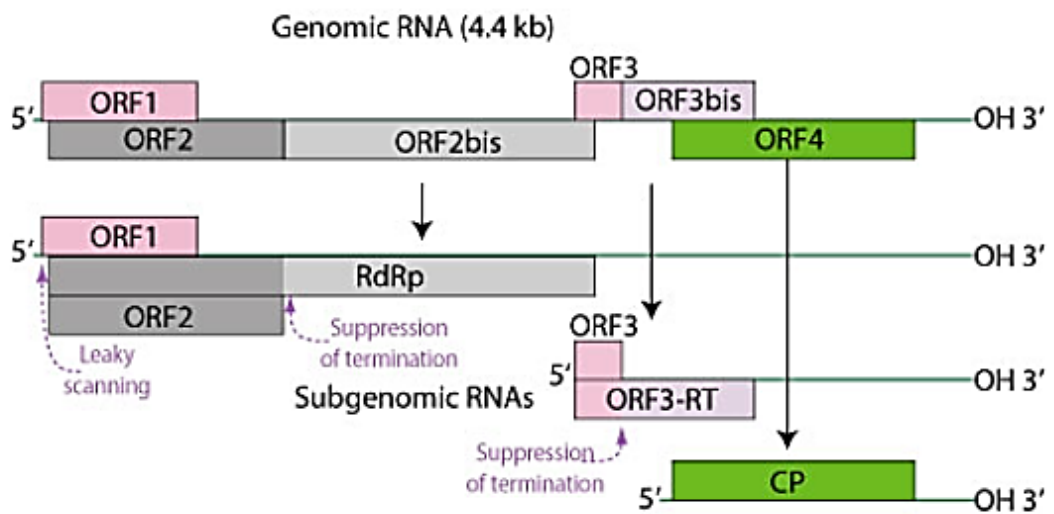
Figure 2. 6: Organization of the MCMV genome (Viralzone, 2018a)

ORF1-Protein of unknown function, ORF2-replicase protein encoding region, ORF4-Movement protein p7 and capsid protein.

2.4.3.   Sugarcane mosaic virus (SCMV)

Sugarcane Mosaic Virus (SCMV) is a potyvirus belonging to the family *Potyviridae*. The virus was first observed in Puerto Rico in 1916 (Brandes, 1919) with the first reports in East Africa's sugarcane and maize in 1973 (Kulkarni, 1973). Currently, SCMV is the main potyvirus implicated in synergistic relationship with MCMV in East Africa during the development of Maize lethal necrosis disease (Wangai *et al*., 2012; Adams *et al*., 2013, 2014; Lukanda *et al*., 2017).

SCMV is made of 9596 nucleotides long RNA and unlike the MCMV, SCMV is surrounded by a non-enveloped capsid with helical symmetry (Figure 2.7). Its length lies between 680 and 900 nm. The genome of SCMV has a single, open reading frame that is translated into large polyprotein of 340–370 kDa which is post-translationally cleaved by Nla proteinase to produce at least 10 products (Shukla *et al*., 1994). Hc-Pro domain serves as an oligomer and plays a role in interaction between the virion and the vectors (Plisson *et al.,* 2003). Additionally, the central part of this protein is involved in viral amplification, suppression of gene silencing, synergism with MCMV as well as symptom development (Thornbury and Pirone 1983). The region serves as the main factor in host response and pathogenesis during the SCMV infection.

The host range for SCMV includes maize, sugarcane, sorghum as well as other members of Poaceae family. SCMV is transmitted in a non-persistent mode by vast members of aphid species including *Aphis maidis, Aphis craccivora, Rhopalosiphum padi, Dactynotus ambrosiae* and *Hysteroneura setariae.* Low seed transmissions at rates of 0.4% to 3.9% have also been reported for SCMV with the virion particles remaining

latent in silks, glumule and kernel. Equally, a low soil transmission rate has been observed for SCMV with an average of 0.7-5.4% (Li *et al*., 2011).
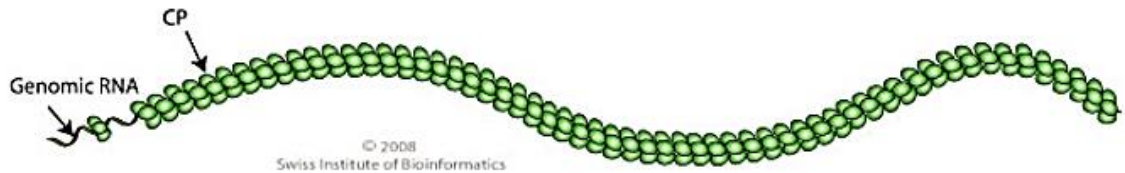
Figure 2. 7:The structure of SCMV, linear ssRNA molecule which is non-encapsulated (Viralzone, 2018b)
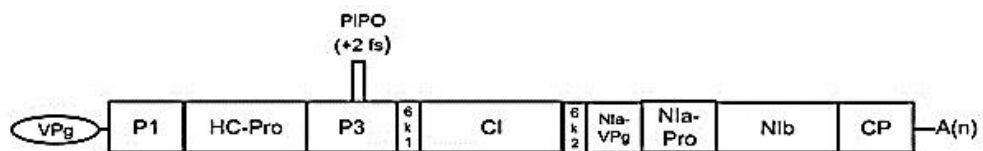
Figure 2. 8:SCMV genome structure showing P1, HC-Pro (a VSR), VPG (recently identified as a VSR) and coat protein coding region (CP) (Viralzone, 2018b)

2.4.4. Host –range of SCMV and MCMV viruses

MCMV and SCMV majorly infects Poaceae members (formerly-Gramineae) with maize and sugarcane being the natural hosts for MCMV and SCMV, respectively (Scheets, 2004). It has been reported that all weeds belonging to the grass family and growing in close proximity to the sorghum eventually end up being infected with SCMV through the insect vectors (Koike and Gillaspie, 1989). MCMV and SCMV has been reported to cause systemic infections in several grass members including Finger millet, (*Eleusine coracana*), prosso millet (*Panicum milliearceum L*), fox tail millet (*Setaria italical*), wheat (*Triticum aestivum L*), sugarcane (*Saccharum officinarum*) and barley (*Hordeum*

vulgare L) (Kusia *et al., 2015;* Castillo and Hebert, 1974; Niblett and Claflin, 1978; Bockelman *et al.*, 1982; Brunt *et al.*, 1996).

2.5.   Management strategies against MLND

MLND   being a viral disease is challenging to control. Management of the disease is through integrated pest management strategies comprising cultural control such as crop rotation, vector control, closed season as well as crop diversification. Pragmatic interventions have also been proposed and this include efficient and effective surveillance programs and diversification of food habits. Uyemoto *et al.* (1980) demonstrated the ability of crop rotation to break the viral cycle. This is done with non-Poaceae members such potatoes, beans, cassava, bulb onions and vegetables (Mahuku *et al.*, 2015). Seeds have been identified as transmission avenue for MLND viruses especially MCMV and therefore, to avoid progressive transmission of the viruses, recycling of seeds should be avoided and instead certified seeds should be used for propagation of new maize plants (Fatma *et al.*, 2016). Imposing a quarantine is also another strategy to prevent the movement of infected maize materials to non-endemic areas. This is however the least effective control since controlling the movement of insect vectorsb is an impossible task. This is equally the same with the management of insect vectors using chemical pesticides which faces a challenge of resistance as well as the diverse nature of MCMV and SCMV and environmental pollution.

The most viable strategy for the control of MLND is breeding for tolerance/resistance. Development of virus-tolerant and resistant varieties is not only environmentally sustainable but also economically viable as a strategy for the control of the disease.

Breeding for tolerance is a long process that involves the establishment of plants in fields, artificial inoculation with MLND causal viruses and scoring for tolerance/resistance based on symptom development and the extent of damage (CYMMIT, 2018; Mahuku *et al*., 2015). Development of tolerant varieties not only require the assurance of disease resistance but also incorporation of agronomically desirable traits such as good yield. Studies by Nelson *et al*, (2011) while demonstrating the potential of developing tolerant varieties reported high levels of tolerance to MLND by tropical inbred varieties even though there was no 100% immunity reported. In Kenya, several lines are being screened for resistance/tolerance by KALRO and CIMMYT in centralized MLN screening stations in Naivasha and Bomet. Unfortunately, many temperate inbred and hybrid lines are showing high levels of susceptibility to the viruses (Wangai *et al*., 2012). Further reports by CYMMIT under 'CIMMYT Global Maize Program' on 2013 screening indicated that 122 of the screened 124 varieties had extremely high levels of susceptibility with additional screening of 62000 lines showing up to 90 % susceptibility (CYMMIT, 2018). As of 2017, only nine highly tolerant varieties namely Bazooka (UH5354), H12ML, H13ML, Meru HB607, WE5135, WE5136, WE5138, WE5139 and WE5140 had been identified (CYMMIT, 2018; Makumbi and Wangai, 2012).

2.6.   Plant immune system and MLN viruses

RNA-mediated gene silencing also known as RNA interference (RNAi) is a conserved sequence-specific gene regulation system, which plays a crucial role in the maintenance of cell genome integrity. In higher plants and insects, it operates as an adaptive inducible antiviral defense mechanism (Basu *et al*., 2014). It is made of

sequence-specific recognition and inhibition of target gene expression by endogenous (gene, development, stress response regulation) or virus-derived short silencing RNAs (vsiRNAs) mechanisms. Target gene expression can be inhibited post-transcriptionally (post transcriptional gene silencing (PTGS)) where target double stranded RNAs (dsRNAs) are recognized by Dicer-like enzymes (DCL) and diced into small interfering RNA duplexes (siRNA) of about 21- to 24 nucleotides (nt), which interact with Argonaute (AGO) and associated proteins to form RNA-induced complexes (RISCs) to target homologous RNAs for destruction (Zhang *et al*., 2015). Plants, through this mechanism of RNA- interference defend themselves against viral attack.

Viruses have however developed a strategy against the RISC mechanism where they use Viral Suppressor of RNA proteins (VSRs) to inhibit the RISC mechanism. VSRs are able to impede host anti-viral responses by interacting with key components of cellular silencing mechanisms, every so often mimicking normal cellular functions. One such protein, suppressor protein p19; spoils plant defence by transporting the silencing signal into the peroxisomes to avoid its systemic spread (Cabanas *et al*., 2013). VSRs proteins include Helper component-proteinase (Hc-pro), protease 21 (p21) and suppressor protein p19. The strategy to overcome MLND through RNAi technology has been rendered inefficient due to the effect of these proteins (Cabanas *et al*., 2013).
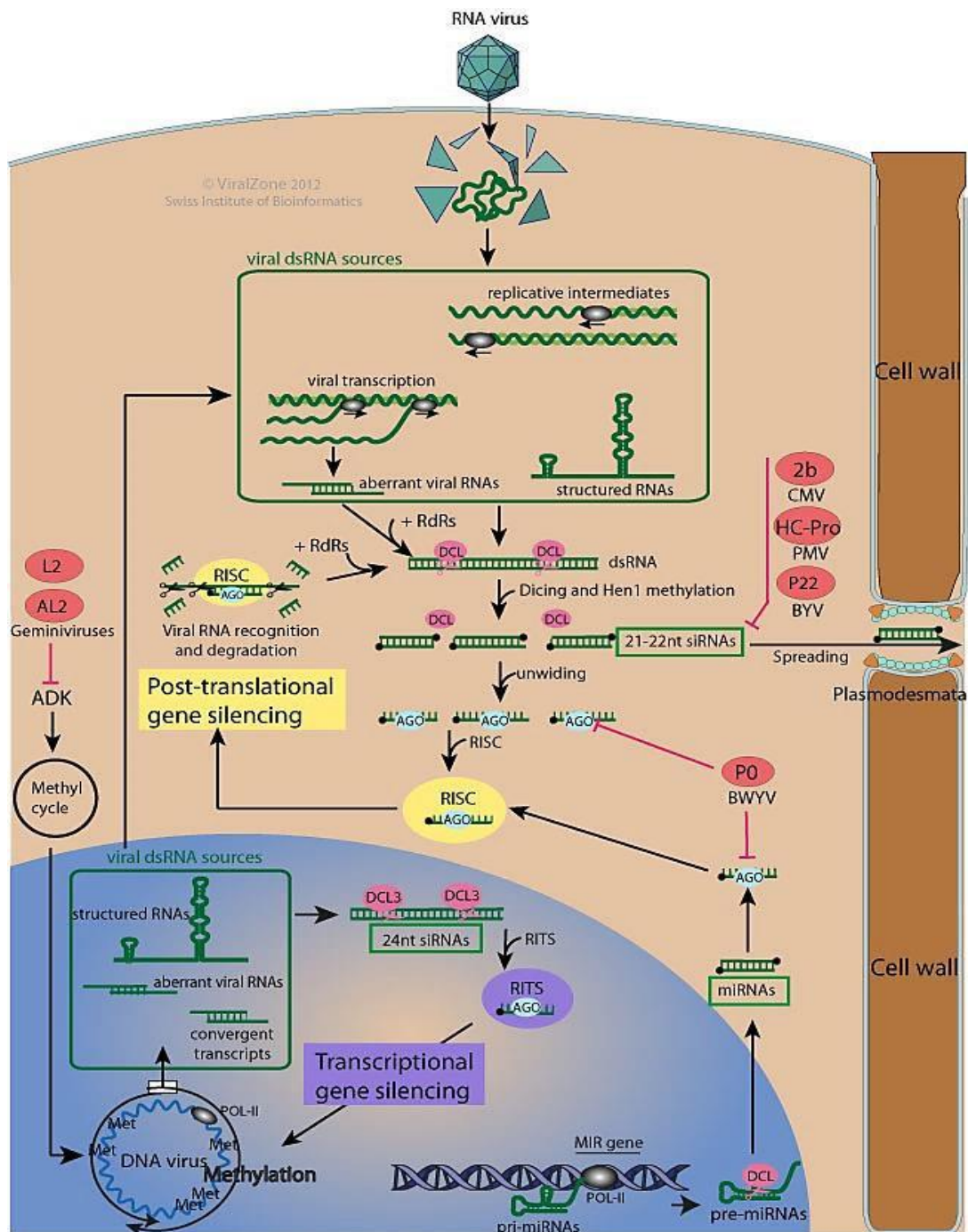
Figure 2. 9:viral suppressors of RNA silencing (VSRs).

(source: https://viralzone.expasy.org/891?outline=all_by_protein)

Plant's immune system consisting of RNA-interference recognizes double-stranded RNA molecules or structured RNA and dice them to small interfering RNA molecules thereby inactivating the virus. Plant viruses have however developed an evasion strategy by encoding viral suppressor proteins including HC-Pro, PO and P22 which

inhibits this system by binding to core proteins in RNA-interference pathway such as argonaute (AGO) and DCL.

2.7. Strategies for evasion of host immunity by MLN causal viruses

2.7.1. Viral suppressor proteins

Three major strategies have been proposed to explain the ability of plant viruses to counter the endogenous RNA-silencing mediated defense which include preventing the generation of small interfering RNAs (siRNAs), inhibiting the incorporation of siRNAs into effector complexes and interfering with one of the effector complexes (Li and Ding, 2006). Three key players in the inhibition strategies by the viruses are proteins P19, P21, helper component- proteinase (Hc Pro) (Voinnet *et al*., 1999). It has been reported that potyvirus-encode helper component proteinase (HC-Pro); the first viral protein identified as Ribonucleic Acid- Silencing-Suppressor (RSS), enhances the replication of un-related viruses. This is achieved by acting through reverse-establishment of calmodulin-like protein (rgs-CaM) which is a cellular negative regulator of post-transcriptional gene-silencing (PTGS) (Kasschau *et al*., 2003). It is encoded by the 5' proximal region of the Tobacco Etch Virus (TEV) and is 1380bp long (app.1). Its characteristics in different potyviruses have been described with one outstanding feature common to all HC-Pros being the difference in accumulation of various micro-RNA (miRNA) molecules and miRNA target transcripts (Li and Ding, 2006). With Hc-Pro being a conserved domain in *Potyviridae* viruses, targeting of the coding sequence for this protein is a feasible approach against these viruses.

P19 is a 19kDa protein (p19) from *tombusviridae* which specifically binds 21-nt ds siRNAs in vitro and in vivo, preventing siRNA incorporation into effector complexes such as RISC (He *et al*, 2004). A series of studies on P19 has shown its mediation in RNA silencing (Figure 2.10).
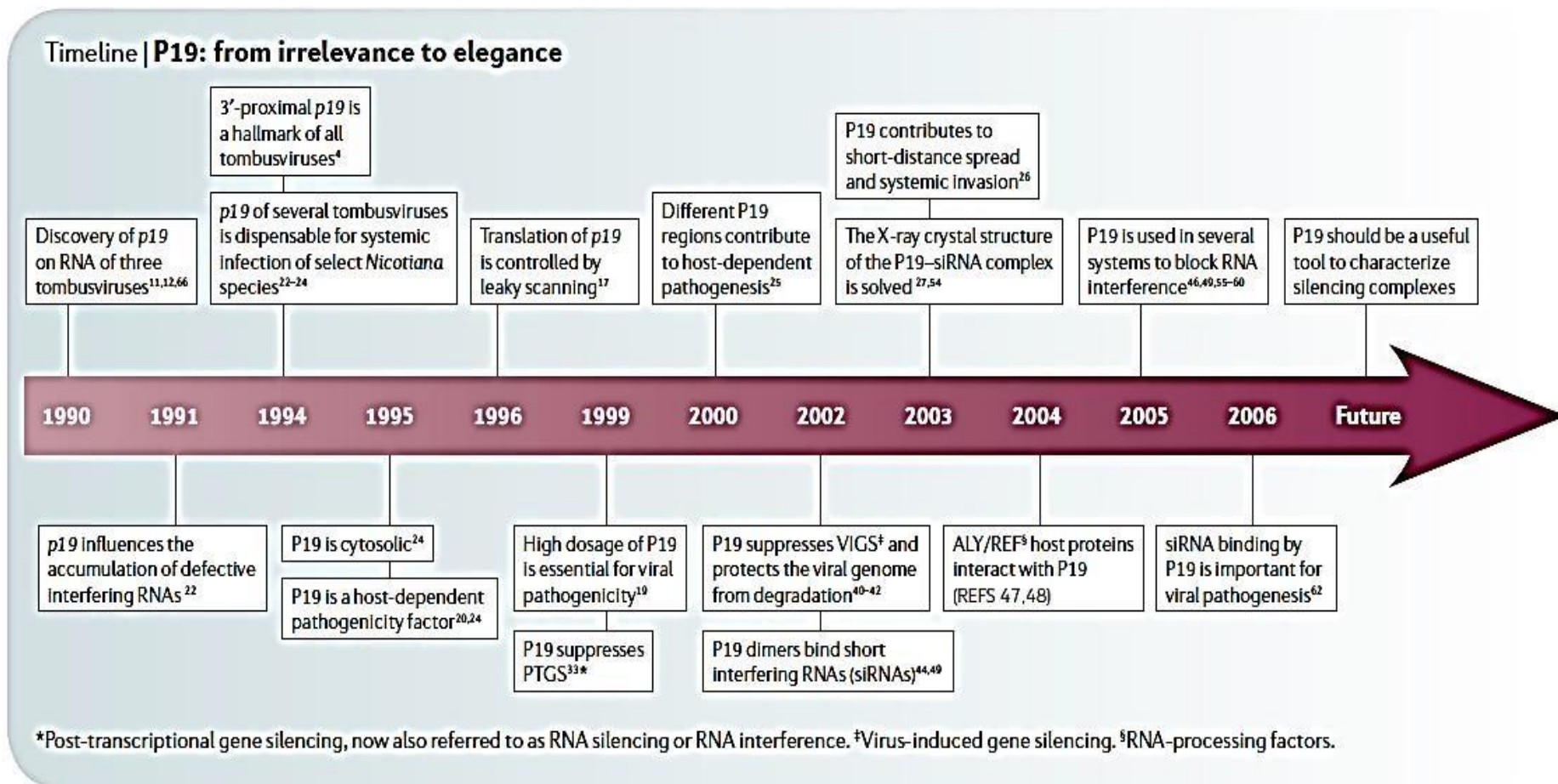
Figure 2. 10:Series for recognition of P19 as the inhibitor of RNAi (He et al., 2004).

Protein p21 was discovered in Beet yellow virus (BYV) counteraction mechanism and inhibits silencing pathways by binding siRNAs or ds miRNA intermediates.

2.7.2. Mechanisms for Suppression of host immune system

The first mechanism entails binding of proteins p19 and p21 to dsRNA (Li and Ding, 2006). These proteins have been shown to bind and inhibit the function of 21-nt siRNAs leading to suppression of RNAi mechanism (He *et al*., 2004). SiRNA which are 21-22 nucleotides long, are genetic elements produced by RISC mechanism to target specific RNAs and form duplexes by complementary binding to allow splicing and degradation by RISC dicer-complex. Although siRNA duplexes of other sizes may also be bound by these two viral suppressors of RNAi, the binding affinity weakens rapidly with amassing size differences. Due to its target specificity, TBSV p19 has been used as a universal RNAi suppressor to explore the molecular mechanism of 21-nt siRNAs and miRNAs.

Second mechanism of VSR entails suppression of siRNA production. This mechanism uses VSR proteins including HC-Pro following the dsRNA binding. Inhibition of viral siRNA production in infected cells suppresses RNAi by preventing Dicer from accessing the viral RNA trigger(s) (Li and Ding, 2006). HC-Pro also inhibits Dicer processing and increased accumulation of unprocessed dsRNA. This has been illustrated in transgenic plants with studies indicating its ability to inhibit the accumulation of the 21-nt siRNAs but does not inhibit, or has a less pronounced effect on, the accumulation of the 24-nt siRNAs (Syller, 2012).

Third mechanism entails sequestration of siRNAs and this is facilitated by p19 protein. Development of immunity against viruses by plants requires recruitment of a

portion of viral double stranded siRNA into RISC system. p19 binds double stranded siRNA to prevent siRNA from being incorporated into siRISC (Li and Ding, 2006). This prevents the plant from developing immunity against the same virus.

2.8. Approaches in detection and characterization of plant viruses

Detection of plant viruses ranges from conventional approaches including Enzyme-Linked Immunosorbent Assay (ELISA) test and Symptomatological analysis, Molecular based techniques such as PCR and genome sequencing and lately the use of small RNAs (Fatma *et al.*, 2016). Serological-based techniques such as ELISA which includes (triple antibody sandwich ELISA (TAS-ELISA), double antibody sandwich ELISA (DAS-ELISA) and direct antigen coating-ELISA (DAC -ELISA) serve as pre-diagnosis methods for plant virus detection including maize lethal necrosis disease causal viruses (MLNV). More accurate and precise nucleic acid-based techniques such as polymerase chain reaction (PCR) and sequencing techniques have also been used to detect MLND viruses (Fatma *et al.*, 2016). Conventional approaches are fast and easy to use. There is however a limitation with the use of these techniques as they do not provide information on evolution and interactions between the host and the viruses since such interactions occurs at molecular level which is beyond their scope. Molecular biology approaches are reliable and can provide this information. PCR analysis and gene sequencing are accurate, reproducible and have the potential to allow full characterization of the viruses. Through these techniques, the underlying basis of disease development and progression can be adequately addressed. Finding remedies to plant viral diseases lies with proper identification of causative agents and understanding of

mechanisms involved during pathogenesis. The error-prone replication and recombination which is common in RNA viruses creates extensive diversity which hinders the potential of serological and symptomatological assays. This requires molecular tools to exploit and understand the evolution dynamics of these viruses (Cross, 2015). Additionally, genetic diversity studies of plant RNA viruses in various host plants provide valuable information on evolution and emergence of new diseases caused by RNA viruses. Cross *et al*. (2015) by analysis of high- throughput sequencing technology products of Rice stripe virus were able to demonstrate a direct evidence that viruses might allow maximal genetic diversity for host adaptation. Relatedness of viruses provides a guide in determination of host-range and can aid in crop management practices including weed management, intercropping and crop rotation. In a phylogenetic study on capsid protein from Wheat streak mosaic virus belonging to genus *Tritimovirus* and Oat necrotic mottle virus of genus *Rymovirus,* analysis showed up to 76% nucleotide identity for isolates from different countries and this explained the ability of the two viruses to infect the same host (Braidwood *et al*., 2018).

2.9. Small RNA and MLN viruses

Small ribonucleic acids (sRNA) are short nucleotide sequences (20-35 nulcleotides) that are non-coding and are implicated in the control of gene expression through post-transcriptional gene silencing. These molecules have the capacity to control gene expression both at nucleus and cytoplasmic levels (Chunxiang Zhang, 2009). There are three classes of small Ribonucleic acids that have been well-defined, including microRNAs (miRNAs), small interfering RNA (siRNAs) and Piwi-interacting RNAs

(piRNAs). Additionally, a special group of sRNA named viral-derived siRNA (vsiRNA) have also been identified as sRNAs responsible for inhibition of the RISC system (Iqbal *et al*., 2017). Small RNAs have also been implicated in processes such as cell proliferation, cell differentiation, apoptosis, migration, metabolism and defense. They also play a critical role in regulation of normal development and physiology (Chunxiang Zhang, 2009).

There is a close relationship between pathogenesis and sRNAs both in plants and animals (Zhang, 2009; Iqbal *et al*., 2017). While in animals sRNAs have been associated with diverse diseases including cardiovascular, stroke, cancer, diabetes, neurodegenerative, liver and kidney diseases, sRNAs in plants infected with viral agents have been implicated in down-regulation of the plant's immune system. It is because of this that sRNAs can serve in transcriptomic analysis as well as diagnosis of diseases and detection of pathogens. Studies by Iqbal *et al*. (2017) on the mechanism of MLND progression indicated the potential host- derived miRNAs in Maize as well as viral source induced RNAs (vsiRNAs). The host derived miRNAs can serve to designate the presence of viral agents as well as indicate the host capability to resist the viruses. This therefore allows for both detection of the viruses as well as a means for screening for tolerance against MLN causal viruses. Viruses especially RNA viruses have a high mutation rate arising from error-prone polymerases and limited RNA proofreading functions (Illumina, 2013). The resulting diversity is essential for adaptive evolution and the capacity of the viruses to cause disease. Analysis of MLN derived vsiRNAs and mRNAs will therefore be an efficient tool for analysis of diversity as well as the interaction among the viruses and the host.

2.10. sRNA sequencing in genetic diversity and identification of plant viruses

sRNAs play a critical role in the host-pathogen interaction during infection by viruses. Sequencing of viral mRNAs provides a wealth of information on the activity of viruses as well as their mechanisms of action during pathogenesis (Illumina, 2013). This information can allow annotation of the viral genome and aid in explaining the virus-host interactions at molecular level during infection and disease progression. The availability of next-generation sequencing makes it simple and efficient to sequence viruses and derive crucial information necessary for full characterization of the viruses. While several studies on diversity of MLN causal viruses focus on the whole genome sequencing, the actively transcribed regions which explain the viral strategies in counteracting the maize immune system is oftenly not exploited (Burgyán and Havelda, 2011; Soitamo *et al.*, 2011; Xia *et al.*, 2016; Zhu *et al.*, 2016). Limited knowledge on viral sequence roles presents a large gap in gaining insights of virus population dynamics. Since both RNA and DNA viruses use miRNAs for the host and viral gene regulation, sequencing and identifying these sRNAs will aid in expanding the current knowledge of virus-host interactions by uncovering genes that manipulate their hosts in unexpected ways (Rosario and Breitbart, 2011).

MLN viruses are ssRNA viruses and undergo high recombination and mutation rates and are considered among the fastest evolving biological entities (Xu *et al.*, 2017). Xu *et al.* (2017) described viral populations within a single tomato host plant containing many non-identical but similar genome sequences (quasispecies). From their study, several recombination events were discovered among the Cucumber mosaic virus and

tomato mosaic virus showing the highest recombination. Guimaraes *et al*. (2015) developed a sequence-independent strategy based on virus-derived small RNAs produced by the host response including the RNA interference pathway. By performing comparison of sequences of small and long RNAs, the study demonstrated that viral sequences are enriched in the small RNA fraction and size profiles provided a unique signature for each virus which means that it can be used to identify novel viral sequences without known relatives in the reference databases.

Qi *et al*., (2009) using deep sequencing analysis of vsiRNAs in tomato mosaic virus demonstrated the potential of vsiRNAs to target specific host sequences for post-transcriptional silencing. Through this study, it was possible to point out that there is an underlying possibility for viral siRNA-mediated virus-host interactions to contribute to viralpathogenicity and host specificity. The extreme richness of small RNA molecules associated with viruses in infected plants allow small RNAs to be used as tools for identification and characterization of plant viruses. Chung *et al*., (2017) while studying diversity, distribution, and evolution of tomato viruses in China used deep small RNA sequencing and was able to detect 22 new single stranded RNA viruses. The ability of both DNA and RNA viruses to initiate the production of virus- derived small interfering RNAs (vsiRNAs) allows characterization of viruses based on length distribution, paired distance and base selection bias of vsiRNAs. Profiling of vsiRNAs sequences allow insight into different plant Dicer- like proteins and Argonautes involved in vsiRNAs biogenesis (Xu *et al*., 2017). Aguiar *et al*. (2015) were able to fully characterize six novel viruses based on sequence-independent characterization patterns of viral small RNAs produced by the host. Several studies have shown that plants' RNA silencing mechanism is actively induced in the presence of a viral attack and

this observation makes plant siRNAs and vsiRNAs a noble tool not only for identification but also for characterization and evolutionary analysis of plant viruses (Canada *et al*., 2018; Chapman and Carrington, 2007; Koenig and Kühn-institut, 1981; Wu *et al*., 2010; Xu *et al*., 2017; Chunxiang Zhang, 2009).

2.11. Application of detached leaf assay in analysis of disease tolerance

Analysis of plant tolerance to viral, fungal and bacterial diseases is conventionally done by whole plant inoculation (WPI). This is the approach applied in screening of maize germplasm against maize lethal necrosis disease in various research institutions. The method is however cumbersome and requires a large screen house or field establishment of the crop. It is also undesirable in screening for multiple disease resistance in plants (Mo, 2007). Detached leaf assay (DLA) is an alternative to the whole plant assessments, which simplifies handling large numbers of genotypes, and allows individual germplasms to be challenged with several foliar pathogens. Additionally, laboratory based DLA analysis requires less space and has been shown to be more sensitive than WPI in measuring relative resistance against plant pathogens (Kuma, 2009).

DLA has been applied in several studies in crop pathology. Kuma (2009) used DLA method to avoid contaminated soil and irrigation water to fully demonstrate the relationship between pathogenic inoculum concentrations of isolates of *Pythium sylvaticum* and *P. ultimum* in soil and the expression of root rot symptoms. DLA has also

been applied in identification of cassava varieties with resistance to cassava anthracnose disease (CAD), a fungal disease caused by *Colletotrichum gloeosporioides* (Stobbs *et al*. 2015). Mo (2007) developed a DLA screening technique to evaluate soybean for resistance to powdery mildew caused by *Microsphaera diffusa*. The study was able to demonstrate the ability of DLA to be used for analysis of response to multiple fungal pathogens in a short period. Browne *et al*. (2006) while detecting components of partial disease resistance investigated a range of wheat cultivars based on DLA and the study concluded that the selection of the first or second growing leaf of wheat is optimal for the use in the DLA using *M. majus* for studying components of partial disease resistance. Assessment of plant tolerance to the bacterial diseases has also been done using the DLA method with the method demonstrating very high efficiency as symptoms are seen 48 hours after inoculation as compared to 6-8 days for whole plant innoculations (Stobbs *et al*., 2015). Results from a study on *Plum pox virus* (PPV) transmission efficiencies further demonstrated the effectiveness of DLA. Detached peach leaves subsequently maintained on viruliferous aphids and transferred to agar layer for three weeks was not significantly different from that for intact seedlings (Stobbs and Samara, 2015). These studies demonstrated the potential of DLA method to validate the reaction of plants against various plant pathogens.

CHAPTER THREE

MATERIALS AND METHODS

3.1. Study site

Sampling of infected maize plants and potential alternate hosts was carried out in three counties in Kenya namely Bomet, Kericho and Nyamira. These counties were selected as they were the most affected by the disease. The entire experiment was carried out in plant transformation facility (PTL)-Kenyatta University. qPCR analysis was done in the Pan African University Laboratory at Jomo Kenyatta University of Agriculture and Technology.

3.2. Study design

Collection of samples for small RNA-next-generation sequencing was done in the three counties listed above selected based on the previous reports on MLND disease severity. From each county, six maize fields were selected randomly and 12 maize samples from the top-most leaves of maize plants showing MNLD symptoms were collected. The twelve samples per field were mercerated and pooled to a single sample to serve as a representative of each farm. For the identification of alternate hosts, a diagonal work pattern was adopted across infected maize fields and plants growing in close proximity to infected maize both symptomatic and asymptomatic were selected. A total of 17 plants were collected. Three tolerant maize genotypes (E1-SS-15-01/CHMLND0093, E2-SS-20-25/CKIR11027 and E3-SS-20-1/CKIR12032) and two susceptible genotypes (H144 and *Namba nane*) were selected

for detached leaf study. Each germplasm was represented by two leaf explants per plate replicated three times and repeated five times. The entire experiment was then repeated twice. Continuous scoring of MLND symptoms from day 1 to day 19 post-infection was done using CIMMYT necrosis scoring chart.

3.3. SRNA profiling using next-generation sequencing for detection of MLND causal virus

3.3.1. Sample collection

Maize leaves showing typical signs of MLND including chlorotic mottling of leaves, premature drying up of cobs, leaf necrosis at the leaf margins and 'dead heart' (Mahuku *et al*., 2015) were collected from the three counties namely Kericho (Buret sub-county), Bomet (Konoin sub-county) and Nyamira (Borabu sub-county). Twelve plants from six different, distant farms per location were selected as representative samples. Samples from each farm were then arranged and mercerated into small pieces, mixed and placed in RNase-free Eppendorf tubes to fill 1/3 of the tube. The tubes were filled with DNA/RNA shield reagent (ZymoResearch, 2018) to stabilize the RNA during transportation and to deactivate any external contaminants. Upon arrival in the laboratory the samples were kept in 4°C refrigerator awaiting total RNA isolation.

3.3.2. Total RNA isolation and sequencing

Total RNA isolation was done using Directzol^{TM} kit from ZymoResearch following the manufacturer's instructions. Briefly, 50mg of leaf tissues were frozen in liquid nitrogen and ground into fine powder by vortexing in steel beads. The powder was

thereafter treated with 500µl Trizol reagent to lyse the cells for 10 minutes and centrifuged at 10000RPM for 5 minutes to separate nucleic acids from plant debris. The aqueous phase was then transferred into clean, RNAse-free eppendorf tubes and processed through column treatment with buffer solutions provided. DNA contaminants were removed by DNAse treatment. The resulting RNA was eluted with 25ul RNAse free water and quantified using a nanodrop spectrometer. Further validation was done using gel electrophoresis. Total RNA samples were then shipped in dry ice for cDNA library construction and sequencing at BGI-Hong Kong. Size PAGE gel was done and 18-30nt fragments were harvested for further analysis. The 3' adapter ligation was achieved using a 5'-adenylated, 3'-blocked single stranded DNA adapter. RT primers were then added to the ligated 3'adapters and dissociative 5' adapters. Ligation of 5'end was achieved using 5'adapters. One strand cDNA synthesis was done using RT-primer through reverse transcription extension followed by PCR amplification and enrichment. This was followed by library quantification and pooling cyclization. Sequencing of the libraries was done based on BGISEQ500 platform, a high-throughput sequencing solution, powered by combinatorial Probe-Anchor Synthesis (CPAs) and DNA Nanoballs (DNB) technology.

3.3.3. Bioinformatics analysis

Upon completion of sequencing, raw data was filtered by removal of raw quality reads at 5' and 3' ends as well as getting rid of reads without the insert tag and those with poly A using cut-adapt. Reads shorter than 15nts 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%. The clean reads were then exported to *sRNAworkbech* pipeline where

host-derived miRNAs were filtered by performing subtractive mapping to *Zea mays* miRGENE and mirBase miRNA databases. To completely get rid of host-derived reads, repeated mapping was done by mapping the cleaned reads to miRNA library followed by genome mode mapping. Further mapping of host-filtered reads done using Bowtie2 tool within Geneious version 11.5 to map the reads onto the host genome under default parameters to further eliminate any leaky host-derived miRNAs.

MCMV NC_003627.1 and SCMV NC_003398.1 Refseq sequences were retrieved from NCBI for mapping, annotation and identification of highly expressed domains of the viruses based on host-filtered siRNAs. Mapping of sRNA reads was done using Bowtie2 under Geneious software and calculation of expression levels was performed based on normalized values of raw read counts with reference to mature peptides domains of the viruses. This was done to avoid bias based on the length disparity with reference to the mature peptide domain (mtr-PD). For MCMV, 7 domains were targeted including the ORFs P31, P50, P111, P7a, P7b, replicase associated protein and P32 CDS. For SCMV, 10 well-characterized domains were used including polyprotein CDS, PI protein, Nla Vpg, Nlb replicase, Nla pro, HC-Pro, 6k2, P3, CI and 6k1 protein. After identification of highly expressed and conserved domains, primers able to amplify these domains were designed. This was done with strict observation of parameters including GC content 40-50%, GC-glamp at 3' end, length 20-25nt, Self-complementarity of less than 2. A Confirmation of product size was done using *Sequence Manipulation Suite* (SMS) with qPCR product size restricted to 75-250bp.

Small RNA *de novo* assembly for host-filtered sRNA from the three samples was performed using *velvet* software within the *Virusdetect pipeline* (Zheng *et al*., 2017). Contigs that had a coverage of more than 75% and at least 10% genome coverage (10% COV) were reported. For each identified virus, one longest continuous contig was selected for phylogenetic analysis. Genomes from viral isolates reported in previous studies from Nebraska, Taiwan, Rwanda, Bomet, KARLO, Ecuador, Elgeyo Marakwet, Kirinyaga and Yunan were retrieved from NCBI's database for determination of evolutionary relatedness with the assembled isolates. For target prediction analysis, plant sRNA target analysis server (*psRNATarget*) was used with the set parameters of 200 top targets, expectation value of 5, penalty of 1, gap opening penalty of 2 and an extension penalty of 0.5, and a translation inhibition range of 10 to 11 nucleotides. A generalized summary of pipeline adopted for bioinformatics analysis is shown in Figure 3.1.
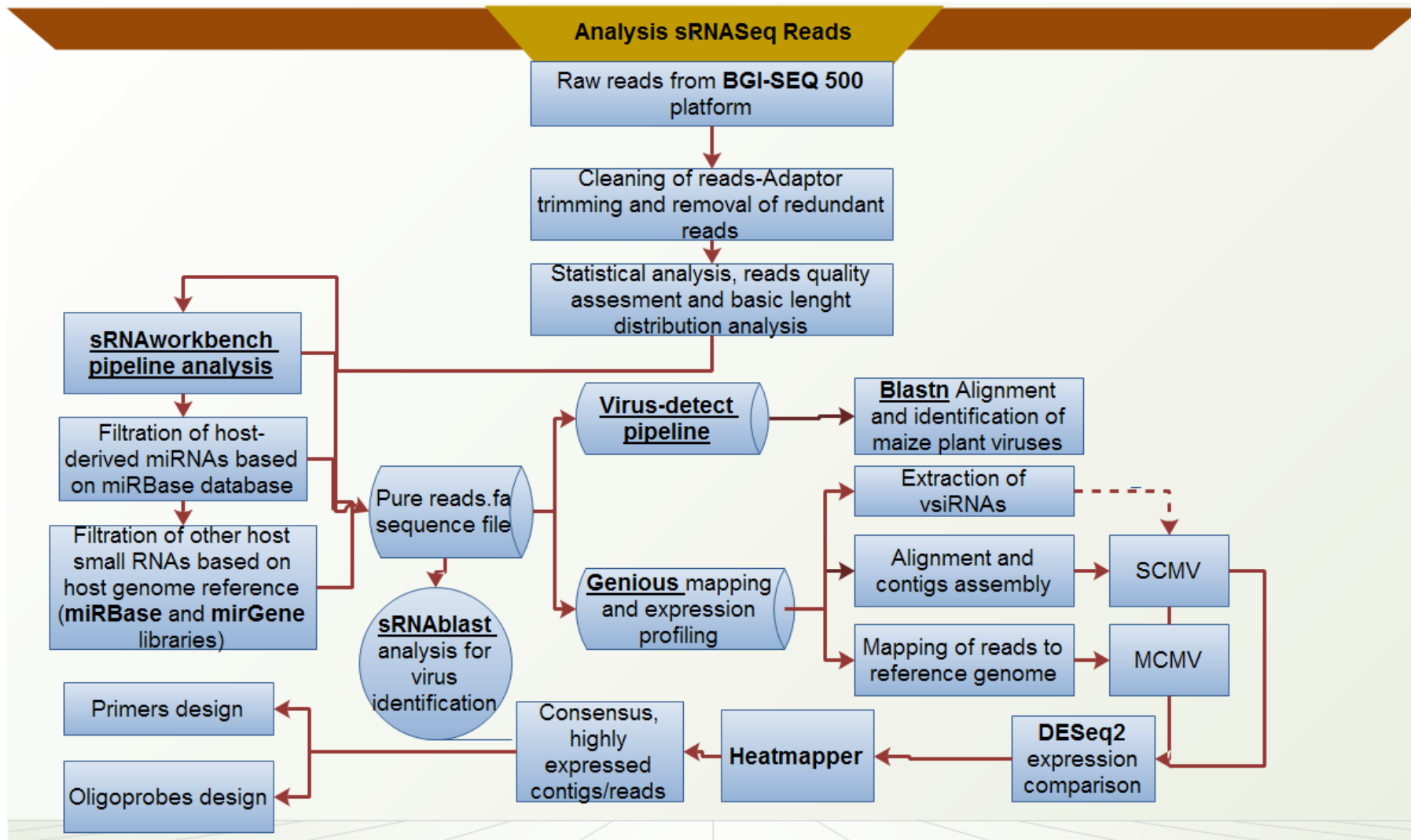
Figure 3. 1: Adopted pipelines and tools for Bioinformatics analysis

3.3.4. qPCR validation of markers

Sample collection from three regions (Bomet, Borabu and Kericho) was done in July 2018 from farms different from original sites where the samples were collected previously for sequencing. Eighteen samples per county were collected from at least six different farms, macerated and kept in DNA/RNA shield solution (ZymoResearch). Isolation of total RNA in triplicates was done using Directzol kit ^{TM} (ZymoResearch) following the manufacturer's instructions. Quality check was done prior to cDNA library construction using nanodrop 260/230 and 260/280 and only samples that had ratios of 1.8 and above were selected for library construction. cDNA preparation was carried out using 5XFirepol kit (Solis Biodyne) following the manufacturer's instructions. The resultant 18 different cDNA from the same regions were pooled into three representative samples per region for qPCR analysis. Confirmation of successful cDNA library construction was done using conventional PCR with amplification of PPDK gene from the maize genome.

 Universal qPCR kit (5x HOT FIREPol® EvaGreen® qPCR Mix Plus) was utilized in running qPCR using Roche Light Cycler 96-well real time qPCR following the manufacturer's instruction. These involved adding 0.5ul of forward and 0.5ul of reverse primer to 2ul of 5x HOT FIREPol mix and introducing 2ul of 50ng/ul cDNA. The volume of the mixture was then topped up to 10ul using RNase-Free-water. Replicase genes including Nlb replicase and replicase CDS in SCMV and MCMV genomic regions were used as reference genes for relative determination of expression levels for domains selected for SCMV and MCMV, respectively. Standard curves were generated by performing 9-times, tenth dilutions of the original samples. This was

repeated for calculation of each of the primer used in the expression profiling. The qPCR cycle parameters were set as: initial denaturation at 95ºC for 15 min, denaturation at 95ºC for 15 s, annealing at 60º-65ºC for 20 s (depending on the marker/primers) and elongation at 72ºC for 20 s. The plate design had three samples per region replicated three times with three negative control columns (4, 8 and 12) as shown in Appendix 4.1.

3.4. Detached Leaf Assay as a Pre- Screening strategy for tolerance to Maize Lethal Necrosis Disease in maize

3.4.1. Test materials, preparation of viral inoculum and maize growth conditions

Three MLND-tolerant maize genotypes CHMLND0093, CKIR11027, CKIR12032 (Das *et al*., 2015) provided by International Maize and Wheat Improvement Center (CYMMIT) Nairobi and susceptible genotypes *Namba nane* and H144 from Kenya Agriculture and Livestock Research Organization (KALRO) were used in this study. Purified and confirmed viral inocula of two MLND-causative viruses (MCMV) and Sugarcane Mosaic Virus (SCMV)) were provided by KALRO, Nairobi. Seeds were surface-sterilized using sodium hypochlorite solution (95% water: 5% NaOCl (Jik)) and sowed in autoclaved soil in pots. Plants were maintained in a glass house and upon tasseling, individual plants were self-pollinated. Seeds were collected at maturity and used to grow plants that were used for the DLA. All plants from which seeds were collected tested negative for both MCMV and SCMV using methods developed by Wangai *et al.,* (2012). To determine the media regime that would keep the detached

leaves in a green state, one tolerant (CKIR11027) and one susceptible (H144) variety were used. However, all 5 genotypes were inoculated *in vitro* with the mixture of the two viruses and evaluated for resistance. All *in vitro* detached leaf culture on media and viral inoculations were done under a laminar flow hood in controlled environment at the Biosafety Level II Plant Transformation Laboratory, Kenyatta University Kenya as stipulated in the user manual on handling of pathogens. In the experiment the 2nd youngest leaf of 28-, 42- and 63-day-old plants grown in autoclaved soil under controlled glasshouse conditions was used.

3.4.2.   Optimization of media for DLA

To optimize culture media regime capable of maintaining the detached maize leaves in green state for a long period of time and inhibit senescence, seven media enrichments were tested. The media comprised phyto agar (Duchefa, CAS. No 9002-18-0, sucrose (Duchefa, CAS No.57-50-1) and were supplemented with 10 mg/L kinetin (Duchefa, CAS No.525-79-1) and varying concentrations of gibberellic acid (Table 3.1). Individual media components were dissolved in distilled water, the pH adjusted to 5.8 then sterilized by autoclaving at 121°C for 15 minutes. Upon cooling, filter-sterilized gibberellic acid (in the indicated concentrations) and 250mg/L carbenicillin were added to the media before dispensing in sterile 90mm petri plates. Carbenicillin (antibiotic) was used to prevent the growth of any contaminating bacteria.

Table 3. 1:Formulation of media used for *in vitro* culture of maize leaves

47

Media ID	Phyto agar (mg/L)	Gibberellic acid (mg/L)	Kinetin (mg/L)	Sucrose (%)
1.	8	0	0	0
2.	8	0	10	0
3.	8	0	0	3
4.	8	10	0	0
5.	8	20	0	0
6.	8	50	0	0
7.	8	10	10	0
8.	8	20	10	0
9.	8	50	10	0
10.	8	10	0	3
11.	8	20	0	3
12.	8	50	0	3
13.	8	10	10	3
14.	8	20	10	3
15.	8	50	10	3
16.	8	0	10	3

3.4.3. Determination of media regime for senescence inhibition

3.4.3.1. Explant preparation, culturing and determination of senescence

Thirty, second youngest leaves of CKIR11027 and H144 genotypes were collected from 28, 42- and 63-day-old plants and washed under running tap water. The leaves were then surface-sterilized using 5% (v:v) sodium hypochlorite (0.384%) with 2 drops of tween 20 solution for 5 minutes. The tissues were rinsed three times using sterile distilled water and blotted dry using sterile filter papers (Whatman® qualitative filter paper, Grade 1). To prepare explants, sterile leaves were sliced into rectangular blocks measuring 3-4cm long and 1.5 to 2cm wide using a sterile scalpel blade under a laminar flow hood. The leaf explants were plated with the adaxial side in contact with the media on the listed media regimes (table 1), plates were sealed with parafilm and incubated in a growth chamber at 26-28°C with a photoperiod of 16hour light and

8hour darkness. Light was supplied by fluorescent tubes (PHILIPS-TL-D 18W/54-765). Three replicates of two explants were cultured per plate with five repetitions per media regime for each of the maize genotypes. Explants were monitored daily and senescence on the leaf blades scored as described by Twizeyimana *et al*. (2007).

3.4.3.2. Determination of chlorophyll content

To determine the amount of chlorophyll in explants maintained on each media regime, the method of Katoch (2011) was used. Both explants in each plate were pooled and 200mg of tissue collected into a clean centrifuge tube. The tissues were ground to a fine powder in liquid nitrogen then chlorophyll extracted using 2ml of 80% acetone. The mixture was vortexed then centrifuged at 10,000 RPM for five minutes. The lysate was transferred into a clean tube then absorbance measured at 645nm and 663nm wavelength for chlorophyll a and b, respectively using a spectrophotometer (Equipment 325-1000 nm 4 nm 722N-JoyFay). Chlorophyll concentration was calculated according to method by Katoch (2011).

$$\text{mg total chlorophyll/g tissue} = 20.2(A645) + 8.02(A663) + \frac{V}{1000 \times W}$$

Where

A = absorbance at specific wavelengths

V = final volume of chlorophyll extract

W = fresh weigh of tissue extracted

3.4.4. Determination of MLND resistance using DLA

3.4.4.1. Inoculum preparation

MLND inoculum was prepared using maize leaves independently identified as positive for SCMV and MCMV (provided by KALRO, Nairobi) following Gowda *et al*. (2015) method. To prepare a 100ml of crude inoculum, 2 grams of SCMV- and 8 grams of MCMV-infected leaves were combined (ratio of 1:4). The leaves were first washed under running tap water for 15 minutes then rinsed three times with autoclaved distilled water. They were then crushed using a mortar and pestle in Potassium Phosphate buffer solution at a pH of 8.0 prepared by dissolving 17.418g of dipotassium hydrogen orthophosphate dibasic (K_2HPO_4) and 13.609g of potassium phosphate (KH_2PO_4) in 100ml of autoclaved distilled water. The resulting inoculum was sieved through four layers of sterile cheese cloth to remove plant debris, aliquoted into 2ml Eppendorf tubes and stored at 4°C until plant inoculation.

3.4.4.2. Inoculation of leaf explants and plating

The best performing media regime (20mg/L gibberellic acid, 10mg/L kinetin supplemented with 3% sucrose) were used for incubation of detached leaf explants of CHMLND0093, CKIR11027, CKIR12032, Namba nane and H144 genotypes. The 2nd leaves were collected from the listed explants at the 28, 42 and 63 days after planting and explants prepared after surface sterilization as earlier described (section 3.4.3.1). A slight abrasion was then created on the leaves by gently brushing the adaxial side of the leaf surface with a fine autoclaved sandpaper (P-120) three times. Twenty (20ul) of the crude viral inoculum earlier prepared (3.4.1) was applied at each wounded site using a sterile pipette tip. The infected leaf explants were then plated with the adaxial side in contact with the media, plates sealed with parafilm and

incubated in a growth chamber under the aforementioned growth conditions. For controls, the abraded explants were inoculated with 20µl of potassium phosphate buffer solution from virus free maize plants. Cultures were monitored on a daily basis and scored for MLND severity as described in the CYMMIT MLND-severity chart (CYMMIT 2014). According to the scale provided in the chart, a score of 1 was given for no chlorotic symptoms, 2 for fine necrotic specs, 3 for chlorotic mottling throughout the explant, 4 for extensive chlorotic mottling and 5 for complete explant necrosis.

3.4.5. Data analysis

A generalized linear model (GLM) in Statistical Analysis Software (SAS) version 9.1 was used to analyse leaf senescence, chlorophyll content and disease severity. Analysis of variance (ANOVA) was used to compare the means and Tukey's HSD test ($p \leq 0.05$) used for mean separations. Graphical representation of the data was displayed in GraphPad prism version 7.0. Chlorophyll concentrations and senescence data were correlated using Pearson's Correlation matrix in GraphPad to confirm the reliability of the visual qualitative scores for senescence and to identify the best senescence inhibiting media enrichment. 'R' version 3.5.1 was used to generate a clustered heat map based on disease severity means.

3.5. Identification of alternate hosts for maize chlorotic mottle virus and sugarcane mosaic virus using small RNA markers

3.5.1. Sample collection

Five maize fields infected with MLND were sampled in Kericho, Nyamira, and Bomet counties. Samples of legumes and wild grasses were collected randomly within the infested fields as well as those proximal (at the edges) to MLN infected maize fields. A diagonal walk pattern was adopted when sampling. Both symptomatic and asymptomatic legumes and grasses were collected. Sampling was done between 1st August and 4th August 2018. Global positioning system (GPS) co-ordinates for the locations where samples were collected were recorded with samples assigned names based on the location as well as the local and scientific name with the help of an experienced botanist and Pl@ntNET android application. Photos were taken before the leaves of the potential alternate were macerated and kept in Eppendorf tubes with DNA/RNA-Shield preservative. A total of seventeen (17) potential alternate were collected (Table 3.2.)

Table 3. 2: List of potential alternate hosts of MLND causing viruses collected and analysed

S/No.	Botanical name	Scientific name	County
1.	oxalis	*Oxalis species*	Kericho, Nyamira, and Bomet
2.	Nut grass	*Cyperus rotundas*	Kericho, Nyamira, and Bomet
3.	Black jack	*Bidens Pilosa*	Kericho, Nyamira, and Bomet
4.	Couch grass	*Cynodon dactylon*	Kericho, Nyamira, and Bomet
5.	Mexican marigold	*Tagetes minula*	Kericho, Nyamira, and Bomet
6.	Macdonald's eye/gallant soldier	*Gallinsoga parviflora*	Kericho, Nyamira, and Bomet
7.	Chinese lantern	*Physalis alkekengi*	Nyamira, and Bomet
8.	Sorghum	*Sorghum bicolor*	Nyamira, and Bomet
9.	Finger millet	*Eleusine coracana*	Nyamira
10.	Wild finger millet	*Eleusine africana*	Nyamira, and Bomet
11.	Sugar cane	*Saccharum officinarum*	Kericho, Nyamira, and Bomet
12.	Common pigweed	*Amaranthus spp*	Kericho, Nyamira, and Bomet
13.	Pumpkin	*Cucurbita moschata*	Bomet
14.	Napier grass	*Pennisetum purpureum*	Kericho, Nyamira, and Bomet
15.	Proso millet	*P. milliacium*	Kericho and Nyamira
16.	Common bean	*Phaseolus vulgaris*	Nyamira, and Bomet
17.	Wandering Jew	*Commelina benghalensis*	Nyamira and Bomet

3.5.2. Total RNA isolation and cDNA library construction

Total RNA isolation was done using Directzol kit (ZymoResearch) following the manufacturer's instructions as described previously (section 3.3.2.). A two-step RT-PCR was adopted for the detection of the viruses in the alternate host. cDNA preparation was carried out using 5XFirepol kit (Solis Biodyne) following the manufacturer's instructions using random hexamers. Briefly, a master mix consisting of 1μl random hexamers (100 μM), 0.5μl of dNTP MIX (20 mM of each), 2μl of 10x RT Reaction Buffer with DTT, 1μl of FIREScript RT, 0.5μl RiboGrip RNase Inhibitor and 13μl of Nuclease-free H2O was made. A 2μl of 50μg/ μl volume of total RNA was then added to the master mix and the mixture ran on a RT-PCR with the following parameters: Primer annealing 25°C for 7 min, reverse transcription 50°C for 30 min and enzyme inactivation 85°C for 5 min. The cDNA was then kept on cold ice for use in the subsequent reactions. Confirmation of the success of the library construction step was done using PPDK gene for maize.

3.5.3. PCR Detection of MCMV and SCMV

Two validated markers from sRNASeq and qPCR including SCMV P3 and MCMV capsid protein marker MCMV116 (section 3.3.4) were then used for the detection of the viruses by using 5x FIREPol® Master Mix Ready to Load kit with specific components being 4μl of the 5xFIREPol, 0.5μl of forward and 0.5 μl of the reverse primer and 4μl of cDNA sample. The mixture was then topped up to 20μl with RNAse free water. PCR specific parameters included Initial denaturation at 95ºC for 3 minutes, Denaturation at 95ºC for 30seconds, annealing at 50ºC for 30seconds, elongation at 72ºC for 1minute and the reaction repeated for 25 cycles. Final

elongation was done at 72°C for 7 minutes. The presence of a distinct 116bp band was translated as positive for MCMV and a 164bp band was considered positive for SCMV. Position of the band was further confirmed by the positive controls for both MCMV and SCMV.

CHAPTER FOUR

RESULTS

4.1. SRNA profiling using next-generation sequencing for detection of MLND causal viruses

4.1.1. Small RNA sequencing statistics

An average of 25 million clean reads were obtained per region after the adaptor, contamination and low-quality reads were removed from raw data (Table 4.1). The reads ranged from 15-50 nucleotides long with Guanine-cytosine content ranging from 50-55%.

Table 4. 1:Sequencing statistics for sRNAseq of MLND samples from Nyamira (Borabu), Kericho and Bomet counties-Kenya (2018) using BGISEQ500 Next-generation sequencing.

Sample Name	Clean Reads	Clean bases	Read length (bp)	Q20 (%)	GC (%)
Nyamira	25,122,822	538,477,059	50	99.13%	53.58%
Kericho	25,590,873	567,780,623	50	99.08%	53.21%
Bomet	24,657,131	531,883,920	50	99.11%	55.09%

Samples originating from the same locale were pooled and therefore each sample represented the county of origin. Q20(%) represents Phred quality score based on the equation $Q = -10\log10(e)$.

Across the three counties, highest percentage of reads ranging from 18 to 24 nucleotides long were observed (Figure 4.1.). Notably, the 21-22 nucleotides long small RNAs covered the highest percentage.
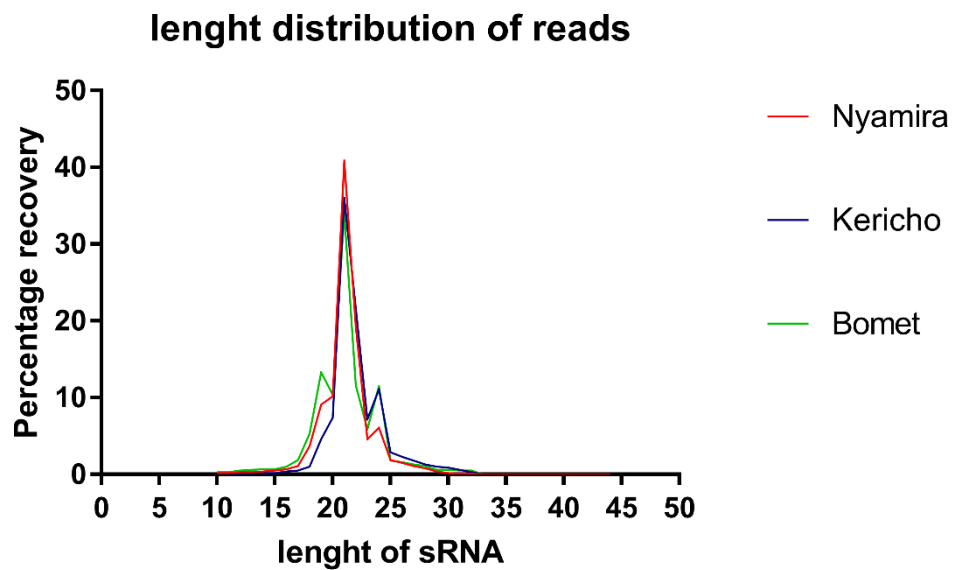
Figure 4. 1: Cleaned reads length distribution for Nyamira, Kericho and Bomet counties-Kenya (2018).

Horizontal axis represents the length of reads while the vertical axis describes the percentage of reads of a given read length. Samples pooled from the three counties Borabu (Red line), Kericho (Blue line) and Bomet (Green line) are represented.

4.1.2. Mapping of small RNAs to MiRbase database

Mapping of small RNA reads to mirBase database returned only 1.05, 1.23 and 1.15% matches to mature sense miRNAs for Bomet, Kericho and Borabu respectively (Table 4.2). Mature sense miRNAs are the only species that had the highest mapping percentages. Even though some of the hairpin antisense, hairpin sense and mature antisense miRNAs were also detected, their relative abundance was extremely low (Table 4.2). From the total mature sense miRNAs, only five classes of microRNAs

57

including zma-miR159a-3p, zma-miR168a-5p, zma-miR166a-3p, zma-miR167e-5p and zma-miR444a were highly expressed consistently across the three regions based on the Read per Kilobase Million (RPKM) values (Table 4.3). zma-miR159a-3p was however the most frequently mapped miRNA.

Table 4. 2:Raw reads mapped and their percentage representation to mirBase database

Name of sRNA	RC			RC perc		
	BOMET	KERICHO	NYAMIRA	BOMET	KERICHO	NYAMIRA
mature sense	257,829	313,777	290,027	1.05	1.23	1.15
hairpin antisense	985	1400	1142	0	0.01	0
hairpin sense	4343	7164	7448	0.02	0.03	0.03
mature antisense	132	187	249	0	0	0
un-assigned	24,391,743	25,266,421	24,822,195	98.93211	98.73958	98.8103

Values represents raw read counts (RC) per region and RC percentages are percentage fractions of the reads mapped to the total number of reads tested. Values assigned to antisense and sense for mature and hairpin miRNAs represents mapped reads. Un-assigned represents reads that did not map to the host *mirBase* miRNA database.

Table 4. 3:Five highly represented miRNAs in Nyamira, Bomet and Kericho counties-Kenya, 2018.

mirBase annotation	Read count (SA)		
	BOMET	NYAMIRA	KERICHO
>zma-miR159a-3p GAGCTCCTATCATTCCAATGA	120409	52348	90890
>zma-miR168a-5p TCGCTTGGTGCAGATCGGGAC	22369	36371	55967
>zma-miR166a-3p GGAATGTTGTCTGGCTCGGGG	11788	24696	20280
>zma-miR167e-5p TGAAGCTGCCAGCATGATCTG	11389	9361	16180
>zma-miR444a TGCAGTTGTTGTCTCAAGCTT	9215	9857	18690

mirBase annotation represents the identity assigned to specific miRNA in *mirBase* database. Values presented per county are the reads that matched the identified miRNA in the database.

4.1.3. PsRNAtarget analysis based on Host-derived mature miRNA

Three main miRNAs derived from the host mature miRNA (zma-miR167b-3p, zma-miR168b-3p and zma-miR528a-3p) were identified using psRNA-Target analysis (Table 4.4.). All the 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 were identified to target genomic regions HC-Pro, capsid protein, CI peptide, 6K2 and P3. Both translation and cleavage inhibition were both present with low unpaired energy requirement (Table 4.5). Six of the miRNAs targeted the HC-Pro region, a viral suppressor protein that have been recognized for its inhibitory activity on RNA interference mechanism.

4.1.4. Families of maize viruses detected through sRNAseq profiling

Assembled contigs from the hist-filtered reads identified five main families of viruses in the sequenced samples. These include Machlomovirus (Maize chlorotic mottle virus), Polerovirus (Maize yellow mosaic virus), Potyvirus (Sugarcane mosaic virus), Mastrevirus (Maize streak virus) and maize associated totivirus (Figure 4.2). Full assembly of genome was achieved for MCMV in all regions and SCMV in Bomet and Nyamira county. A complete maize streak virus genome was also assembled from samples collected in Kericho and partially for samples from Bomet and Nyamira

county. Maize yellow mosaic virus and maize associated totivirus genome assembly was generally poor with no complete assembly obtained in samples collected from all the three regions.

4.1.5.   Phylogenetic analysis of MCMV and SCMV isolates

Alignment of selected contigs of SCMV from each region to previous isolates from Kenya, Rwanda, U.S.A and China showed a high evolutionary divergence (Figure 4.3). A close relationaship between isolates from China, Bomet and Nyamira whereas the isolate from Kericho was distantly related to the rest of the SCMV isolates. MCMV isolated showed close relationships with all isolates clustering together (Figure 4.4).

Table 4. 4: Prediction of Host derived miRNAs targeting MCMV domains for samples collected from Nyamira, Kericho and Bomet counties-Kenya, 2018.

miRNA_Acc	miRNA_aligned_fragment	Target_aligned_fragment	Exp	UPE$	Inhibition	Target_Desc.
zma-miR167b-3p	GAUCAUGCUGUGACAGUUUCACU	CAACAAAC-GUCACGGUAUGUUC	4.5	-1	Cleavage	[locus_tag=MCMV_ [protein=replicase]
zma-miR167b-3p	GAUCAUGCUGUGACAGUUUCACU	CAACAAAC-GUCACGGUAUGUUC	4.5	-1	Cleavage	[locus_tag=MCMV_ [protein=replicase]
zma-miR167b-3p	GAUCAUGCUGUGACAGUUUCACU	CAACAAAC-GUCACGGUAUGUUC	4.5	-1	Cleavage	[locus_tag=MCMV_ [protein=replicase]
zma-miR168b-3p	CCCGCCUUGCAUCAAGUGAA	UGGAUUUGAUGCAAGUCGGU	5	-1	Cleavage	[locus_tag=MCMV_ [protein=replicase]
zma-miR528a-3p	CCUGUGCCUGCCUCUUCCAUU	UGUGGACGAAGAAGGCACAGU	5	-1	Translation	[locus_tag=MCMV_ [protein=replicase]
zma-miR528a-3p	CCUGUGCCUGCCUCUUCCAUU	UGUGGACGAAGAAGGCACAGU	5	-1	Translation	[protein=replicase-associated protein]
zma-miR528b-3p	CCUGUGCCUGCCUCUUCCAUU	UGUGGACGAAGAAGGCACAGU	5	-1	Translation	[locus_tag=MCMV_ [protein=replicase]
zma-miR528b-3p	CCUGUGCCUGCCUCUUCCAUU	UGUGGACGAAGAAGGCACAGU	5	-1	Translation	[protein=replicase-associated protein]

miRNA-Acc are the miRNA accession in *mirBase* database. Expectation value (Exp) is the penalty for the mismatches between mature small RNA and the target sequence. Unpairing Energy (UPE) is the maximum energy required to unpair the target structure and allow accessibility of the mRNA target site to small RNA.

Table 4. 5: Target prediction of Host-derived miRNAs targeting SCMV domains for MLND infected samples collected from Nyamira, Kericho and Bomet counties-Kenya, 2018.

. miRNA_Acc.	miRNA_aligned_fragment	Target_aligned_fragment	Exp	UP E$	Inhibition	Target_Desc.- locus_tag
zma-miR156i-3p	GCUCACUGCUCUAUCUGUCACC	AGCGAGAGUUAGCGCAGUGAGG	4.5	-1	Translation	SCMV_ [protein=P3-Pro Peptide]
zma-miR166k-5p	GGAUUGUUGUCUGGCUCGGGG	GGCCAAGUUAUACAAUAAUCU	4.5	-1	Translation	SCMV_[protein=HC-Pro protein peptide]
zma-miR166n-5p	GGAUUGUUGUCUGGCUCGGUG	GGCCAAGUUAUACAAUAAUCU	4.5	-1	Translation	SCMV_ [protein=HC-Pro protein peptide]
zma-miR2275a-3p	UUUGUUUUCCUCCAAUAUCUCA	GACGAUGAUGGACGGAGACGAA	4.5	-1	Translation	SCMV_ [protein=Capsid protein]
zma-miR166j-5p	GGUUUGUUUGUCUGGUUCAAGG	UGCGACACUAGACGAGCAAAUC	5	-1	Cleavage	SCMV_ [protein=Nla-Vpg]
zma-miR167b-3p	GAUCAUGCUGUGACAGUUUCAC	CCAAAAACUUUUAC-GCAUGAUU	5	-1	Cleavage	SCMV_ [protein=HC-Pro protein peptide]
zma-miR169o-3p	GGCAGGUCUUCUUGGCUAGC	UUUGGCAAAGAAAACCUCCC	5	-1	Cleavage	SCMV_ [protein=P3-Pro Peptide]
zma-miR172b-5p	CAGCACCAUCAAGAUUCACA	AGUAAAGUUUGAUGGUUCUA	5	-1	Cleavage	SCMV_ [protein=HC-Pro protein peptide]
zma-miR172d-5p	CAGCACCAUCAAGAUUCACA	AGUAAAGUUUGAUGGUUCUA	5	-1	Cleavage	SCMV_[protein=CI Peptide]
zma-miR2275d-3p	UUUGUUUUCCUCUAAUAUCUCA	GACGAUGAUGGACGGAGACGAA	5	-1	Translation	SCMV_[protein=Capsid protein]
zma-miR395l-5p	GUUCCUUCCAAACACUUCACCA	CUUGGAAGAGUU-GGAAGGAAC	5	-1	Translation	SCMV_[protein=CI Peptide]
zma-miR395o-5p	GUUCUCUUCAAGCACUUCACGA	AGAGGAAGUUCUUGGUGAGAAU	5	-1	Cleavage	SCMV_ [protein=6K2]
zma-miR396c	UUCCACAGGCUUUCUUGAACUG	CCAUUGAGCAAAGCUUGUGUGA	5	-1	Cleavage	SCMV_ [protein=HC-Pro protein peptide]
zma-miR396d	UUCCACAGGCUUUCUUGAACUG	CCAUUGAGCAAAGCUUGUGUGA	5	-1	Cleavage	SCMV_ [protein=HC-Pro protein peptide]
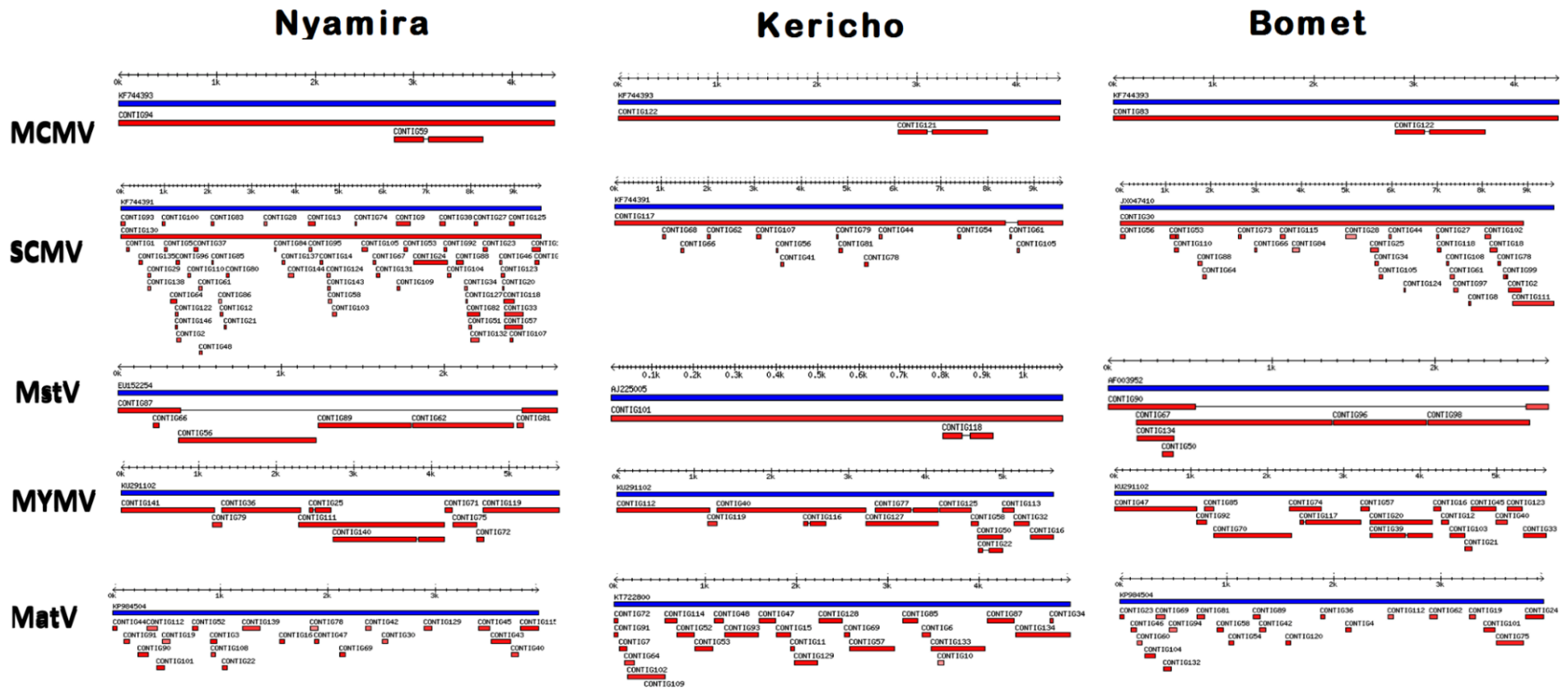
Figure 4. 2: Maize viruses detected by alignment of assembled contigs from sRNA *de novo* assembly

Blue bar represents the Refseq genome of the detected virus while red bars represents the *de novo* assembled contigs. Five families of maize viruses namely Maize Chlorotic Mottle Virus (MCMV), Sugarcane Mosaic Virus (SCMV), Maize –associated totiviruses (MatV), Maize Streak Virus (MSV) as well as Maize Yellow Mosaic Virus (MYMV) were identified. Several sugarcane mosaic virus strains were detected across the three regions.
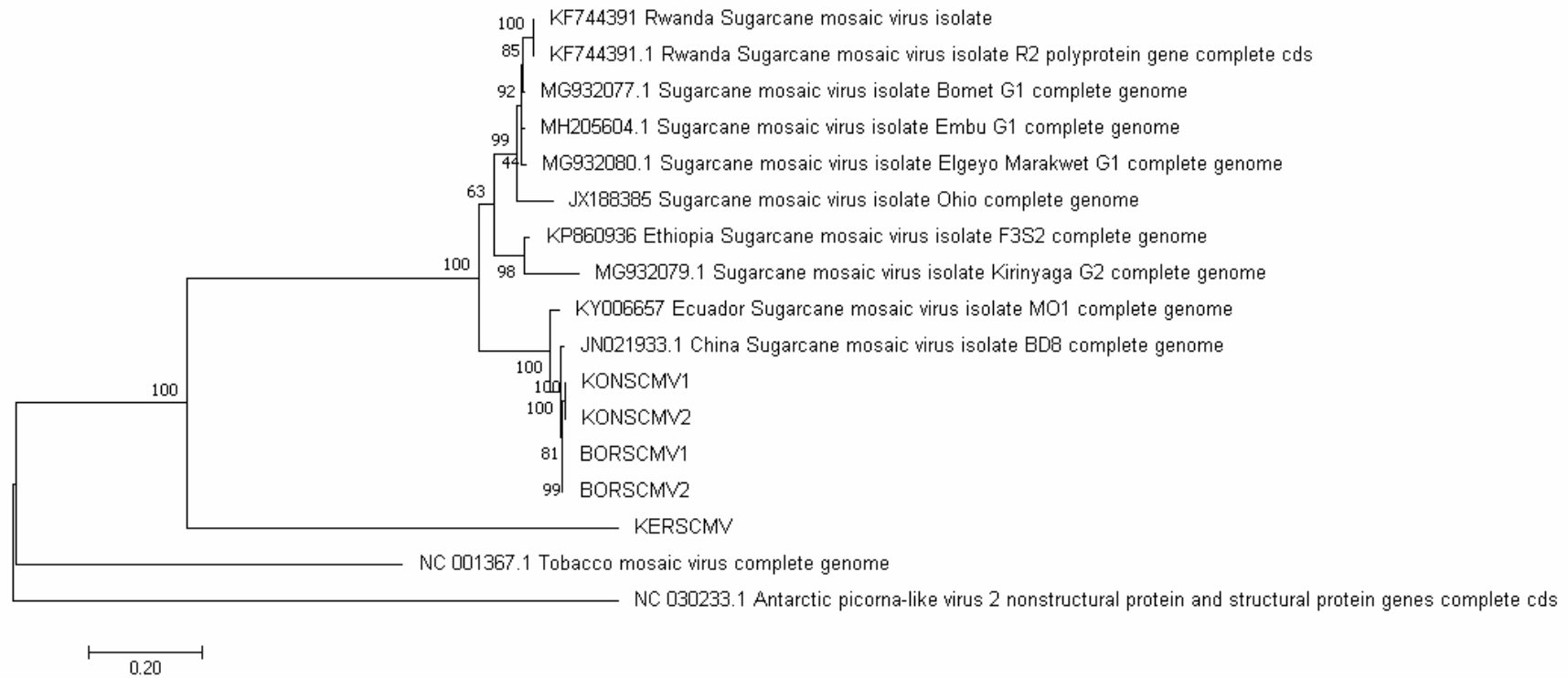
Figure 4. 3:Molecular Phylogenetic analysis of assembled genomes of SCMV from Nyamira, Kericho and Bomet counties- Kenya, 2018.

Bomet SCMV (KONSCMV), Nyamira SCMV (BORSCMV) and Kericho SCMV (KERSCMV). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 17 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 5422 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.
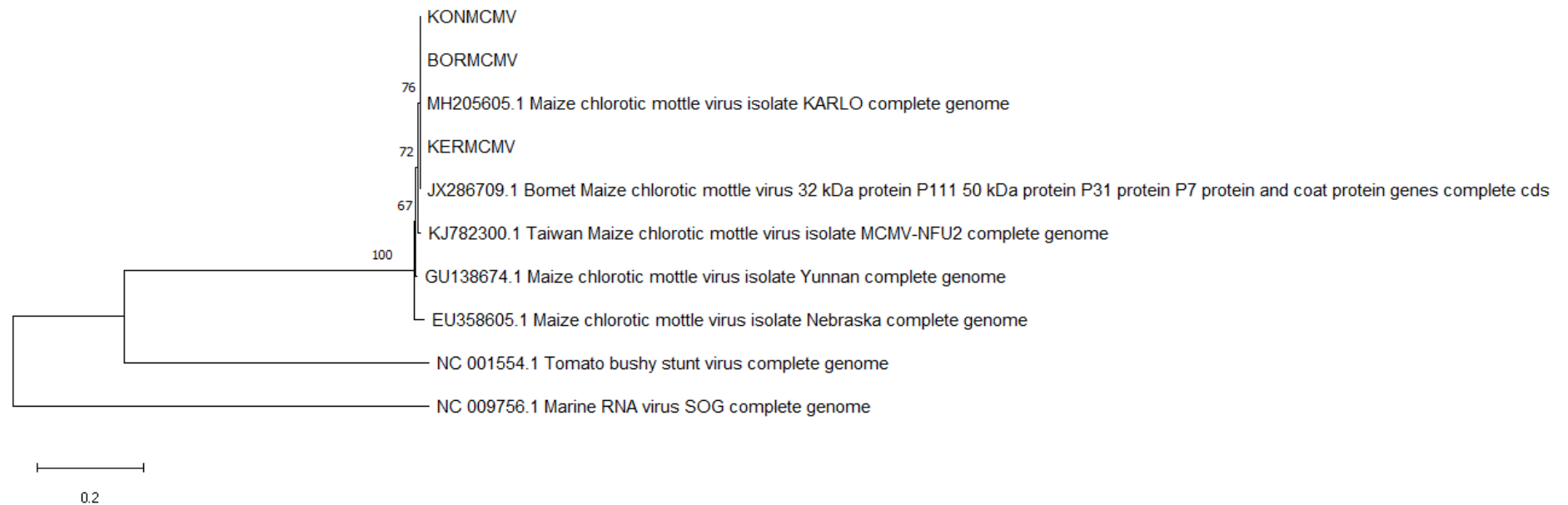
Figure 4. 4: Molecular Phylogenetic analysis of assembled genomes of MCMV from Nyamira, Kericho and Bomet counties- Kenya, 2018.

Bomet MCMV (KONMCMV), Nyamira MCMV (BORMCMV) and Kericho MCMV (KERMCMV).

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-17922.68) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 10 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 4226 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

4.1.6. Identification of highly expressed domains of Viruses

4.1.6.1. P7 and capsid protein CDS domains upregulated in MCMV

A total of 3046799, 6537375 and 4865520 reads from Bomet, Nyamira and Kericho respectively mapped onto MCMV *Refseq* genome. Small RNA reads mapped across the entire genome of MCMV differently with coat, p7a and p7b CDS being the most frequently recalled domains (Figure 4.5). Replicase domain, which spans the replicase CDS and p32 CDS domains on average, was moderately expressed. Individually, replicase-associated peptide and P32 had the lowest expression.

4.1.6.2. Expression of P3 and capsid protein domains of the SCMV

A total of 1474576, 1174148, 506721 raw reads mapped to the SCMV genome with PIPO CDS, a domain within P3 CDS read through and the capsid protein were the most frequently mapped domains in all the three regions. This is indicated by the heat map red-coded colour sections (Figure 4.6). Nlb replicase peptide was moderately expressed in samples collected from the three counties whereas Nla domains and 6K2 had the lowest expression.
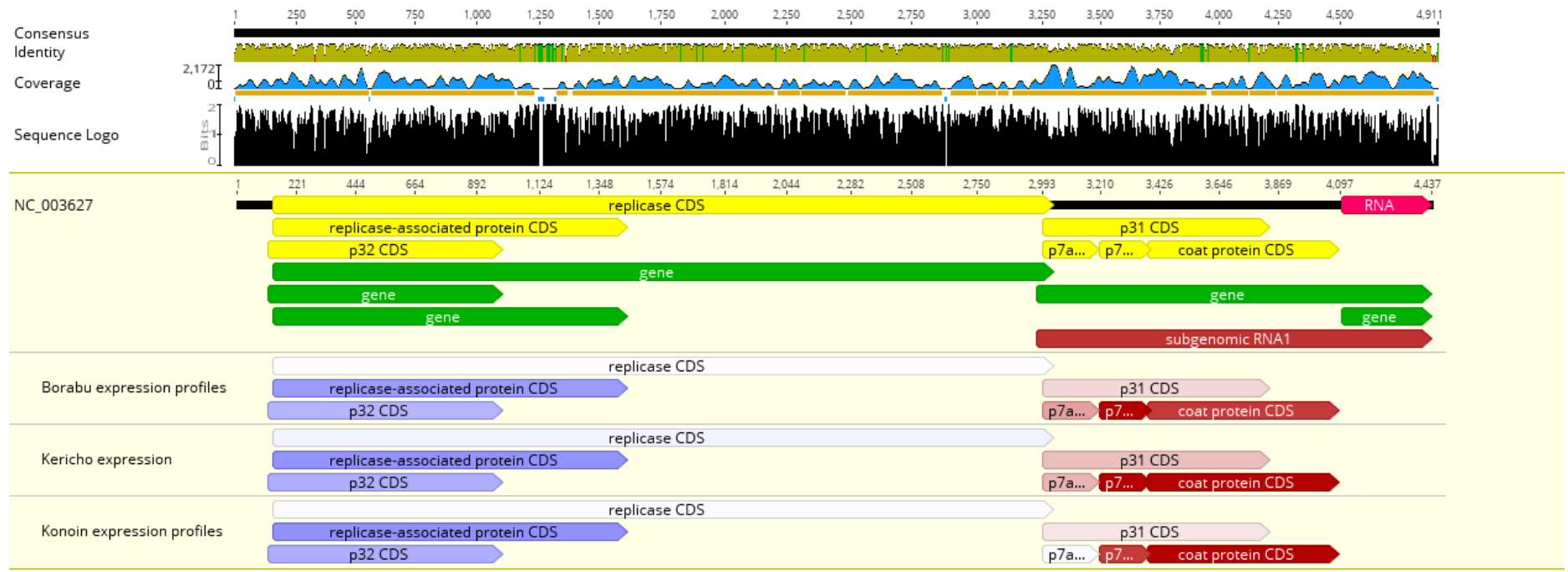
Figure 4. 5:Differential expression of different MCMV domains based on RPKM values from sRNA mapping

NC-003627-*Refseq* for Maize chlorotic mottle virus' genomic CDS is represented by the yellow bars. Expression profiling was done with reference to the replicase-associated protein CDS, p32 CDS, P7a, P7b and coat protein CDS. Red-color coded regions are the up-regulated domains while blue-color coded regions are the least expressed domains.
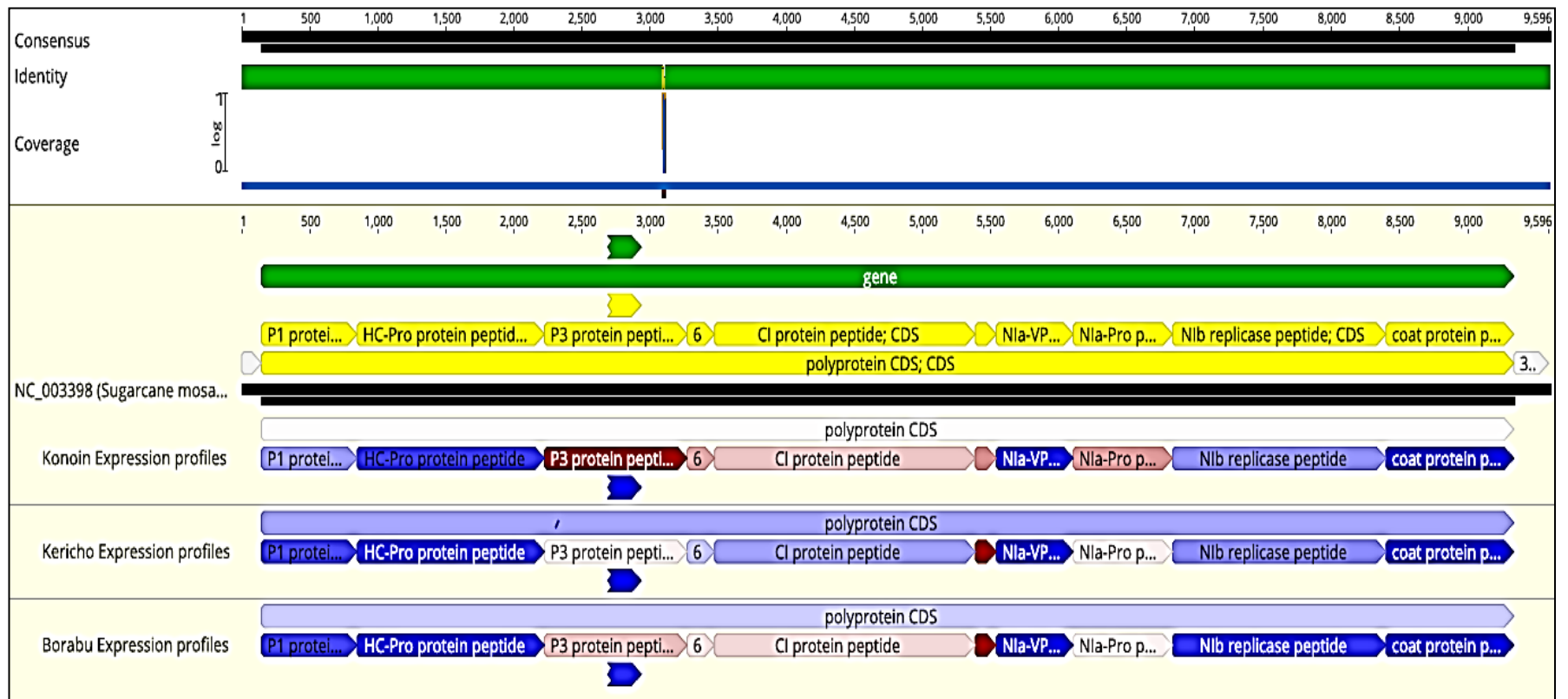
Figure 4. 6: Differential expression of different SCMV domains based on RPKM values from sRNA mapping.

NC-003398, the annotated *Refseq* genome of Sugarcane mosaic virus is illustrated in the first row (yellow and green bars). Mapping of small RNAs were done with reference to the P1 pro, HC-Pro, P3 protein, CI protein, 6K2, Nlb, Nla-pro peptide and coat protein peptide CDS. Highly expressed and conserved domains in Bomet (Konoin), Nyamira (Borabu) and are represented by red-color coded CDS.

Table 4. 6: Mapping statistics for MCMV and SCMV per coding region of the genome   (CDS)

Sugarcane mosaic virus

Name	RPKM		
	BOMET	**NYAMIRA**	**KERICHO**
capsid protein peptide	142012.3	111250.9	121076.4
NIb replicase peptide	55548.97	46676.24	51511.89
NIa-Pro protein peptide	18659.39	30817.11	48993.4
NIa-VPg protein peptide	59455.91	57550.03	18815.86
6K2 protein peptide	282.21	38991.62	57423.09
CI protein peptide	39093.29	50211.04	59458.85
PIPO CDS	176190.5	190226.4	206289.6
polyprotein CDS	66951.05	67408.44	63872.32

Maize chlorotic mottle virus

Name	RPKM		
	BOMET	**NYAMIRA**	**KERICHO**
capsid protein CDS	214250.7	183621.1	195761
p7b CDS	284915	247756.5	249106.5
p31 CDS	178264.6	164304.6	162350.9
p7a CDS	151277.5	189597.9	152675.9
replicase CDS	143047.3	151468.9	150401.4
replicase-associated protein CDS	89050.27	94397.44	95034.43
p32 CDS	91060.21	92634.34	95673.62

4.1.7. Molecular marker identification, design and qPCR validation

A total of eight highly expressed and conserved regions (Table 4.7) were used to design primers. For MCMV, markers spanning capsid protein, P31 and p7b domains were designed and sent for synthesis. PIPO (P3), CI, p70 and Nla pro domains of SCMV were highly expressed and were therefore selected for the design of markers.

Table 4. 7:sRNA Markers from mapping statistics

Virus	S/No.	DOMAIN	Forward Primer 5'-3'	Reverse primer 5'-3'
MCMV	1.	P7b (116)	CAGTGTGTGCAGGGATTGAC	GCGCACAGAGTTGAACACAAT
	2.	Capsid-Protein (245)	TACTCCGTGTCCTACCGTGG	ACACTTTGGATTGGCAGGAC
	3.	Capsid-Peptide (113)	GCCGCTGACAAGCAAATGTA	CCACGGTAGGACACGGAGTA
	4.	P31(113)	GCCGCTGACAAGCAAATGTA	CCACGGTAGGACACGGAGTA
	5.	P31(139)	GTTGCCGTTTATCACGAGCC	GGATTTTGCTTCCACGCTACG
SCMV	6.	P3(164)	GCGTGTGTTCAGAGTCAA	ACCGTTGCGGAGTTTGGTTA
	7.	P3(88)	CACCCCCTACCGTGCTTATC	GCGACTCCTTGGTTCTTGAC
	8.	P70(94)	GACCCACAAAAGAGCGATGC	TGACTGTTGGGAATGGCACA
	9.	CI-Peptide (83)	GAACAGTGCCACACTACCGA	AGCGAGTTCAATGCTGACCA
	10.	CI-Peptide (136)	TGGCGAGCGTATTCAGAGAC	TGCGAAACACATGAAAGCCG
	11.	Nla-pro (79)	TCAAGGGTGTGCTTAGTCGG	GTGCTGTGACGCTACTCTC C
	12.	Nla-pro (76)	GGGTGTGCTTAGTCGGAGTG	GGTGCTGTGACGCTACTCTC

All the listed markers, which served as primers for qPCR, had GC content of 47-52%, self-complementarity of less than 2 and a product size of less than 250bp (showed in brackets). The length of each marker was 21-22 nucleotides.

4.1.8. Standard curve for reference replicase genes

a. Replicase CDS MCMV reference gene   **b.** Nlb Replicase SCMV reference gene
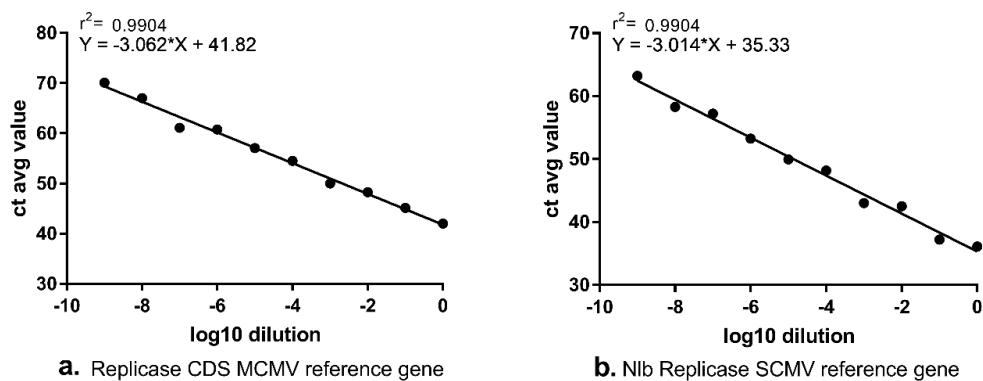
Figure 4. 7: Efficiencies of primers for reference genes selected

4.1.9. Primer efficiencies

The quantitative PCR efficiencies for individual primer sets were determined using the standard curve slopes achieved from pooling cDNA from each of the three regions to generate a set of three independent samples per gene (Table 4.8). Regression analysis of the data obtained showed high efficiencies (0.94-1.14) and high linearity of r2>0.90 as shown in the table

Table 4. 8:Efficiencies of individual primers tested as markers

Gene	Efficiency	R^2	Slope	Y-intercept
MCMVp7b	0.946607	0.9645	-3.458	12.986
MCMVCP245	1.052783	0.9905	-3.2016	14.402
MCMVCP113	1.039847	0.9913	-3.23	12.5194
MCMVCP139	1.086838	0.9911	-3.13	15.695
SCMVNla	1.053063	0.9867	-3.201	32.615
SCMVP3	1.016506	0.9989	-3.283	23.249
SCMV6KI	1.127971	0.9768	-3.0491	30.575
CI	1.18903	0.9557	-2.939	35.677

Analysis of variance on the relative expression ($2^{\Delta\Delta Ct}$) showed no significant differences (p= 0.0857) in relative expression for the tested markers for MCMV normalized to the replicase gene of MCMV (Figure 4.8a.). Additionally, no significant difference in terms of regions (p= 0.4044) was observed. Significant differences (p=0.0035) in expression of the tested domains for SCMV normalized to Nlb replicase gene of SCMV was observed (figure 4.8b.) with no differences (p=0.5054) in terms of regions.
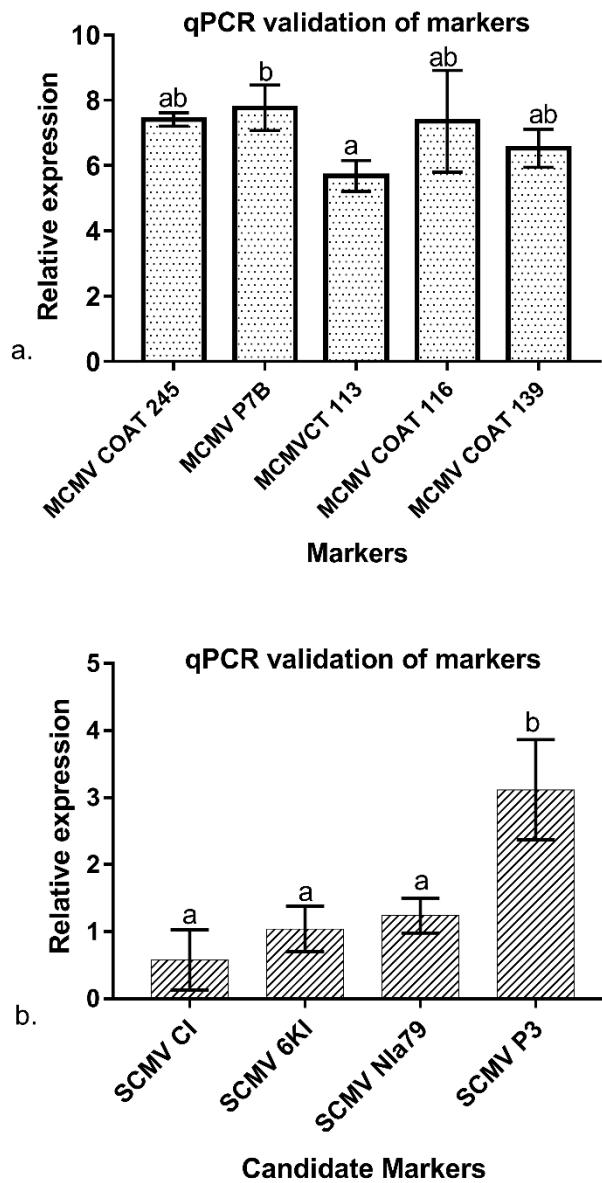
Figure 4.8: Relative expression for Marker-target Genes to replicase control genes. (a)-MCMV markers, (b)-SCMV markers

No significant differences observed for MCMV marker candidates' relative expressions even though MCMVP3 had the highest ratio levels as compared to the rest of the genes. SCMV marker SCMVP3 was differentially expressed as compared to other domains.

4.1.10. Specificity test for selected markers MCMVP7B, MCMV113 and SCMVP3
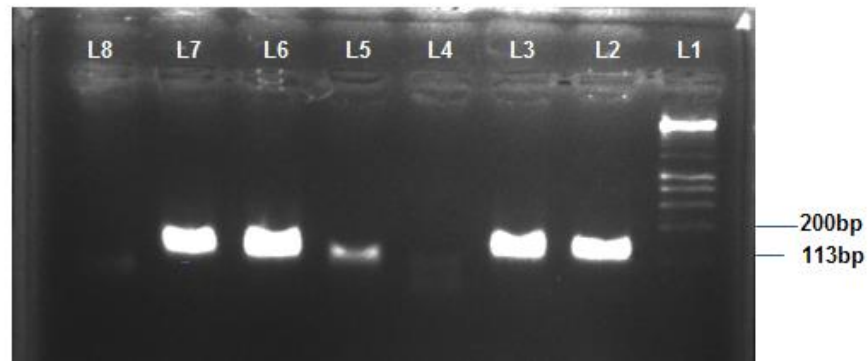
Plate 4.1: Gel product for qPCR for MCMV markers MCMV139, MCMV113 and MCMVP7B

L8 is an empty well, L7, L6 represents a 139bp product of MCMV139 marker, well L5, L4 represents qPCR product of MCMV113 and L3, L2 is a qPCR product of MCMVP7B with a size of 116bp.
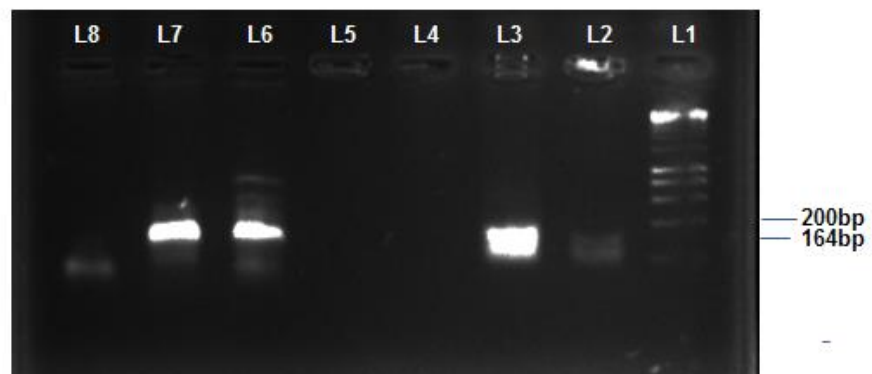
Plate 4. 2: Gel product for qPCR of SCMVP3 and SCMV6KI marker. L2, L3 are 136bp products of SCMV6CI marker, L4, L5 are SCMVCI products while L6, L7 are the 164bp products of SCMVP3 marker. L8 is a 79bp product of SCMV Nla79.

4.2. Detached Leaf Assay as a Pre- Screening strategy for tolerance to Maize Lethal Necrosis Disease in maize

4.2.1. Effect of media regimes on senescence inhibition

The 7 media formulations used showed varied effects on leaf senescence up to 21 days, post detachment. For instance, leaves incubated on culture media not treated with any hormones started senescing by showing characteristic yellowing and browning as early as three days post detachment. The leaves plated on media treated with sucrose, agar and without GA3 or kinetin showed accelerated senescence in comparison to those incubated on media treated with hormones. Leaves maintained on media with a combination of GA$_3$, sucrose and kinetin exhibited delayed senescence. The rate of senescence varied greatly depending with the specific media. Leaves plated on media with 20mg/L GA3, 10mg/L kinetin and sucrose remained green hence showing delayed senescence beyond 21 days of after culture (Plate 4.3). Percent senescence was scored further at 21 days after incubation noted significant differences (p≤0.05) among the media regimes tested (Figure 4.9). In the susceptible genotype H144, media formulations G20K, AS, G50K, AW and AK resulted in the highest leaf senescence of 91.67 ±3.445%, 84.167 ±3.786%, 81.67 ±2.706%, 75.83 ±4.84% and 85.83 ± 4.345%, respectively at 21 days, post incubation. On the other hand, G20KS resulted in the lowest leaf senescence producing an average of 8.33 ±1.978% at 21 days, post incubation (Figure 4.9). For the tolerant genotype CKIR11027, media AS, G50K, G20K, AK and AW resulted in the highest levels of leaf senescence of 91.67 ±3.445%, 85.83 ±4.345%, 84.167 ±3.786%, 83.33 ±3.761% and 86.667 ±2.706%, respectively at 21 days, post incubation. The leaves of genotype

CKIR11027 on G20KS media resulted in 7.083±1.681% senescence which was the lowest level reported in this study (Figure 4.9). For this reason, media formulation G20KS was selected for subsequent experiments involving infection with SCMV and MCMV
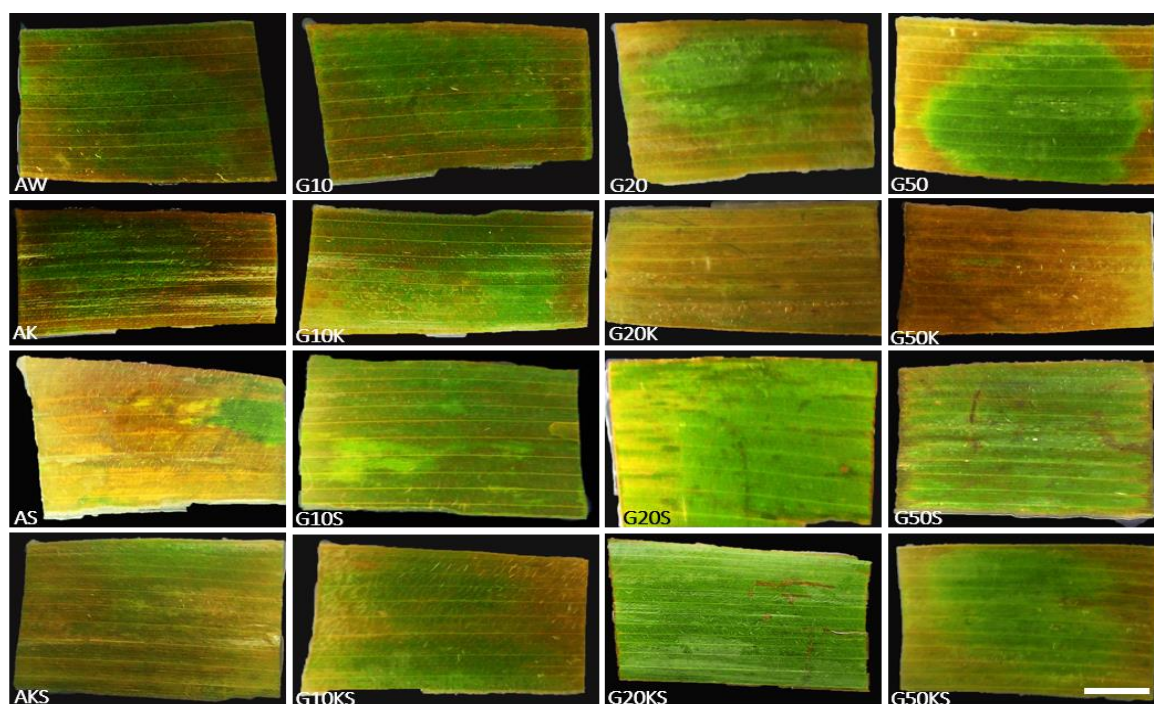
Plate 4. 3: Profile of maize leaves (H144) cultured on various media regimes showing levels of senescence 21 days post-detachment.

AW-media without any enrichment. Media with Gibberellic acid at 10mg/L (G10), 20mg/L (G20) 50 mg/L (G50). Media with kinetin 10mg/L (AK), kinetin at 10mg/L with 10mg/L gibberellic acid (G10K), kinetin at 10mg/L with 20mg/L gibberellic acid (G20K), kinetin at 10mg/L with 50mg/L gibberellic acid (G20K). AS-media enriched with 3% sucrose, G10S-3% sucrose with 10mg/L GA3, G20S-3% sucrose with 20mg/L GA$_3$ and G50S 3% sucrose with 50mg/L GA3. AKS- media enriched with 3% sucrose and 10mg/L kinetin, G10KS- 3% sucrose and 10mg/L kinetin, G20KS- and sucrose with 20mg/L kinetin and G50KS- sucrose with 50mg/L kinetin. Bar=1cm.
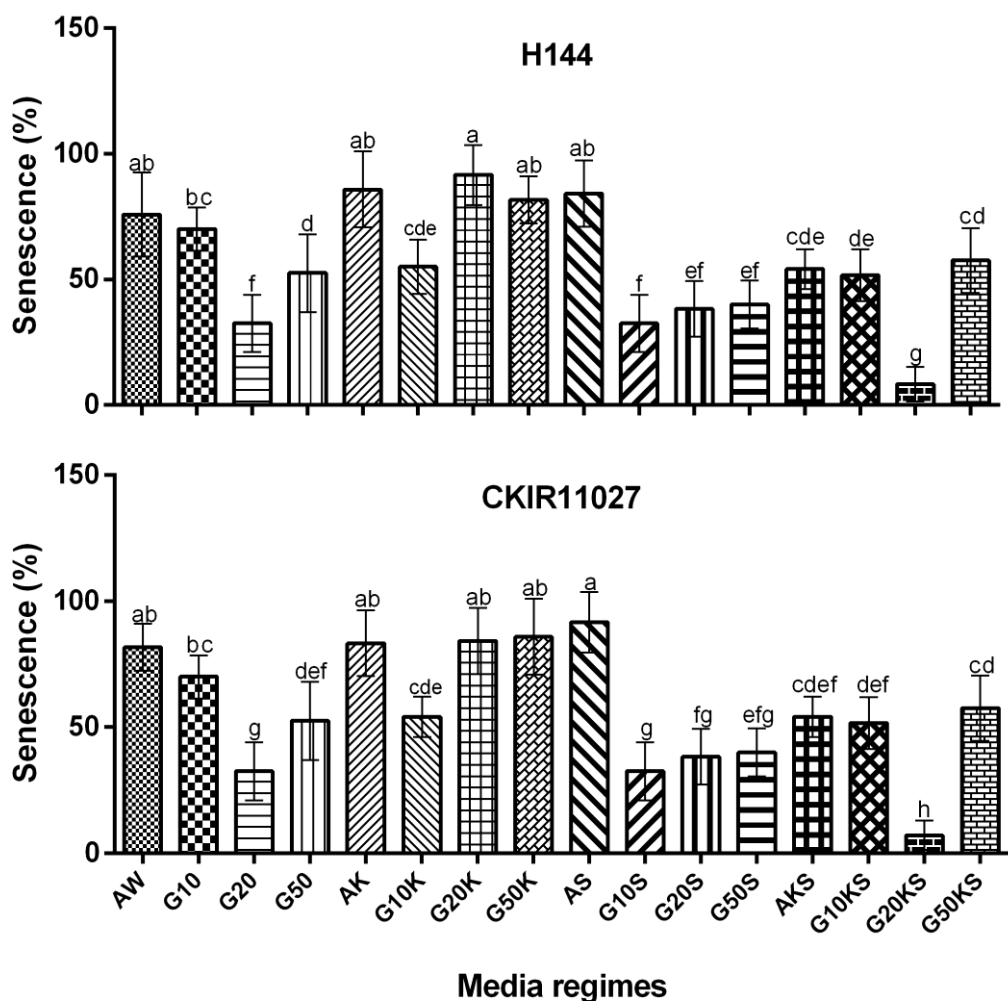
Figure 4.9: Effect of media formulations on senescence of detached leaves of 2 maize genotypes H144 and CKIR11027 21 days post detachment.

Data represents means with standard deviations of 3 replicates (n=30) analysed using SAS version 9.2. Letters on top of each bar indicate mean separations according to Tukey's HSD test (p≤ 0.05). Means followed by the same letter are not significantly different. AW-media without enrichment. Media with Gibberellic acid at 10mg/L (G10), 20mg/L (G20) 50 mg/L (G50). Media with kinetin 10mg/L (AK), G10K-kinetin at 10mg/L with 10mg/L gibberellic acid, G20K-kinetin at 10mg/L with 20mg/L gibberellic acid, G50K- kinetin at 10mg/L with 50mg/L gibberellic acid. AS-media enriched with 3% sucrose, G10S-3% sucrose with 10mg/L GA_3, G20S-3% sucrose with 20mg/L GA_3 and G50S 3% sucrose with 50mg/L GA_3. AKS- media enriched with 3% sucrose and 10mg/L kinetin, G10KS- 3% sucrose and 10mg/L kinetin, G20KS-and sucrose with 20mg/L kinetin and G50KS- sucrose with 50mg/L kinetin.

4.2.2. Effect of media regimes on total Chlorophyll content in leaf explants

The total chlorophyll content in detached leaves from H144 and CKIR11027 was quantified and compared across media regimes used in this study. It was hypothesized that senescence could have an effect on the amount of chlorophyll in the leaf after detachment. In H144, analysis of variance on the total chlorophyll content showed that detached discs on media formulation G20KS had the highest chlorophyll content (18.27 ± 0.493 mg/g of tissue) followed by G20 (15.03 ± 0.9682 mg/g of tissue) and G10S (15.02 ± 0.9682 mg/g of tissue). The lowest chlorophyll in this genotype was recorded in leaf discs cultured on G20K with an average of 5.33 ± 0.3433 mg/g of tissue (Figure 4.10). Significantly lower chlorophyll contents were recorded in the other media regimes (Figure 4.10). A similar trend was observed in genotype CKIR11027 with detached leaf discs cultured on G20KS (17.20 ± 0.3688 mg/g of tissue) producing the highest chlorophyll content followed by G20 and G10S at 11.01 ± 0.4222 mg/g of tissue and 11.00 ± 0.4222 mg/g of tissue, respectively. The lowest chlorophyll in this genotype was recorded in leaf discs cultured on AS and G50K with means of 2.90mg/g of tissue and 4.17mg/g of tissue, respectively (Figure 4.10). A Pearson correlation between percent senescence and total chlorophyll content revealed a significantly strong negative correlation (R^2=0.9) with the media formulations that produced the highest percent senescence showing the lowest chlorophyll contents in both genotypes (Figure 4.11).
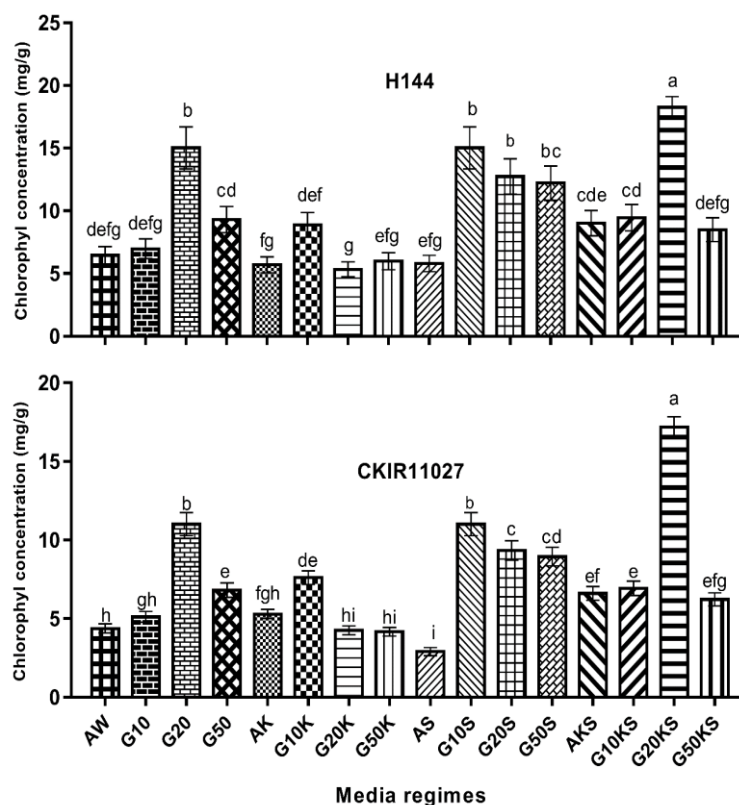
Figure 4. 10: Effect of media regimes on the concentration of chlorophyll in leaves of H144 and CKIR11027 at day19 post-detachment

Data represents means with standard deviations of 3 replicates (n=30). Letters on top of each bar indicate mean separations according to Tukey's HSD test (p≤ 0.05). Means followed by the same letter are not significantly different. AW-media without any enrichment. Media with Gibberellic acid at 10mg/L (G10), 20mg/L (G20) 50 mg/L (G50). Media with kinetin 10mg/L (AK), G10K-kinetin at 10mg/L with 10mg/L gibberellic acid, G20K-kinetin at 10mg/L with 20mg/L gibberellic acid, G50K- kinetin at 10mg/L with 50mg/L gibberellic acid. AS-media enriched with 3% sucrose, G10S-3% sucrose with 10mg/L GA_3, G20S-3% sucrose with 20mg/L GA_3 and G50S 3% sucrose with 50mg/L GA_3. AKS- media enriched with 3% sucrose and 10mg/L kinetin, G10KS- 3% sucrose, 10mg/L GA_3 and 10mg/L kinetin, G20KS-20mg/L GA_3 and 3% sucrose with 20mg/L kinetin and G50KS-50mg/L GA_3 and 3% sucrose with 10mg/L kinetin
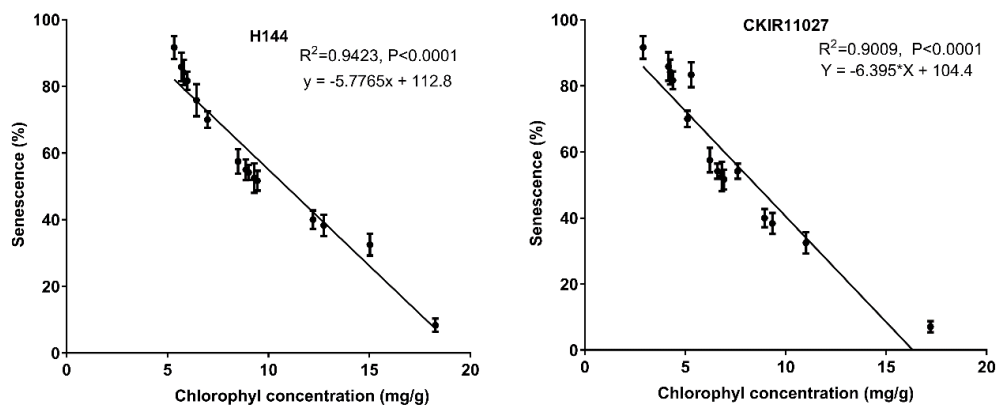
Figure 4.11:Correlation analysis between chlorophyll concentration and percent senescence detached leaf

Correlation analysis between chlorophyll concentration and percent senescence in leaves of H144 and CKIR11027 genotypes cultured on different media regimes at 19 days after detachment. The vertical axis represents percentage senescence while the horizontal axis represents chlorophyll concentration. Correlation was done using the Pearson matrix ($p \leq 0.05$) in GraphPad Prism version 6.

The best performing media regime; G20KS- comprising of 10mg/L of kinetin, 20mg/L of gibberellic acid and 3% sucrose was selected for assay of *in vitro* resistance to SCMV and MCMV co-infection through observation and scoring of disease symptoms. Abrasion and inoculation of the leaf explants with crude extracts of both viruses resulted in development of typical MLND symptoms in all genotypes studied although rate and severity of the symptom development varied across the different germplasm used (Plate 4.4). The initial MLND symptoms were first observed after five days' post-inoculation and progressed with time. The disease was more pronounced in the susceptible genotypes H144 and *Namba nane* (Plate 4.4). The tolerant genotypes retained most of the green color throughout the leaf discs with the abraded regions remaining free of chlorotic lesions (yellow and necrotic spots). At the 19th day post-inoculation, *Namba nane* showed extensive necrosis throughout the leaf with limited areas remaining free of yellow and necrotic spots (Plate 4.4).
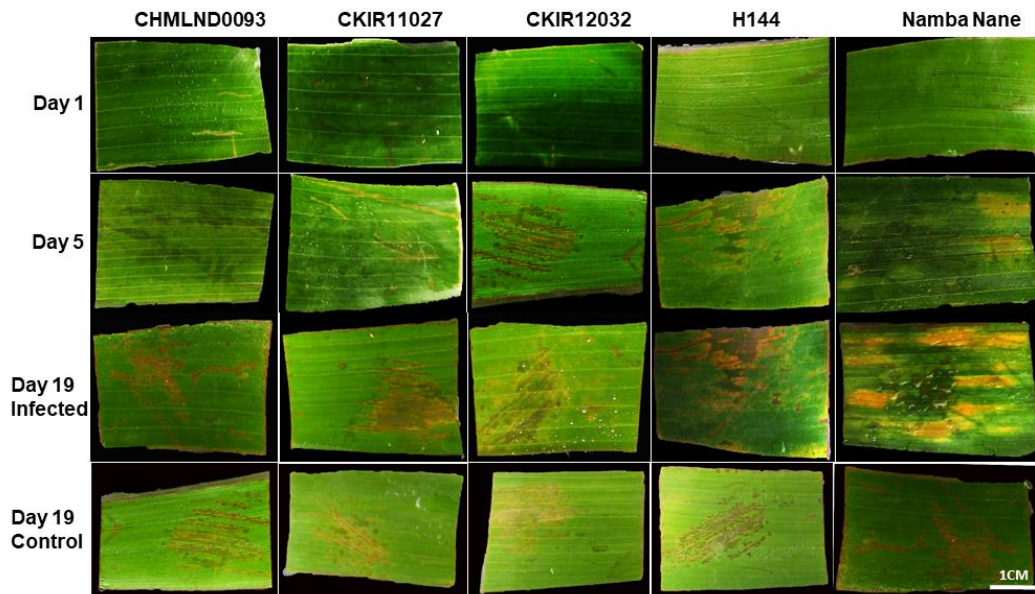
Plate 4. 4: Profile of five maize genotypes assayed for MLND symptoms.

Profile of five maize genotypes assayed for MLND symptoms and disease progression using detached leaves on culture media. Photographs were taken at day 1, 5 and 19 following inoculation with a 20µl of inoculum containing MCMV and SCMV. Each column represents a single genotype while the rows indicate time points. Bar=1cm.

Das *et al*. (2015) scale for scoring MLND severity was used to score for disease severity to confirm susceptibility and tolerance of these genotypes. Analysis of variance of disease severity revealed significant differences (p≤0.0001) among the germplasm. A heatmap generated using these disease indices clustered the genotypes into susceptible and tolerant categories (Figure 4.12). The highest disease severity index was recorded in *Namba nane* and H144 (average 2.997±0.481 and 2.700 ± 0.480, respectively) while CHMLND0093, CKIR12032 and CKIR11027 showed tolerance exhibited by significantly lower mean severities of 2.071 ± 0.482, 2.097 ± 0.481 and 2.173 ± 0.480, respectively (Figure 4.12). Analysis of variance for the 28, 42- and 63-day old plants at p≤0.05 showed complete distinction between resistant and susceptible (Supplementary Figure. 4.13*)*
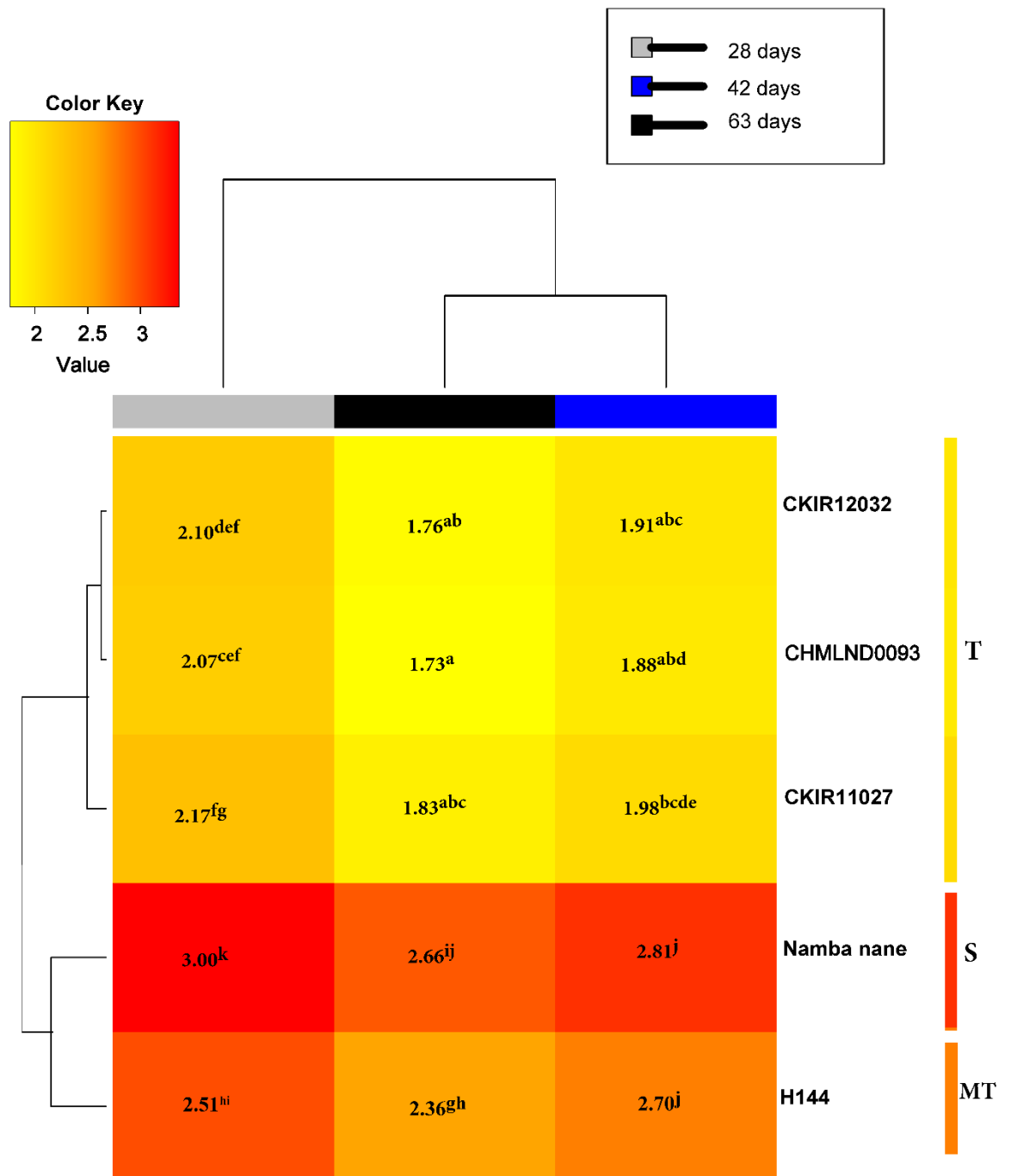
Figure 4. 12: Comparison of resistance among detached leaves of two susceptible and three tolerant varieties of maize at 28,42 and 63-day post-detachment and infection.

Clustering of the five-germplasm based on disease severity scores using leaf explants collected at different ages of 28,42 and 63days after planting. The heatmap was generated using mean severity scores at the 19-day post inoculation with letters indicating significant differences (p≤0.05). T= tolerant, S =Susceptible, MT=Modereate tolerance.
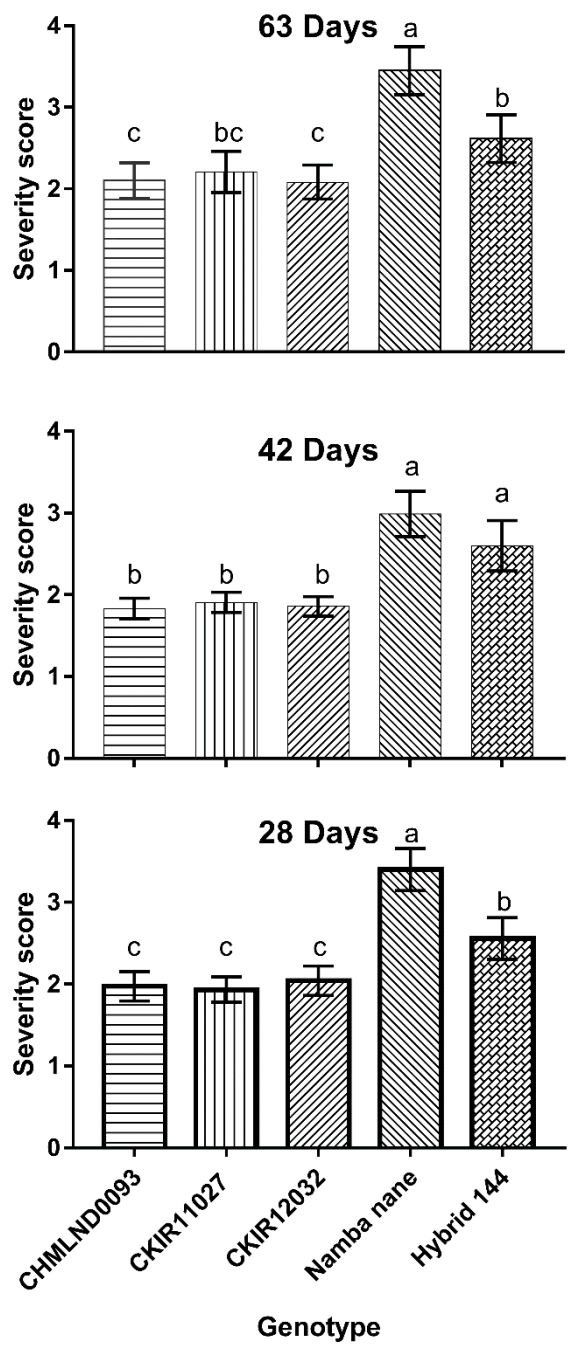
Figure 4. 13: Analysis of resistance against MLND for explants from 28, 42 and 63-day old plants.

4.3. Identification of alternate host for maize chlorotic mottle virus and sugarcane mosaic virus

4.3.1. Maize chlorotic mottle virus detected in grass family members and *Commelina benghalensis*

Most of the grass family members collected in close proximity to infested fields tested positive for MCMV. Sugar cane *(Saccharum officinarum),* Napier grass *(Pennisetum purpureum)* Proso millet *(Panicum miliaceum)*, Sorghum *(Sorghum versicolor)* and Finger millet *(Eleusine coracana)* all tested positive for the presence of MCMV. Except for the Sugar cane *(Saccharum officinarum),* all the samples that tested positive were not showing any symptoms of the disease (Plate 4.5).
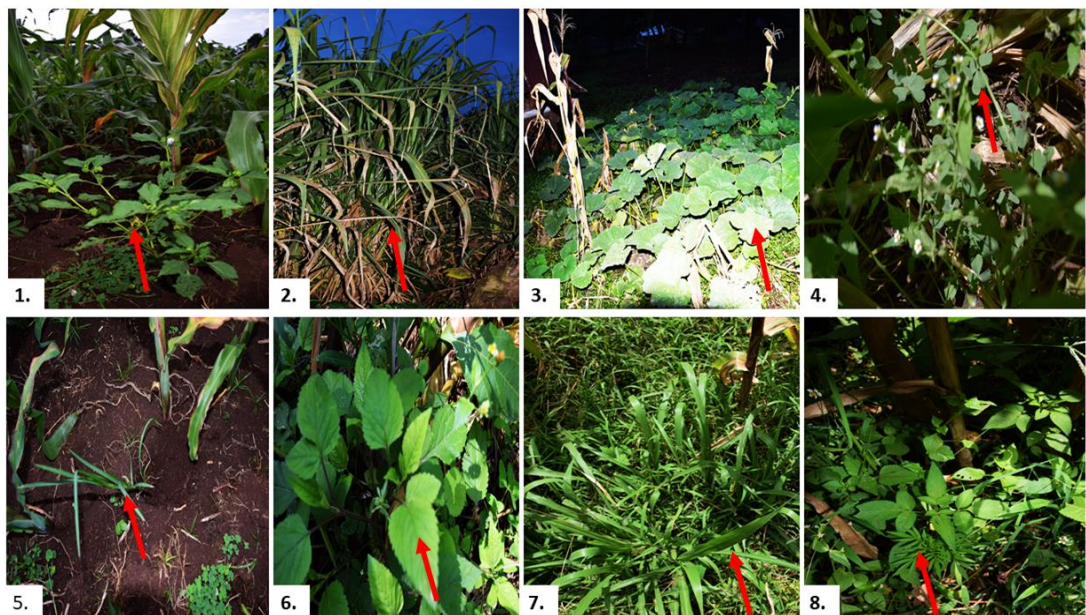
Plate 4. 5:Some of the alternate hosts tested for MCMV and SCMV collected from Nyamira, Kericho and Bomet counties-Kenya, 2018.

1.) *Physalis alkekengi 2.) Saccharum officinarum 3.) Cucurbita moschata 4.) Oxalis species 5.) Eleusine coracana 6.) Bidens pilosa 7.) Pennisetum purpureum 8.) Tagetes minula*

The majority of non-Poaceae members showed negative results for MCMV. These included Common bean (*Phaseolus vulgaris*), oxalis (*Oxalis species*), macdonald's eye (*Gallinsoga parviflora*), Chinese lantern (*Physalis alkekengi*), common pigweed (*Amaranthus spp*) and pumpkin (*Cucurbita moschata*). A surprising observation was the wandering Jew (*Commelina benghalensis*) which was positive for MCMV (Plate 4.6).
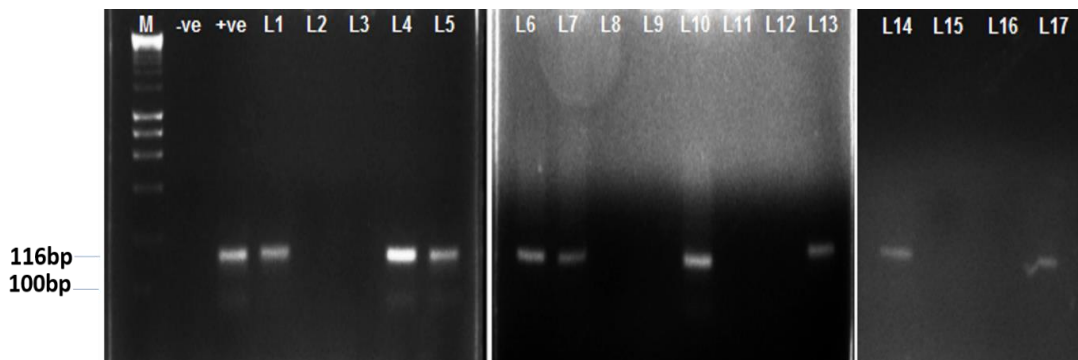
Plate 4. 6:Identified alternate hosts for MCMV represented by 116bp gel bands for non-maize plants collected from Nyamira, Kericho and Bomet counties-Kenya, 2018.

Lanes: negative control (-ve), maize positive control(+ve), *Commelina benghalensis (*L1)*, Phaseolus vulgaris* (L2), *Oxalis species (*L3), *Saccharum officinarum (*L4)*,*

Pennisetum purpureum (L5)*, Panicum miliaceum* (L6)*, Sorghum versicolor* (L7),

Gallinsoga parviflora (L8)*, Physalis alkekengi* (L9), *Eleusine coracana* (L10),

Amaranthus spp. (L11), *Cynodon aethiopicus* (L12), *Pennisetum clandestinum* (L13),

Eleusine africana (L14), *Cucurbita moschata* (L15), *Bidens pilosa* (L16), *Cynodon*

dactylon (L17)

4.3.2. SCMV detected in 5 members of grass family and *Commelina benghalensis*

Only five samples confirmed positive for SCMV. These include Sugar cane *(Saccharum officinarum),* Napier grass *(Pennisetum purpureum)* Prosso millet *(Panicum miliaceum)*, Sorghum *(Sorghum versicolor)* and Finger millet *(Eleusine coracana).* Wandering Jew (*Commelina benghalensis)* tested positive for SCMV (Plate 4.7).
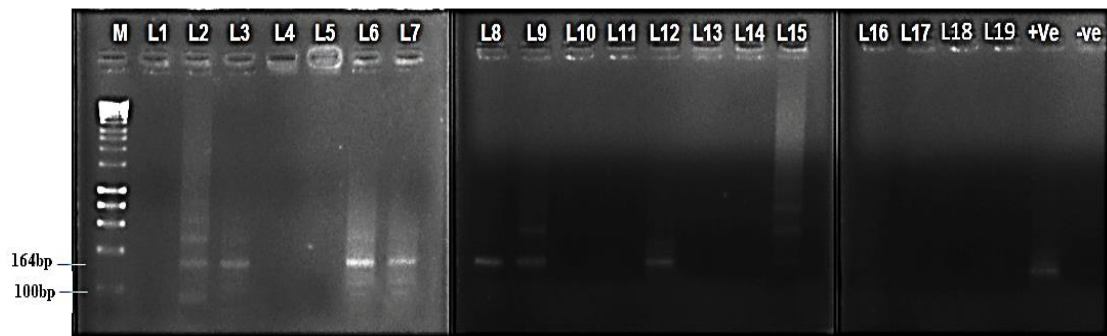
Plate 4. 7: Identified alternate hosts for SCMV represented by 164bp gel bands for non-maize plants collected from Nyamira, Kericho and Bomet counties-Kenya, 2018.

Lanes: negative control (L1), maize positive control(L2), *Commelina benghalensis*(L3)*, Phaseolus vulgaris* (L4), *Oxalis species (*L5), *Saccharum officinarum (*L6)*, Pennisetum purpureum* (L7)*, Panicum miliaceum* (L8)*, Sorghum versicolor* (L9)*, Gallinsoga parviflora* (L10)*, Physalis alkekengi* (L11), *Eleusine coracana* (L12), *Amaranthus spp.* (L13), *Cynodon aethiopicus* (L14), *Pennisetum clandestinum* (L15)*, Eleusine africana* (L16), *Cucurbita moschata* (L17), *Bidens pilosa* (L18), *Cynodon dactylon* (L19).

Plate 4. 8:Sugarcane plants showing typical symptoms of SCMV/MCMV infection on leaves.

A sugarcane (*Sacharum officinarum*) in Bomet county growing in field where maize was previously cultivated. Clear symptoms of necrotic spots on the lower and upper leaves. The samples collected from these plants were positive for both MCMV and SCMV.
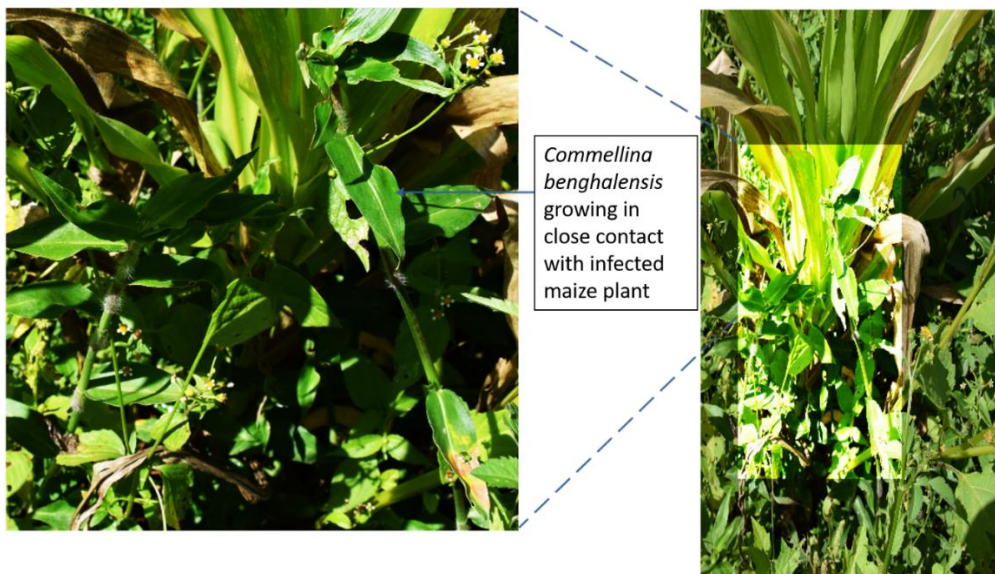
Plate 4. 9: *Commelina benghalensis* growing in contact with maize infected with MLND, Nyamira county-Kenya.

Wandering jew (*Commelina benghalensis*) weed growing in contact with an infected maize plant. The weed is aasociated with persistence in the field and even though the plant tested positive for MCMV and SCMV, it did not present any signs of lethal necrosis disease.

CHAPTER FIVE

DISCUSSION

5.1.   sRNA profiling using next-generation sequencing for detection of MLND causal virus

5.1.1.   Subtractive mapping, miRNAs identification and PsRNAtarget Analysis

In all the three regions sampled, length distribution of small RNAs showed the highest number of molecules to be between 21-22 nucleotides. These are classes of RNAs which have been characterized to include microRNAs (miRNAs), small interfering RNA (siRNAs) and Piwi-interacting RNAs (piRNAs) as well as viral-sourced small RNAs (vsiRNAs) as shown in *Appendix 4* (Iqbal *et al.*, 2017). The main sources of miRNAs are post-transcriptional gene silencing (PTGS) where they are involved in regulation of gene expressions. Piwi-Interacting RNAs have also been associated with gene silencing especially of transposon genes (Saito, 2013). SiRNAs share high similarities with miRNAs but they have very specific targets and are not as broad targeting as miRNAs (Lam *et al.*, 2015).

VsiRNAs are small RNA species which are sourced from the infecting virus in the host organism and despite their biogenesis being synchronized with that of siRNAs and miRNAs, they serve completely different roles. While miRNAs are implicated in PTGS, vsiRNAs are involved in inhibition of RNA-silencing system hence allowing the replication and further spread of the virus within the host (Iqbal *et al.*, 2017) . Low

percentages of miRNAs mapped to the annotated miRGENE and mirBase database of maize. Of the five identified miRNAs sourced from the host, zma-miR159a-3p is the only microRNA that have been previously identified to target MCMV virus. Small RNA target analysis for all the identified host-derived miRNAs by PsRNAtarget analysis demonstrated an excellent target site accessibility based on low unpaired energy (UPE) required to open the secondary structure around the target site. For MCMV, only three miRNAs were identified whereas SCMV had 14 miRNAs with all having UPE values of -1 with two domains of MCMV including replicase and replicase-associated domains being the only identified targets. 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 were however 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 in plant cell.

5.1.2. Machlomaviruses, Potyviruses, Mastrevirus, Poleroviruses as well as un-accounted viruses detected

Similarity search for *de novo* assembled contigs from the host-filtered reads detected a total of five maize viruses including Maize Chlorotic Mottle Virus (MCMV), Sugarcane Mosaic Virus (SCMV), Maize –associated totiviruses (MatV), Maize Streak Virus (MSV) as well as Maize Yellow Mosaic Virus (MYMV). Of all the detected viruses, only MCMV genome was fully assembled while SCMV genome was

assembled up to 90% across the three regions. This explains the typical MLND symptoms; the resultant disease due to co-infection of the detected viruses that were evident during collection of the samples. SCMV, the most characterized potyvirus in MLND development encodes protein HC-pro which has been previously identified as a viral suppressor (VSR) of maize immune system (Kasschau *et al*., 2003; Plisson *et al*., 2003; Soitamo *et al*., 2011; Stenger *et al*., 2007). The upregulation of vsiRNAs associated with MCMV which could have enabled complete assembly of the viral genome is an observation that has been made by studies on synergistic interactions between the two viruses where the potyvirus agent promotes the expression of the co-infecting virus (Xia *et al*. 2016). Maize streak virus (MSV), a mastrevirus was also detected and fully assembled for samples from Kericho and partially assembled for samples from Nyamira and Bomet. The assembly of MSV was complete for samples from Nyamira county and up to 50% assembly for samples from Kericho and Bomet counties. This raises a question as to whether this virus is involved in the development of the MLND. Indeed, a metagenomics project for MLND causal viruses done in Kenya equally detected the presence of this virus in MLND positive samples (Wamaitha *et al*., 2018). The development of lethal necrosis under a co-infection process has solely been associated with VSRs proteins produced by SCMV and to date, no viral suppressors have been identified with MSV even though studies by Xu *et al*. (2017) on Wheat Dwarf Virus belonging to the same family as MSV encodes a replicase (Rep) protein that acts as a VSR of PTGS. This points out to a possibility that MSV may be harbouring a VSR that allows its contribution to MLND development and synergistic relations thereof. MYMV was also detected and despite their assembly being relatively poor, their presence could be contributing to the disease since the said

viruses are able to code for a PO protein that inhibits both local and systemic RNA silencing (Chen *et al*., 2016).

5.1.3.   The diversity of SCMV and MCMV

Pylogenetic analysis of strains identified from the assembled contigs showed divergence among the regional isolates of Sugarcane Mosaic Virus. SCMV isolated from Kericho displayed variations from the rest of isolates with Konoin and Nyamira (Borabu) showing close evolutionary relation to Chinese and Ecuador isolates. Previously isolated SCMV from Kenya, specifically Bomet region was surprisingly distantly related to the current isolates which demonstrated the highly evolving and mutating nature of SCMV. These observations are in line with the findings by single nucleotide polymorphism analysis studies carried out by Wamaitha *et al*. (2018) which identified genetically distinct groups of SCMV in Kenya. In contrast, all the MCMV isolates clustered in one clade showing slight evolutionary divergence. All the previously isolated MCMV strains from Bomet and KALRO clustered with the current MCMV isolates demonstrating a relatively high genetic stability of MCMV. Previous reports have equally indicated the low levels of evolution in MCMV strains with various studies demonstrating very low divergence (Braidwood *et al*., 2018; Kumar and Road, 2017).

5.1.4.   Conserved and highly expressed domains of SCMV and MCMV

Capsid protein CDS region and the two p7 domains namely p7a and p7b were the most frequently mapped domains of MCMV genome. These domains are all encoded by sub-genomic RNA (sgRNA) with capsid protein being expressed at the 3'-proximal end of the sgRNA (Lommel *et al*., 1991). Capsid proteins of various viruses have been

associated with several functionalities including the role in translation of viral RNA, suppression of RNA silencing, virulence during viral infection as well as in dictation of specificity of vectors for virus transmission. A well characterized role of MCMV's capsid protein is its terminal encoded amino acids which allows subcellular localizations of MCMV (Zhan *et al*., 2016). P3 protein peptide, 6K2 and Nla protein regions from SCMV were identified as the frequently mapped domains. The SCMV genome encodes a single, large polypeptide encoding ten proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb, CP) which are cleaved by self-encoded proteases (Akbar *et al*., 2017; Mbega *et al*., 2016). P3 is a movement protein playing a role in enlarging the pore size of plasmodesmata and translocation of the viral genetic material into the neighboring cell thus promoting the local and systemic distribution of the virus within the host (Cabanas *et al*., 2013). The protein domain 6K2 has been associated with replication while Nla Pro, a small nuclear inclusion protein is the main protease involved cleavage of the polypeptide to individual proteins (Land *et al*., 1994).

The expression levels were confirmed by qPCR analysis which were consistent with the small RNA mapping statistics. Further, specificity of the selected markers/primers were confirmed based on the single band gels and melting curve (*appendix 1.0).* Primers targeting P3 domain of SCMV, capsid protein and P7b for MCMV were selected for further modifications for use in development of markers.

5.2. Detached Leaf Assay as a Pre- Screening strategy for tolerance to Maize Lethal Necrosis Disease in maize

Maize lethal necrosis disease is currently the most devastating disease of maize throughout sub Saharan Africa (Makone *et al*., 2014; Fatma *et al*., 2016; Jardine, 2017). The causal viruses are spread by insect vectors as well as infected seed materials and plant debris (Mahuku *et al*., 2015). Despite there being no complete immunity to MLND in maize, screening many germplasms for tolerance against these viruses offers great promise as a strategy to manage the disease. Institutions including CYMMIT and KALRO (Das *et al*., 2015) are involved in a number of field screening assays that rely on whole plant inoculation (WPI). WPI screening process has several challenges which include; the risk of spreading causative viruses to previously unaffected regions, it is labour intensive and requires large parcels of land to establish trials; making this strategy a difficult undertaking. Additionally, variations in inoculation pressure coupled with the effect of environmental factors during WPI often leads to false results in tolerance identification. To circumvent these drawbacks and identify tolerance and/or susceptibility in maize at an early age, the study optimized a detached leaf assay for rapid screening of large germplasm as a pre-screening technique for tolerance against MLND.

First, an *in vitro* media regime that can maintain the green colour of detached leaf explants for a long period of time to allow analysis for tolerance by monitoring leaf senescence was optimized (Kone *et al*., 2017). Senescence, a process that is naturally age-dependent, is the gradual loss of chlorophyll molecules coupled with protein degradation in the leaves and has been shown to be induced by various stresses

including detachment. The findings of this study showed that a medium G20KS enriched with 20mg/L gibberellic acid, 10mg/L kinetin and 3% sucrose was the best in sustaining the green color of detached leaf explants for more than 21 days. Correlation between senescence and chlorophyll concentrations in the detached leaves revealed a strong inverse relationship between these two parameters. This could be due to action of GA_3 which is well documented for its ability to reduce the build-up of reactive oxygen species (ROS) associated with increase in the rate of senescence. Rosenvasser *et al.* (2006) demonstrated the potential of GA_3 to reduce the rate of chlorophyll breakdown, suppress the levels of ROS as well as interfere with senescence development. It has also been reported that combining GA_3 with abscisic acid and/or cytokinins improves senescence inhibition (Chen *et al.*, 2001).

G20KS media regime was adopted for screening for tolerance in five maize genotypes that have previously shown tolerance and susceptibility to MLND. Genotypes CHMLND0093, CKIR11027 and CKIR12032 were developed by CYMMIT and have shown field-based MLND tolerance Das *et al.* (2015). Whole plant inoculation with causative viruses revelaed no significant diffferences among the genotypes with each scoring an average of 2.0 characterised by fine chlorotic streaks on lower leaves (Das *et al.*, 2015). The other 2 genotypes (*Namba nane* and H144) used as susceptible checks in this study have not been studied under field trials but have shown to be the most suscetible varieties under natural infection in Western Kenya. In the current study, the tolerant genotypes showed tolerance consistent with that recorded under WPI while the susceptible checks also showed susceptibility. Screening the genotypes using leaf explants from the 28, 42 and 63-day plants showed that tolerance to MLND increased with age as explants sourced from four-week old plants recorded the highest

levels of susceptibility followed by those from the sixth and ninth week. Previous studies have reported this phenomenon in resistance assays with (Develey-Rivière and Galiana, 2007; Mo, 2007) demonstrating an increase in tolerance to tobacco mosaic virus by plants with age. Developmental induction of defence mechanisms have been proposed as an explanation for the positive correlation between resistance and plant developmental stage (Develey-Rivière and Galiana, 2007). Infected leaf explants showed a rapid onset and progression of MLND symptoms in the susceptible genotypes while the tolerant ones showed slow development of the symptoms and less severity. This could be due to the nature of systemic infection of MCMV and SCMV as they synthesise proteins that ensure effective viral spread from cell to cell. MCMV uses the p7a and p7b movement proteins that are involved in cell to cell movement whereas SCMV uses P3-overlap of trans-frame Potyvirus protein (P3N-PIPO) which has been found to colocalize to plasmodesmata, where it serves in conjunction with cylindrical inclusion protein (CI) to mediate cell-to-cell spread of the virus (Scheets, 2016;Addy *et al*., 2017).

Co-infection by these viruses interferes with starch grains in the chloroplasts and disrupt mitochondria leading to reduced respiration and photosynthesis which eventually leads to leaf necrosis (Wang *et al*., 2017). High necrosis and severity scores in the susceptible genotypes could also be due to increased sensitivity of detached leaves, an observation earlier reported by Dong *et al*. (2015) who illustrated a strong correlation between WPI and detached leaf assay during analysis of tolerance in *Rosa spp.* against black spot disease. This work has demonstrated the potential use of the detached leaf assay technique as a pre-screening tool for identification of tolerant

germplasm against MLND and allow selection of germplasm that show tolerance/resistance before whole plant inoculations in the field.

5.3. Identification of alternate host for maize chlorotic mottle virus and sugarcane mosaic virus using small RNA markers

Sugarcane mosaic virus and maize chlorotic mottle viruses are the most characterized and identified causes of MLND in Kenya and East Africa as a whole. Results from this study confirmed the presence of these viruses in hosts other than maize. This explains the difficulty in managing MLND through agronomic practices such as crop rotation and weed management. Most Poaceae members including *Saccharum officinarum, Pennisetum purpureum, Panicum miliaceum, Sorghum versicolor* and *Eleusine coracana* showed positive results for MCMV and SCMV. These results are in line with several reports on the ability of the two viruses to infect members of grass family including millet, sorghum (Toler *et al.,* 1985) and other grasses (Scheets *et al*. ,2004). Wandering Jew (*Commelina benghalensis) is* a *monocotyledonae* belonging to order *Commelinales* in family *Commelinaceae* and is a weed associated with persistence in cultivated fields and the difficulty associated with its management. The positive results observed for both SCMV and MCMV in this plant which does not belong to grass family was unexpected and this raises questions on the previously reported host-range of these viruses being poacea members. Tracing the relatedness of the plant to maize showed divergence at the level of Order where maize belongs to cyperales while *Commelina benghalensis* belong to *commenilase*. To further rule out the possibility of the markers' unspecificity, the forward and reverse primer was

BLASTed to *Commelina benghalensis* genome (NCBI taxid:37143) and this returned a highest similarity of 1.2% to *matK* gene with an E-value of 9.8 thus ruling out the possibility that the primer was amplifying the host's genome (*Appendix 4.8 and 4.9*). Negative results were observed in all the tested legumes and this further affirms the findings that MCMV and SCMV are mainly co-hosted by Poaceae members. However, the existence of MCMV and SCMV in both grass and non-grass weeds that grow in proximity with or in maize fields calls for new approaches to manage MLND.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

6.1.1. sRNASeq transcriptome profiling for detection of MLND causal viruses

Small RNA sequencing and profiling was able to distinctly detect MLND causal viruses as well as other viruses in maize samples presenting MLND symptoms. The relatively high abundance of vsiRNA species compared to host-derived miRNAs provided a strong platform for detection of MCMV and SCMV. Frequently mapped domains which were considered as the highly expressed and conserved regions in the genomes of SCMV and MCMV allowed identification and development of markers whose potential to detect the causative viruses were confirmed.

6.1.2. Detached leaf assay as a pre-screening technique for identification of MLND tolerant germplasm

Detached leaf assay is an easy, rapid and cheaper technique that has a unique potential as a pre-screening strategy for tolerance against maize lethal necrosis disease. Adoption of the technique will allow selection of germplasm that show tolerance/resistance before whole plant inoculations in the field. A medium formulation comprising 20mg/L gibberellic acid, 10mg/L Kinetin and 3% sucrose was the best in inhibiting senescence and maintaining the green color of the leaves long enough to allow for analysis of the disease severity. As a general observation, any

media supplemented with gibberellic acid in combination with 3% sucrose relative to other media retained the green color of the detached leaf discs for extended time.

6.1.3. Identification of alternate host for MCMV and SCMV

Grass family members including Sugarcane (*Saccharum officinarum*), Napier grass (*Pennisetum purpureum*) Prosso millet (*Panicum miliaceum*), Sorghum (*Sorghum versicolor*) and Finger millet (*Eleusine coracana*) were confirmed as the alternate hosts for the MLND causal viruses. MCMV was detected in more grasses than SCMV. Except for Sugarcane, all the alternate hosts did not show typical symptoms of MLND. *Commelina benghalensis* returned positive results for both MCMV and SCMV.

6.2. Recommendations

i. Five distinct viruses including MCMV, MSV, SCMV, MYMV and MATV were identified in this study. The genome assemblies were however not complete for the MSV, MYMV and MATV. The complete assembly of MSV genome from Kericho isolate and close to complete assembly for Nyamira and Bomet counties as well as previous reports of the same virus raises a question on its contribution in the development of the disease. Further investigations on this virus especially its ability to interact with MCMV and /or with SCMV in development of MLND will be important in understanding MLND dynamics.

ii. Clear evolutionary differences were evident for sugarcane mosaic virus with Kenyan isolates showing clear differences. Detection strategies should take into consideration this especially for tools targeting antibodies.

iii. Detached leaf assay technique demonstrated its efficiency in detecting MLND. The technique was tested on five germplasm only and therefore; more germplasm should be tested using the same technique to further validate the reported results.

iv. Resistance against coinfection by MCMV and SCMV improved with age and false tolerance can be identified on using explants from mature plants. It is my recommendation therefore to use explants from young plants, preferably 28 days after planting.

v. Majority of the screened Poaceae members are alternate hosts for the virus. Cultural management strategies such as crop rotation should also involve removal of members of grass family in the vicinity of maize fields.

vi. More research into the ability of Wandering Jew (*Commelina benghalensis*) to host MCMV and SCMV should be carried out. The plant belongs to monocotyledon family together with maize but is not a grass. It is associated with persistence in the field and therefore, if further studies confirm it as an alternate host, management strategies should also focus on this weed.

REFERENCES

Adams, I., Braidwood, L., Stomeo, F., Phiri, N., Uwumukiza, B., Feyissa, B., … Boonham, N. (2017). Characterising maize viruses associated with maize lethal necrosis symptoms in sub Saharan Africa. *BioRxiv*, 161489.

Adams, I.P., Harju, V.A., Hodges, T., Hany, U., Skelton, A., Rai, S., Daka, M.K., Smith, J., Fox, A., Uzayisenga, B., Ngaboyisonga, C., Uwumukiza, B.,Rutikanga, Rutikanga, A., Mutherford, M., Ricthis, B., Phiri, N., Boonham, N. 2014. First report of maize lethal necrosis disease in Rwanda. New Disease

Adams, I. P., Miano, D. W., Kinyua, Z. M., Wangai, A., Kimani, E., Phiri, N., et al. 2013. Use of next-generation sequencing for the identification and characterization of maize chlorotic mottle virus and sugarcane mosaic virus causing maize lethal necrosis in kenya. Plant Pathol. 62:741–749.

Addy, H., Nurmalasari, Wahyudi, A., Sholeh, A., Anugrah, C., Iriyanto, F., … Sugiharto, B. (2017). Detection and Response of Sugarcane against the Infection of Sugarcane Mosaic Virus (SCMV) in Indonesia. *Agronomy*, *7*(3), 50.

Akbar, S., Tahir, M., Wang, M.-B., & Liu, Q. (2017). Expression Analysis of Hairpin RNA Carrying *Sugarcane mosaic virus* (SCMV) Derived Sequences and Transgenic Resistance Development in a Model Rice Plant. *BioMed Research International*, *2017*, 1–10.

Andersson, J., & Halvarsson, M. (2011). The economic consequences of Striga hermonthica in maize production in Western Kenya, (669), 35. Retrieved from http://stud.epsilon.slu.se/3078/%5Cnhttp://stud.epsilon.slu.se/3078/4/andersson_j_etal_110725.pdf

Audy, P., Palukaitis, P., Slack, A., & Zaitlin, M. (1994). Replicase-mediated resistance to Potato virus Y in transgenic tobacco. *Molecular Plant-Microbe Interactions*.

Basu, S., Sharma, V. K., Bhattacharyya, D., & Chakraborty, S. (2014). An Overview of Antiviral RNA Silencing in Plant: Biogenesis, Host–Virus Interaction and Potential Applications BT - Approaches to Plant Stress and their Management. In R. K. Gaur & P. Sharma (Eds.) (pp. 317–337). New Delhi: Springer India.

Beckingham, C., & NSW, ,. (2007). Commodity growing guides-Sweet corn. *Of Primary Industries Available Online at of Primary Industries Website*.

Bockelman, D. L., Claflin L. E., & Uyemoto, J. K. (1982). Host Range and Seed-Transmission Studies of Maize Chlorotic Mottle Virus in Grasses and Corn. *Plant Disease*.

Braidwood, L., Quito-Avila, D. F., Cabanas, D., Bressan, A., Wangai, A., & Baulcombe, D. C. (2018). Maize chlorotic mottle virus exhibits low divergence between differentiated regional sub-populations. *Scientific Reports*, *8*(1), 1–9.

Brandes EW (1919) The mosaic disease of sugarcane and other grasses. Bull US Dept Agric 829:1–26

Brunt, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J., Watson, L. and Zurcher, E. J. (eds.). 1996 onward. Plant Viruses Online: Descriptions and Lists from the VIDE Database: Version: 20th August 1996. http://image.fs.uidaho.edu/vide/refs.htm

Cabanas, D., Watanabe, S., Higashi, C. H. V, Bressan, A., Bockelman, D. L., Uyemoto, J. K., … Zhou, X. P. (2013). Dissecting the mode of maize chlorotic mottle virus transmission (Tombusviridae: Machlomovirus) by Frankliniella williamsi (Thysanoptera: Thripidae). *Journal of Economic Entomology*, *106*(1), 16–24.

Canada, V. T., Osmond, G., Type, A., Collection, C., Drive, P., and Forstwirtschaft, B. B. L. 2018. REVIEW Guidelines for the Identification and Characterization of Plant Viruses. :223–241.

Chen, S., Jiang, G., Wu, J., Liu, Y., Qian, Y., & Zhou, X. (2016). Characterization of a novel polerovirus infecting maize in China. *Viruses*, *8*(5).

Chen, W. S., Liao, L. J., Chen, C. Y., & Huang, K. L. (2001). Kinetin, gibberellic acid and sucrose affect vase life in Oncidium spp. *Acta Botanica Gallica*, *148*(3), 177–181.

Chung. X., Sun, X., Taylor, A., Jiao, C., Xu, Y., Cai, X., … Wang, Q. (2017). Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing. Journal of Virology, 91(11), e00173-17. https://doi.org/10.1128/JVI.00173- 17

Castillo J., Hebert T. (1974). A new virus disease of maize in Peru. Fitopatologia 9, 79–84.

CYMMIT. (2018). MLN Diseases portal » MLN Tolerant/Resistant germplasm. Retrieved September 17, 2018, from http://mln.cimmyt.org/mln-resistanttolerant-germplasm/

Dallwitz, L. W. & M. J. (1992). The Grass Genera of the World 1038 pp. Wallingford: CAB International. *The Journal of Agricultural Science*, (120(3)), 421–421.

Das, B., Makumbi, D., Beyene, Y., & Regasa, M. (2015). Breeding for MLN Tolerance, (March).

Degani, O., Weinberg, T., & Graph, S. (2014). Chemical control of maize late wilt in the field. *Phytoparasitica*, *42*(4), 559–570.

Develey-Rivière, M. P., & Galiana, E. (2007). Resistance to pathogens and host developmental stage: A multifaceted relationship within the plant kingdom. *New Phytologist*, *175*(3), 405–416.

Donatelli M, Magarey RD, Bregaglio S, Willocquet L, Whish JPM, Savary S. Modelling the impacts of pests and diseases on agricultural systems. Agric Syst.

2017;155:213-224.

Family. (2012), 1111–1138. https://doi.org/10.1016/B978-0-12-384684-6.00096-3

FAO. (2017). FAO GIEWS Country Brief on Kenya -. Retrieved September 15, 2018, from http://www.fao.org/giews/countrybrief/country.jsp?code=KEN

FAO. (2018a). GIEWS - Global Information and Early Warning System. Retrieved May 8, 2018, from http://www.fao.org/giews/countrybrief/country.jsp?code=KEN

FAO. (2018b). Enhancing the fight against Fall Armyworm in Kenya | FAO in Kenya | Food and Agriculture Organization of the United Nations. Retrieved September 16, 2018, from http://www.fao.org/kenya/news/detail-events/en/c/1068542/

Farrell, A. P., & Stecyk, J. A. W. (2007). The heart as a working model to explore themes and strategies for anoxic survival in ectothermic vertebrates. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *147*(2), 300–312.

Fatma, H. K., Tileye, F., & Patrick, A. N. (2016). Insights of maize lethal necrotic disease: A major constraint to maize production in East Africa. *African Journal of Microbiology Research*, *10*(9), 271–279.

Frenken, A. (2013). Maize Lethal Necrosis Disease (MLND) A snapshot (FSNWG), 2.

Gingery, R. E., Bradfute, O. E., Gordon, D. T., & Nault, L. R. (1983). Maize chlorotic dwarf virus. Retrieved from http://www.dpvweb.net/dpv/showdpv.php?dpvno=194

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., … Prasanna, B. M. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theoretical and Applied Genetics*, *128*(10), 1957–1968.

He, Zhengying and Erik J Sontheimer. ""siRNAs and miRNAs": a meeting report on RNA silencing" RN(New York, N.Y.) vol. 10,8 (2004): 1165-73.

ICTV. (2018). "Virus Taxonomy: 2018 Release." Retrieved November 01, 2018, from https://talk.ictvonline.org/taxonomy/

Iqbal, M. S., Jabbar, B., Sharif, M. N., Ali, Q., Husnain, T., & Nasir, I. A. (2017). In silico MCMV Silencing Concludes Potential Host-Derived miRNAs in Maize. *Frontiers in Plant Science*, *8*(March), 1–9.

Jardine, D. J. (2017). Factsheets: Corn Lethal Necrosis. *Available Online at Http://Www.Plantpath.k-State.Edu/Extension/Publications/Cornlethalnecrosis.Pdf*. Retrieved from http://www.plantpath.ksu.edu/pages/extension/factsheets CORN

Jiang, X. Q., Meinke, L. J., Wright, R. J., Wilkinson, D. R., & Campbell, J. E.

(1992). Maize chlorotic mottle virus in Hawaiian-grown maize: vector relations, host range and associated viruses. *Crop Protection*, *11*(3), 248–254.

Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., & Carrington, J. C. (2003). P1 / HC-Pro , a Viral Suppressor of RNA Silencing , Interferes with Arabidopsis Development and miRNA Function, *4*, 205–217.

Katoch, R. 2011a. Analytical Techniques in Biochemistry and Molecular Biology.

Kay.S, 1998. Maize chlorotic mottle machlomovirus and Wheat streak mosaic rymovirus concentrations increase in the synergistic disease corn lethal necrosis. Virology, 242(1): 28-38.

Kiptanui, C. (2013). Determining Factors that Affect Maize Production in Turbo Constituency, Kenya. *Seed Science Research*, *1*(2), 45–52. Retrieved from http://erepository.uonbi.ac.ke/bitstream/handle/11295/56008/Chumo_Determini ng Factors That Affect Maize Production In Turbo Constituency.pdf;sequence=3

Kirimi, L., Sitko, N., Jayne, T. S., Karin, F., Muyanga, M., Sheahan, M., … Bor, G. (2011). Policy and development wps 44 / 2011 a farm gate-to-consumer value chain analysis of kenya ' s maize marketing system a farm gate-to-consumer value chain analysis of kenya ' s maize marketing system.

Koenig, R., & Kühn-institut, J. (1981). Guidelines for the Identification and Characterization of Plant Viruses, (November 2014). https://doi.org/10.1099/0022-1317-54-2-223

Kone, N., Asare-Bediako, E., Silue, S., Kone, D., Koita, O., Menzel, W., & Winter, S. (2017). Influence of planting date on incidence and severity of viral disease on cucurbits under field condition. *Annals of Agricultural Sciences*, *62*(1), 99– 104.

Kulkarni, H.Y., 1973. Notes on East African Plant Virus Diseases: 5. Identification and Economic Importance of Sugarcane mosaic virus in maize in East Africa. East African Agricultural and Forestry Journal, 39(2): 158-164.

Kumar, P.L. (ed.). 2009. Methods for diagnosis of plant virus diseases: a laboratory manual. IITA, Ibadan, Nigeria. 90 pp.

Kumar, L., & Road, O. (2017). http://repository.ruforum.org, *14*(14), 488–493.

Kumela, T., Simiyu, J., Sisay, B., Likhayo, P., Mendesil, E., Gohole, L., & Tefera, T. (2018). Farmers' knowledge, perceptions, and management practices of the new invasive pest, fall armyworm (Spodoptera frugiperda) in Ethiopia and Kenya. *International Journal of Pest Management*, 1–9.

Kusia, E. S., Subramanian, S., Nyasani, J. O., Khamis, F., Villinger, J., Ateka, E. M., & Pappu, H. R. (2015). First Report of Lethal Necrosis Disease Associated With Co-Infection of Finger Millet With Maize chlorotic mottle virus and Sugarcane mosaic virus in Kenya. *Plant Disease*, *99*(6), 899.

Lam, J. K. W., Chow, M. Y. T., Zhang, Y., & Leung, S. W. S. (2015). siRNA versus

miRNA as therapeutics for gene silencing. *Molecular Therapy - Nucleic Acids*, *4*(9), e252.

Land, C., Uwc, B. O., & Us, A. L. O. G. (1994). *Distribution and multiplication of iris severe mosaic potyvirus in bulbous Iris in relation to metabolic activity Implications for ISMV detection*.

Langyintuo, A. S., Mwangi, W., Diallo, A. O., MacRobert, J., Dixon, J., & Bänziger, M. (2010). Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private-public intervention to promote growth. *Food Policy*, *35*(4), 323–331.

Li, F., & Ding, S.-W. (2006). NIH Public Access. *Annual Review of Microbiology*, 503–531. https://doi.org/10.1146/annurev.micro.60.080805.142205.Virus

Li, N., Hauser, R., Holford, T., Zhu, Y., Zhang, Y., Bryan, A., … Zheng, T. (2015). Maize lethal necrosis (MLN), an emerging threat to maize-based food 1 security in sub- 2 Saharan Africa, *508*, 1–43.

Li. X., Zhang, J., Wang, Q., Meng, C., Hong, J., & Zhou, X. ping. (2011). Characterization of Maize Chlorotic Mottle Virus Associated with Maize Lethal Necrosis Disease in China. Journal of Phytopathology, 159(3), 191–193.

Lommel, S. A., Kendall, T. L., Xiong, Z., & Nutter, R. C. (1991). Identification of the Maize chlorotic mottle virus capsid protein cistron and characterization of its subgenomic messenger RNA. *Virology*, *181*(1), 382–385. https://doi.org/10.1016/0042-6822(91)90509-A

Lukanda, M., Owati, A., Ogunsanya, P., Valimunzigha, K., Katsongo, K., Ndemere, H. and Kumar, P.L., 2017. First report of Maize chlorotic mottle virus infecting maize in the Democratic Republic of the Congo. Frontiers in Plant Science, 8.

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., … Redinbaugh, M. G. (2015). Maize Lethal Necrosis (MLN), an Emerging Threat to Maize-Based Food Security in Sub-Saharan Africa. *Phytopathology*, *105*(7), 956–965. https://doi.org/10.1094/PHYTO-12-14-0367-FI

Makone, S. M., Menge, D., & Basweti, E. (2014). Impact of Maize Lethal Necrosis Disease on Maize Yield: a Case of Kisii, Kenya. *International Journal of Agricultural Extension*, *2*(3), 211–218. Retrieved from http://escijournals.net/index.php/IJAE/article/view/891

Makumbi, D., & Wangai, A. (2012). Maize lethal necrosis (MLN) disease in Kenya and Tanzania: Facts and actions. Retrieved from http://www.fao.org/fileadmin/user_upload/drought/docs/1 Maize Lethal Necrosis-(MLN) Facts and Actions.pdf

Mbega, E. R., Ndakidemi, P. A., Mamiro, D. P., & Mushongi, A. A. (2016). Role of Potyviruses in Synergistic Interaction Leading to Maize Lethal Necrotic Disease on Maize. *International Journal of Current Microbiology and Applied Sciences*, *5*(6), 85–96.

McCutcheon, A. (2007). Bioethanol in Victoria. *State of Victoria, Department of Primary Industries*, *G1314*, 1–3.

Mo, J. Y. (2007). Detached Leaf Assay : A Rapid Screening Technique to Study Foliar Diseases of Soybean Study Foliar Diseases of Soybean.

Nelson, S., Brewbaker, J., & Hu, J. (2011). Maize Chlorotic Mottle. *Plant Disease*, *79*(December), 1–6. Retrieved from http://scholarspace.manoa.hawaii.edu/handle/10125/32440

Ngoune Tandzi, L., Mutengwa, C., Ngonkeu, E., & Gracen, V. (2018). Breeding Maize for Tolerance to Acidic Soils: A Review. *Agronomy*, *8*(6), 84.

Niblett, C. L., & Claflin, L. E. (1978). Corn lethal necrosis - a new virus disease of corn in Kansas. *Plant Disease Reporter*, *62*(1), 15–19.

Nyoro, J. K., Kirimi, L., & Jayne, T. S. (2004). Competitiveness of Kenyan and Ugandan Maize Production: Challenges for the Future. Retrieved from http://agris.fao.org/openagris/search.do?recordID=US2016201728

Omarov, R., Sparks, K., Smith, L., Scholthof, H. B., & Zindovic, J. (2006). Biological Relevance of a Stable Biochemical Interaction between the Tombusvirus-Encoded P19 and Short Interfering RNAs Biological Relevance of a Stable Biochemical Interaction between the Tombusvirus-Encoded P19 and Short Interfering RNAs, *80*(6), 3000–3008. https://doi.org/10.1128/JVI.80.6.3000

Pataky, J. K., & Snetselaar, K. M. (2006). Common smut of corn. *The Plant Health Instructor*. https://doi.org/10.1094/PHI-I-2006-0927-01

Patancheru, I., & Bengtson, G. D. (1980). International Crops Research Institute for the Semi-Arid Tropics. *India*, *502*(January), 11–15.

Plisson, C., Drucker, M., Blanc, S., German-Retana, S., Le Gall, O., Thomas, D., & Bron, P. (2003). Structural characterisation of HC-Pro, a plant virus multifunctional protein. *Journal of Biological Chemistry*, *278*(26), 23753–23761.

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).

Quito-Avila, D. F., Alvarez, R. A., & Mendoza, A. A. (2016). Occurrence of maize lethal necrosis in Ecuador: a disease without boundaries? *European Journal of Plant Pathology*, *146*(3), 705–710.

Redinbaugh, M. G., & Zambrano, J. L. (2014). *Control of Virus Diseases in Maize*. *Advances in Virus Research* (Vol. 90).

RL, G, HR, & AD. (2000). Abiotic stresses affecting maize. *In Eds Tropical Maize Improvement and Production Food and Agriculture Organization of the United Nations Rome Pp*, 93–103.

Roche, P., Maillet, F., Plazanet, C., Debellé, F., Ferro, M., Truchet, G., … Dénarié, J. (1996). The common nodABC genes of Rhizobium meliloti are host-range determinants. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(26), 15305–15310.

Rosario, K., and Breitbart, M. 2011. Exploring the viral world through metagenomics. Curr. Opin. Virol. 1:289–297.

Rosenvasser, S., Mayak, S., & Friedman, H. (2006). Increase in reactive oxygen species (ROS) and in senescence-associated gene transcript (SAG) levels during dark-induced senescence of Pelargonium cuttings, and the effect of gibberellic acid. *Plant Science*, *170*(4), 873–879. https://doi.org/10.1016/j.plantsci.2005.12.010

Saito, K. (2013). The epigenetic regulation of transposable elements by PIWI-interacting RNAs in Drosophila. *Genes & Genetic Systems*, *88*(1), 9–17.

Sastry, K. S. (2013). *Introduction to plant virus and viroid disease in the tropics. Plant Virus and Viroid Diseases in the Tropics* (Vol. 1). https://doi.org/10.1007/978-94-007-6524-5

Scheets, K. (1998). Maize Chlorotic Mottle Machlomovirus and Wheat Streak Mosaic Rymovirus Concentrations Increase in the Synergistic Disease Corn Lethal Necrosis. *Virology*, *242*(1), 28–38. https://doi.org/10.1006/viro.1997.8989

Scheets, K. (2016). Analysis of gene functions in Maize chlorotic mottle virus. *Virus Research*, *222*, 71–79. https://doi.org/10.1016/j.virusres.2016.04.024

Scheets, K., Lapierre, H., & Signoret, P. (2004). Maize Chlorotic Mottle. *In A Eds Viruses and Virus Diseases of Poaceae Gramineae Institut National de La Recherche Agronomique Paris*, 642–644.

Shukla, D. D., Ward, C. W., & Brunt, A. A. (1994). The Potyviridae. Retrieved from http://www.cabdirect.org/abstracts/19952304874.html

Sibhatu, K. T., & Qaim, M. (2017). Rural food security, subsistence agriculture, and seasonality. *PLoS ONE*, *12*(10), 1–15. https://doi.org/10.1371/journal.pone.0186406

Soitamo, A. J., Jada, B., & Lehto, K. (2011). HC-Pro silencing suppressor significantly alters the gene expression profile in tobacco leaves and flowers.

Stenger, D. C. D., Young, B. A. B., Qu, F., Morris, T. T. J., & French, R. (2007). Wheat streak mosaic virus lacking helper component-proteinase is competent to produce disease synergism in double infections with Maize chlorotic mottle virus. *Phytopathology*, *97*(10), 1213–1221.

Stobbs, L. W., Lowery, D. T., Samara, R., Greig, N., Vickers, P. M., & Bittner, L. A. (2015). Development of a detached leaf procedure to evaluate susceptibility to Plum pox virus infection by the green peach aphid (Myzus persicae (Sulzer)) in peach. *Canadian Journal of Plant Pathology*, *37*(2), 230–236.

Stocks, M. B.; Moxon, S.; Mapleson, D.; Woolfenden, H. C.; Mohorianu, I.; Folkes, L.; Schwach, F.; Dalmay, T. & Moulton, V. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets.Bioinformatics , Vol. 28, pp. 2059-2061.PubMed.

Syller, J. 2012. Facilitative and antagonistic interactions between plant viruses in mixed infections. Mol. Plant Pathol. 13:204–216.

syngenta. (2018). Fall Army Worm Invasion in Kenya| Syngenta. Retrieved September 30, 2018, from https://www.syngenta.co.ke/fall-army-worm-invasion-kenya

Tamra Jackson-Ziems. (2016). Ear and Stalk Rot Diseases Becoming More Common in Corn Fields | CropWatch | University of Nebraska–Lincoln. Retrieved September 16, 2018, from https://cropwatch.unl.edu/2016/ear-and-stalk-rot-diseases-becoming-more-common-corn-fields

Thornbury, D. W., & Pirone, T. P. (1983). Helper components of two potyviruses are serologically distinct. *Virology*, *125*(2), 487–490.

Toler, R. W. (1985). Maize dwarf mosaic, the most important virus disease of sorghum. *Plant Disease*, *69*(11), 1011–1015.

Twizeyimana, M., Ojiambo, P., & Ikotun, T. (2007). Comparison of field, greenhouse, and detached-leaf evaluations of soybean germplasm for resistance to Phakopsora pachyrhizi. *Plant ...*, (September), 1161–1169.

Uyemoto, J.K., 1980. Detection of Maize chlorotic mottle virus serotypes by enzyme-linked immunosorbent assay. Phytopathology, 70(4): 290-292.

viralzone. 2018. Weblet Importer. Available at: https://viralzone.expasy.org/634 [Accessed September 16, 2018].

Voinnet, O., Pinto, Y. M., & Baulcombe, D. C. (1999). Suppression of gene silencing : A general strategy used by diverse DNA and RNA viruses of plants, *96*(24).

Wamaitha, M. J., Nigam, D., Maina, S., Stomeo, F., Wangai, A., Njuguna, J. N., … Garcia-Ruiz, H. (2018). Metagenomic analysis of viruses associated with maize lethal necrosis in Kenya. *Virology Journal*, *15*(1), 1–19.

Wang, Q., Zhang, C., Wang, C., Qian, Y., Li, Z., Hong, J., & Zhou, X. (2017). Further characterization of Maize chlorotic mottle virus and its synergistic interaction with Sugarcane mosaic virus in maize. *Scientific Reports 7*.

Wangai, a W., Kinyua, Z. M., Otipa, M., Miano, W. D., Kasina, J. M., Leley, P. K., & Mwangi, T. N. (2012). Maize lethal necrosis disease. *KARI Information Brochure*, (December), 1–2.

Wangai, A. W., Redinbaugh, M. G., Kinyua, Z. M., Miano, D. W., Leley, P. K., Kasina, M., … Jeffers, D. (2012). First Report of *Maize chlorotic mottle virus* and Maize Lethal Necrosis in Kenya. *Plant Disease*, *96*(10), 1582–1582.

Weblet Importer. Available at: http://viralzone.expasy.org/637?outline=all_by_species [Accessed October 16, 2018].

Xia, Z., Zhao, Z., Chen, L., Li, M., Zhou, T., Deng, C., … Fan, Z. (2016). Synergistic infection of two viruses MCMV and SCMV increases the accumulations of both MCMV and MCMV-derived siRNAs in maize. *Scientific Reports*, *6*(November 2015), 20520.

Xu, C., Sun, X., Taylor, A., Jiao, C., Xu, Y., Cai, X., … Wang, Q. (2017). Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing. *Journal of Virology*, *91*(11), e00173-17. https://doi.org/10.1128/JVI.00173-17

Xu, P., Zhao, C., & Zhang, Z. (2017). Identification of an RNA silencing suppressor encoded by Lily symptomless virus. *International Journal of Agriculture and Biology*, *19*(5), 1259–1264. https://doi.org/10.17957/IJAB/15.0434

Zhan, B., Lang, F., Zhou, T., & Fan, Z. (2016). Nuclear import of Maize chlorotic mottle virus capsid protein is mediated by importin-α. *European Journal of Plant Pathology*, *146*(4), 881–892. https://doi.org/10.1007/s10658-016-0966-y

Zhang, C. (2009). Novel functions for small RNA molecules. *Current Opinion in Molecular Therapeutics*, *11*(6), 641–651. https://doi.org/10.1007/978-3-540-93824-8_4523

Zhang, C., Wu, Z., Li, Y., & Wu, J. (2015). Biogenesis, function, and applications of virus-derived small RNAs in plants. *Frontiers in Microbiology*, *6*(NOV), 1–12. https://doi.org/10.3389/fmicb.2015.01237

Zhao, B., Lin, X., Poland, J., Trick, H., Leach, J., & Hulbert, S. (2005). A maize resistance gene functions against bacterial streak disease in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15383–15388. https://doi.org/10.1073/pnas.0503023102

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., … Fei, Z. (2017). VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology*, *500*, 130–138. https://doi.org/10.1016/j.virol.2016.10.017

ZymoResearch. (2018). DNA/RNA Shield[TM] Reagent, Zymo Research | VWR. Retrieved September 17, 2018, from https://us.vwr.com/store/product/21876469/dna-rna-shield-reagent-zymo-research

LIST OF APPENDICES
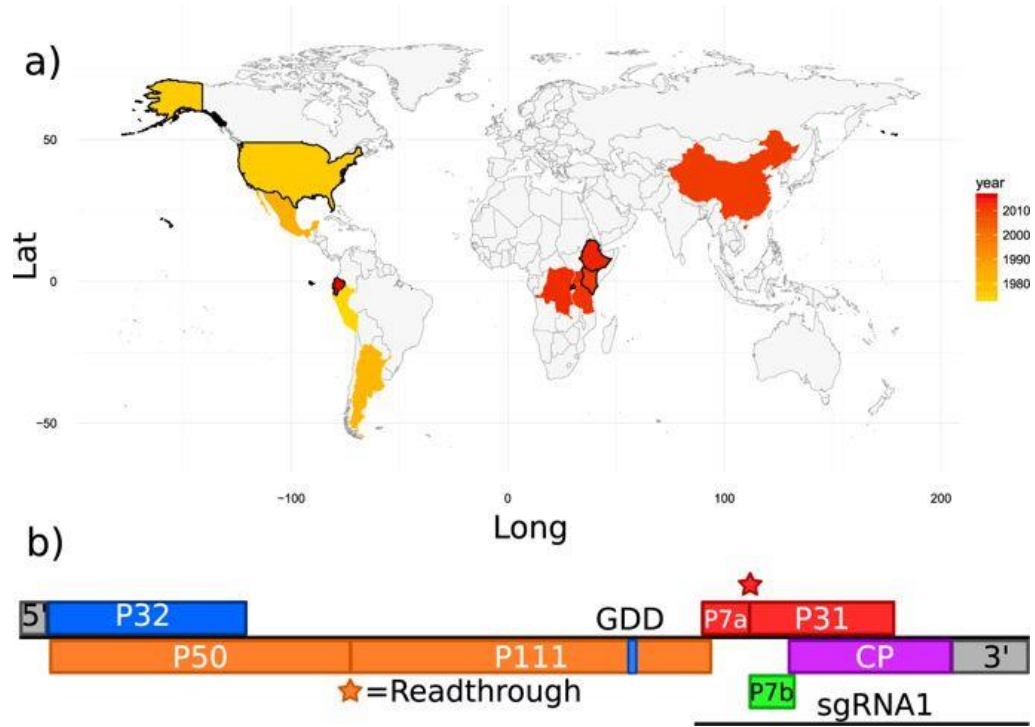
Appendix 2. 1: Global distribution of maize chlorotic mottle virus (MCMV), coloured by year of first report. Countries sampled in this study are outlined in black. (b) MCMV genome structure.

(Available from: https://www.nature.com/articles/s41598-018-19607-4 [accessed 4 Nov, 2018])
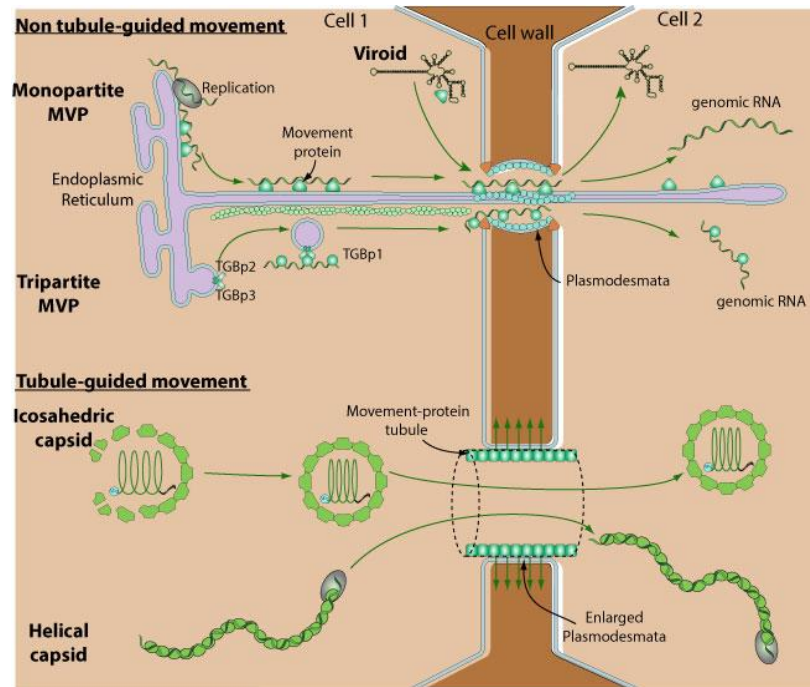
Appendix 2. 2:The role of P3 protein in cell to cell distribution of viruses

P3 acts as an accessory protein, allows cell to cell movement of the SCMV in a host upon infection.

*	1	2	3	4	5	6	7	8	9	10	11	12
A	U Sample 1 BMT MCMV COAT	U Sample 1 BMT MCMV COAT	U Sample 1 BMT MCMV COAT	- NEGATIVE BMT MCMV COAT	U Sample 2 BMT MCMV COAT	U Sample 2 BMT MCMV COAT	U Sample 2 BMT MCMV COAT	- NEGATIVE BMT MCMV COAT	U Sample 3 BMT MCMV COAT	U Sample 3 BMT MCMV COAT	U Sample 3 BMT MCMV COAT	- NEGATIVE BMT MCMV COAT
B	U Sample 4 KCO MCMV COAT	U Sample 4 KCO MCMV COAT	U Sample 4 KCO MCMV COAT	- NEGATIVE KCO MCMV COAT	U Sample 5 KCO MCMV COAT	U Sample 5 KCO MCMV COAT	U Sample 5 KCO MCMV COAT	- NEGATIVE KCO MCMV COAT	U Sample 6 KCO MCMV COAT	U Sample 6 KCO MCMV COAT	U Sample 6 KCO MCMV COAT	- NEGATIVE KCO MCMV COAT
C	U Sample 7 NYA MCMV COAT	U Sample 7 NYA MCMV COAT	U Sample 7 NYA MCMV COAT	- NEGATIVE NYA MCMV COAT	U Sample 8 NYA MCMV COAT	U Sample 8 NYA MCMV COAT	U Sample 8 NYA MCMV COAT	- NEGATIVE NYA MCMV COAT	U Sample 9 NYA MCMV COAT	U Sample 9 NYA MCMV COAT	U Sample 9 NYA MCMV COAT	- NEGATIVE NYA MCMV COAT
D	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None
E	U Sample 1.2 BMT P7B	U Sample 1.2 BMT P7B	U Sample 1.2 BMT P7B	- NEGATIVE BMT P7B	U Sample 2.2 BMT P7B	U Sample 2.2 BMT P7B	U Sample 2.2 BMT P7B	- NEGATIVE BMT P7B	U Sample 3.2 BMT P7B	U Sample 3.2 BMT P7B	U Sample 3.2 BMT P7B	- NEGATIVE BMT P7B
F	U Sample 4.2 KCO P7B	U Sample 4.2 KCO P7B	U Sample 4.2 KCO P7B	- NEGATIVE KCO P7B	U Sample 5.2 KCO P7B	U Sample 5.2 KCO P7B	U Sample 5.2 KCO P7B	- NEGATIVE KCO P7B	U Sample 6.2 KCO P7B	U Sample 6.2 KCO P7B	U Sample 6.2 KCO P7B	- NEGATIVE KCO P7B
G	U Sample 7.2 NYA P7B	U Sample 7.2 NYA P7B	U Sample 7.2 NYA P7B	- NEGATIVE NYA P7B	U Sample 8.2 NYA P7B	U Sample 8.2 NYA P7B	U Sample 8.2 NYA P7B	- NEGATIVE NYA P7B	U Sample 9.2 NYA P7B	U Sample 9.2 NYA P7B	U Sample 9.2 NYA P7B	- NEGATIVE NYA P7B
H	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- NEGATIVE None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- NEGATIVE None	-- EMPTY None	-- EMPTY None	-- EMPTY None	-- EMPTY None

Appendix 4. 1:qPCR plate design

Melting Curves for selected markers

- MCMV113
- MCMV116
- MCMV245
- SCMVP3

(y-axis: -dF/dT, 0.000 to 0.003; x-axis: Temperature, 60 to 110)
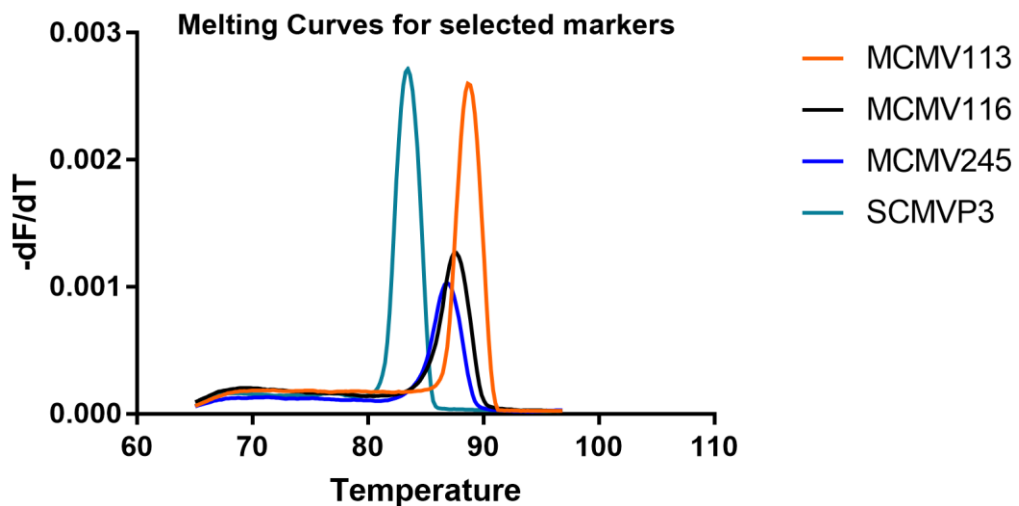
Appendix 4. 2:Melting curves for selected markers

112

Melting curves of the best performing MCMV and SCMV. A single peak illustrates the high specificity of the marker in amplifying the viral genome.

miRNA Profiles

The classification of Host-derived miRNAs. Complementary DNA (cDNAs), non-coding RNA (ncRNA), Piwi-interacting RNA (piRNA), ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), transfer RNA (tRNA).
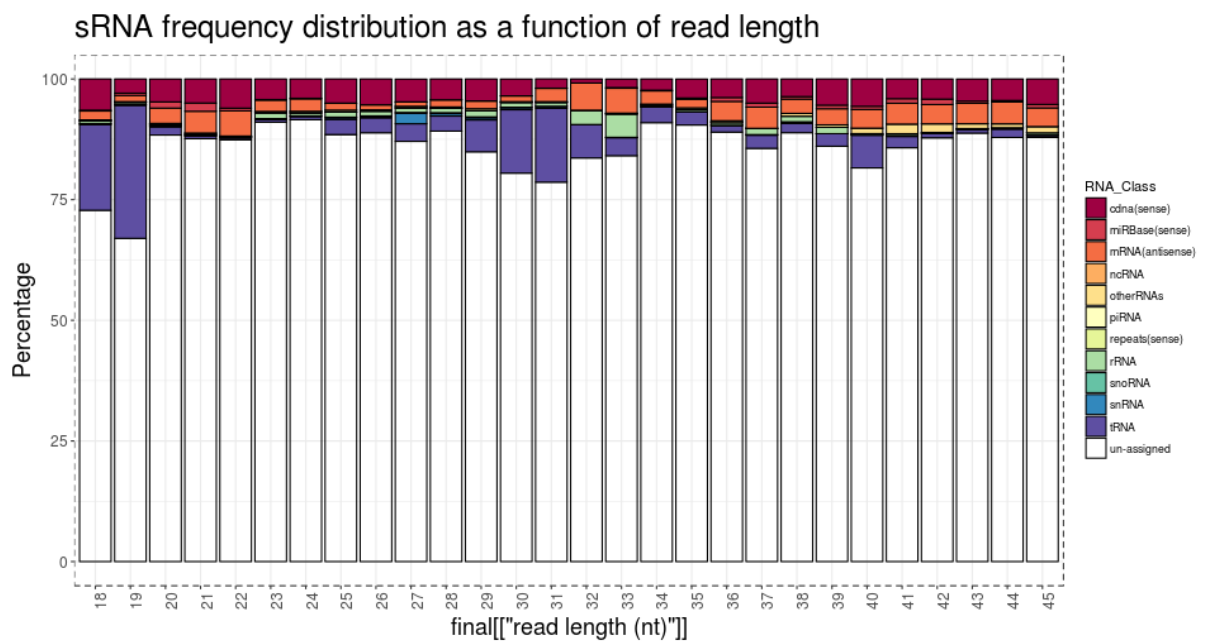
Appendix 4. 3: Classification of host derived miRNAs for Nyamira county
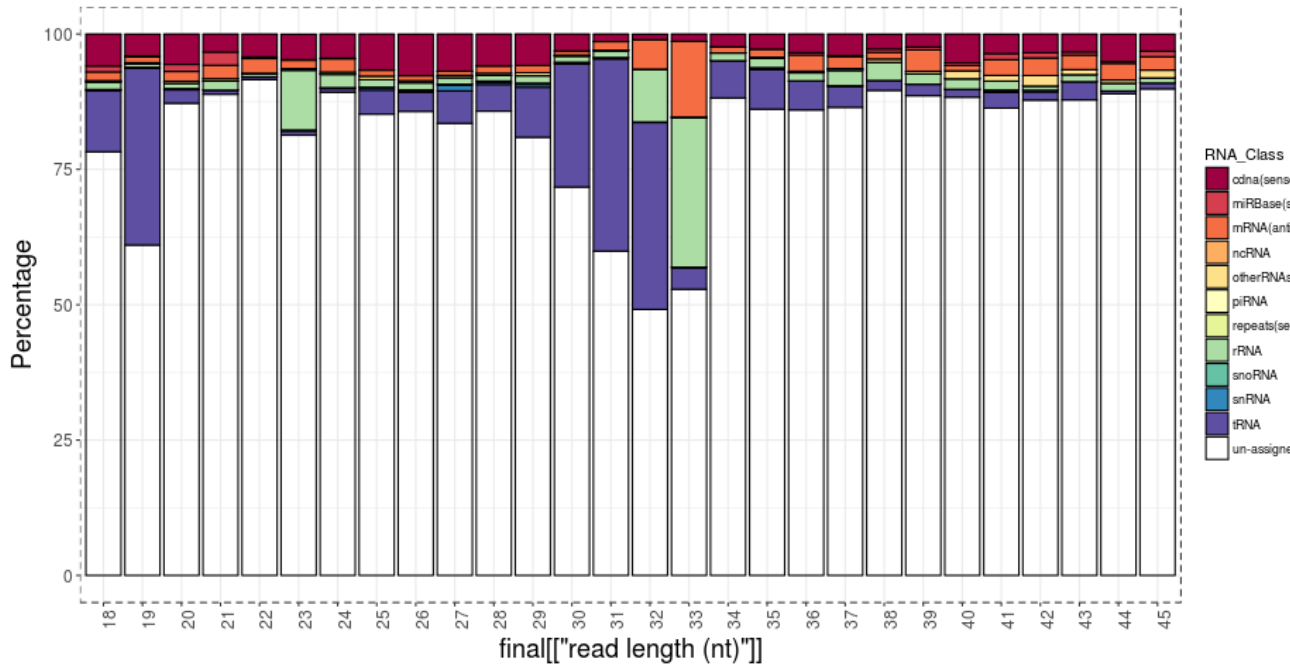
Appendix 4. 4: Host-derived miRNA profiles for Kericho samples
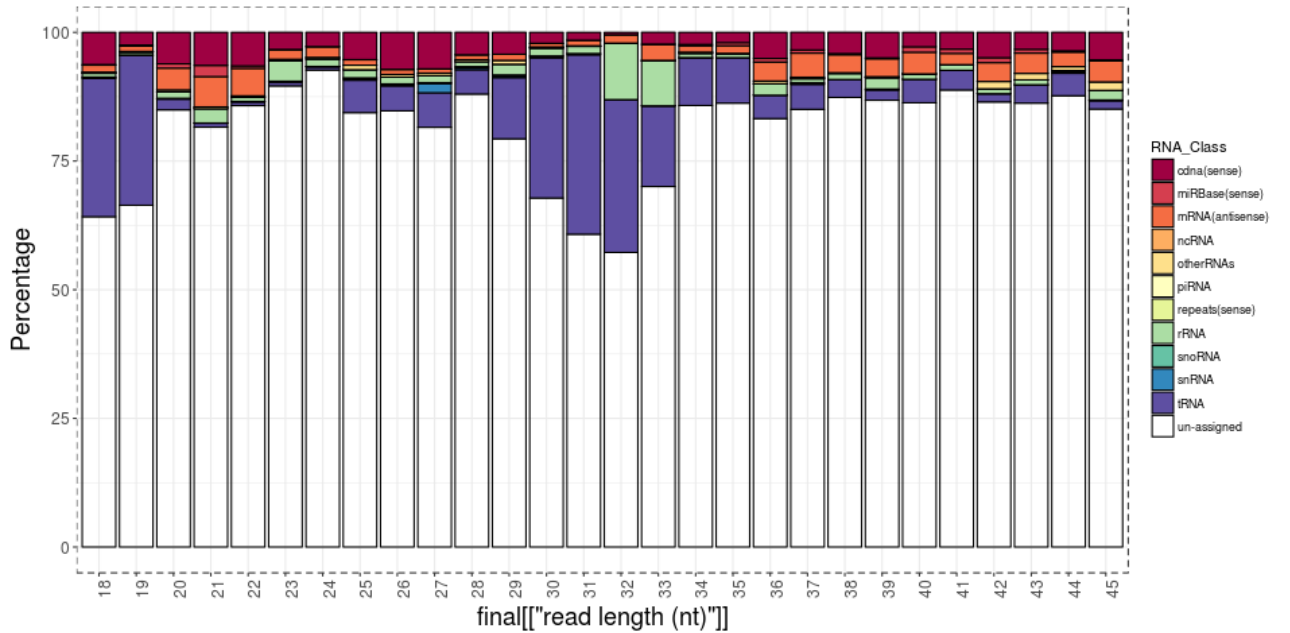
Appendix 4. 5:Host-derived miRNAs profiles for Bomet county
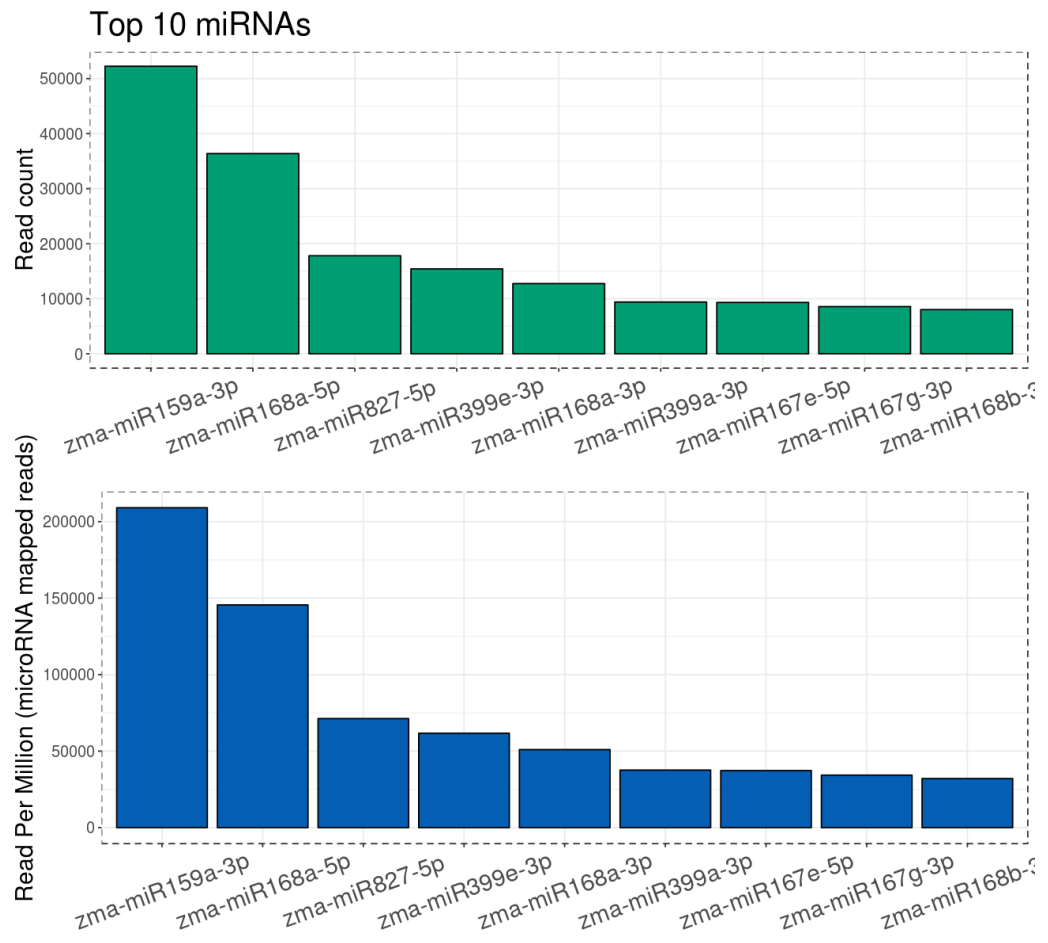
Appendix 4. 6:Top ten profiles of miRNAs identified
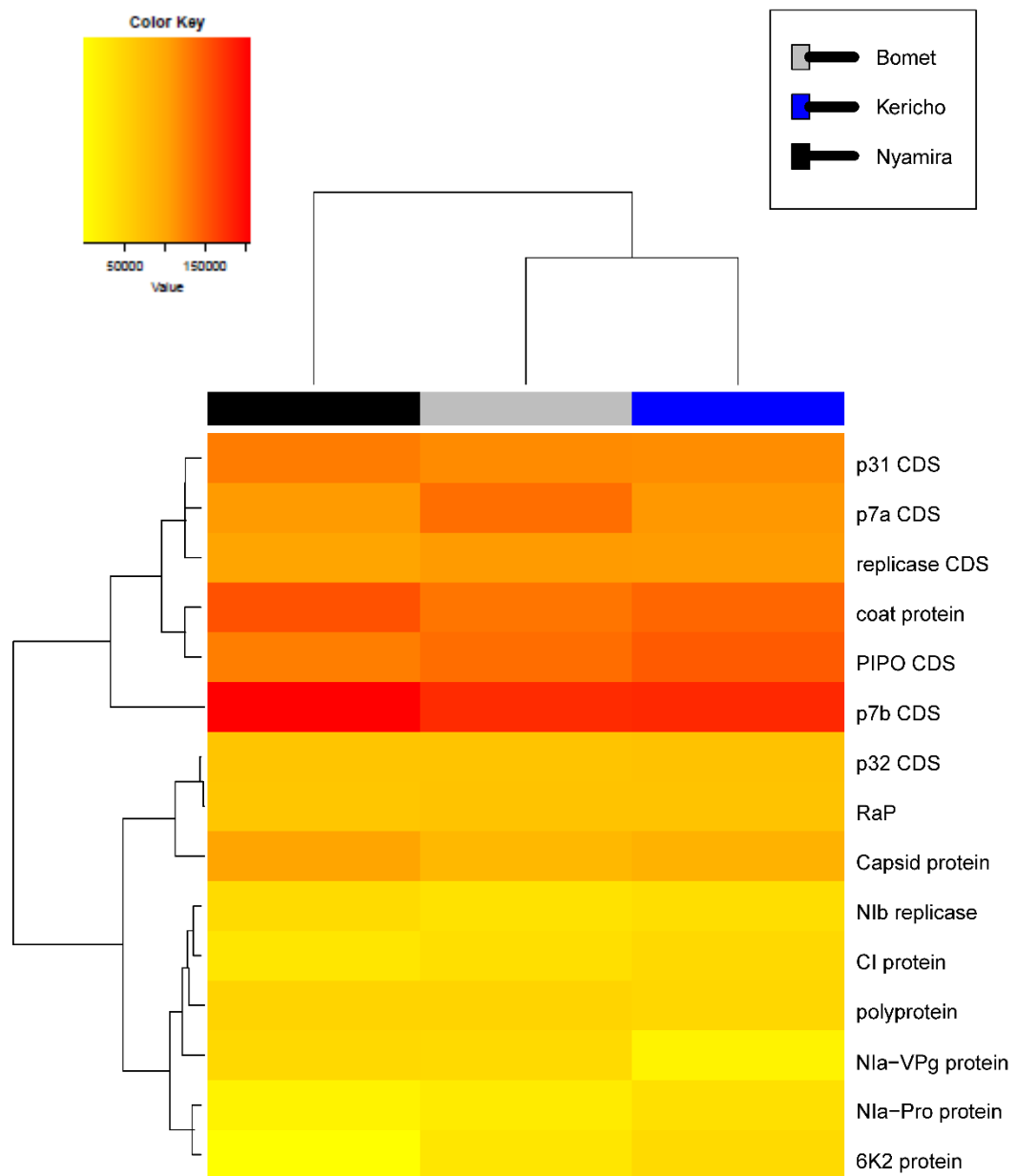
Appendix 4. 7: Combined heatmap for differential expression of MCMV and SCMV genomic regions based on RPKM values

The red-color coded regions represents the highly upregulated regions of MCMV genome. P7b domain was the highly expressed domain across the three regions.
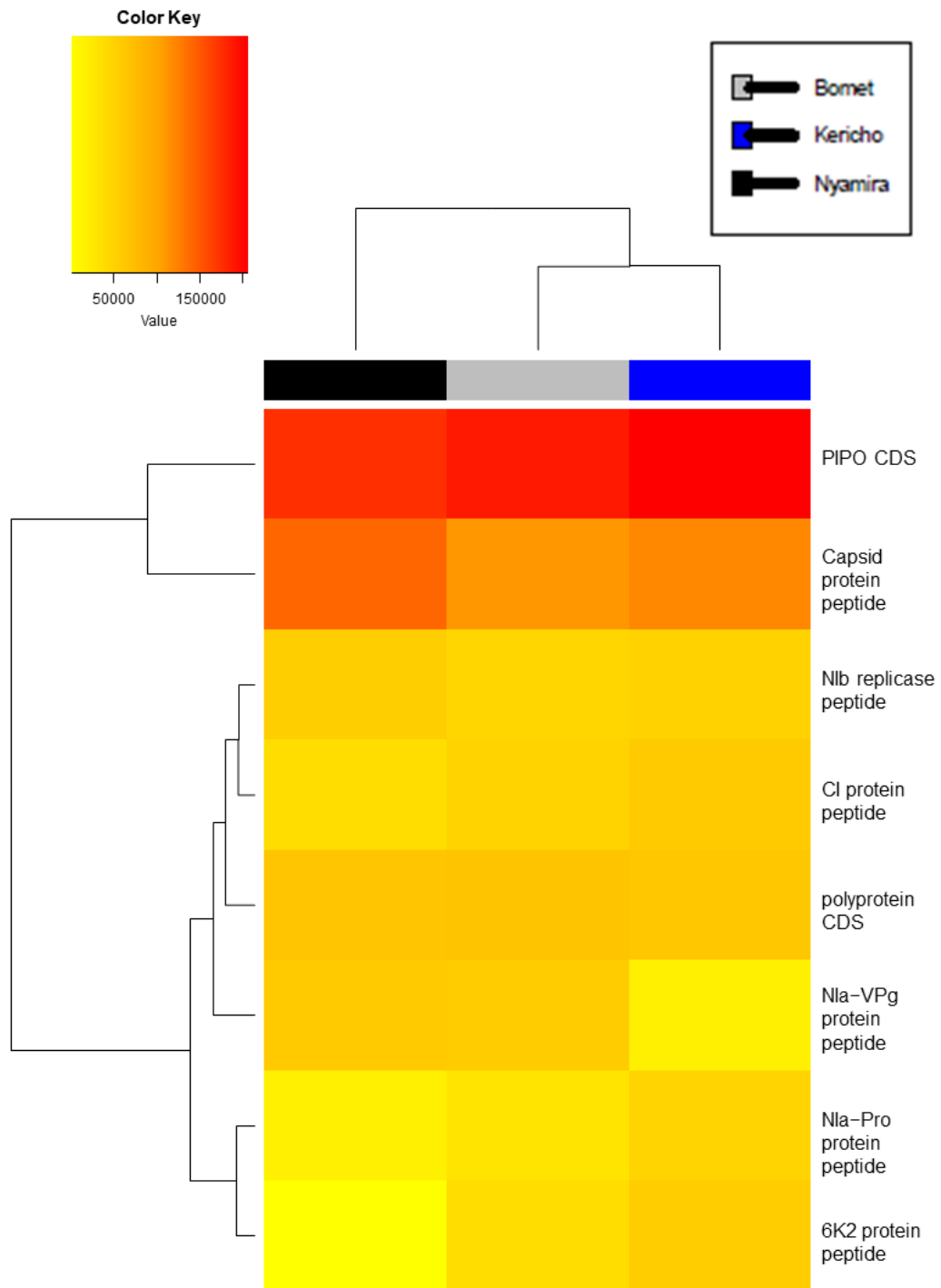
Appendix 4. 8:Heatmap-Differential expression of SCMV genomes based on RPKM values

The red-color coded regions represents the highly upregulated regions of SCMV genome. PIPO (P3) domain was the highly expressed domain across the three regions.
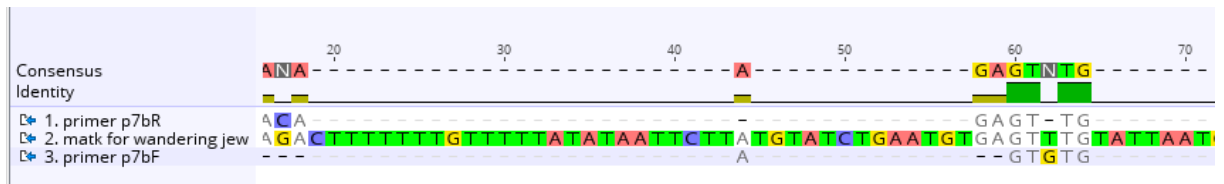
Appendix 4. 9:Alignment of MCM marker to matK gene of Commelina benghalensis
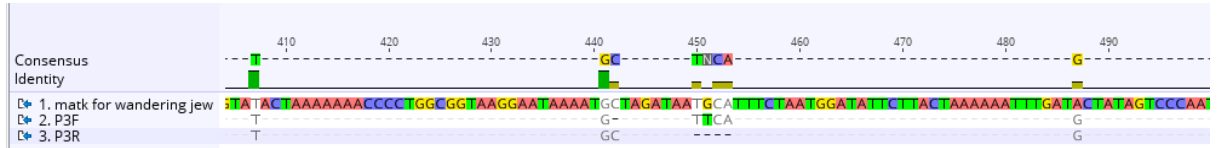
Appendix 4. 10:Alignment of SCMV marker to matK gene of Commelina benghalensis

Zea mays cultivar B73 chloroplast, complete genome
Zea mays subsp. mays cultivar B73 chloroplast, complete genome
Zea mays subsp. mays cultivar inbred line B37C chloroplast, complete genome
Zea mays subsp. huehuetenangensis plastid, complete genome
Zea mays isolate SM10 chloroplast sequence
Zea mays complete chloroplast genome
Zea mays subsp. mays cultivar inbred line B37T chloroplast, complete genome
Zea mays subsp. mays cultivar inbred line B37S chloroplast, complete genome
Zea mays subsp. mays cultivar A188 chloroplast, complete genome
PREDICTED: Zea mays inactive poly [ADP-ribose] polymerase RCD1 (LOC103654264), transcript variant X5, mRNA
lcl|Query_139427
PREDICTED: Zea mays uncharacterized LOC103635615 (LOC103635615), mRNA
PREDICTED: Zea mays Protein RETICULATA-RELATED 6 chloroplastic (LOC100191628), transcript variant X1,
Zea mays Protein RETICULATA-RELATED 6 chloroplastic (LOC100191628), mRNA
Zea mays full-length cDNA clone ZM_BFc0124N07 mRNA, complete cds
Zea mays isolate Zheng58 clone 1 GRMZM2G051891-like gene, partial sequence
Zea mays isolate 8112 clone 1 GRMZM2G051891-like gene, partial sequence
Zea mays isolate 5003 clone 1 GRMZM2G051891-like gene, partial sequence
Zea mays isolate 478 clone 1 GRMZM2G051891-like gene, partial sequence

0.1

Appendix 4. 11: Alignment of Commelina benghalensis to maize to establish their evolutionary relationship

119

Sequence Listings

gcggcaacag accccaacgc gctagacacg accttacagt tgactcacgt cttgcatcct   60

gtgagagctt tagcgtggga atttgcccct gactggcaat caggtttcat gccctctccg   120

tgcacatatg gcgaccctac cttcaattca cgcattttgg aagctgtggt ggccgatgtt   180

ctcggaggaa cggaagatga cggtggtcca agccttgagg aatggtttga cgcgcaaact   240

ctttcagatt acacgaattg cgcaacagat cctccgatgg ccaccgcaca tacgcgagag   300

aatgacatca agtcttggac ggagcttagt gaaaactttc cagaccttgt taggtatccg   360

gaaagtcttg tcgaaactct gcttgacacg gaacatgaat gtggacattt ttacgatgct   420

cctgattcct ttcagatttc tgtcacagct atgttcacag atcgttgcaa gtgccaggca   480

tgtgatccag actttcaagc acggtctctc tcgcgggctc tgcttggccc gttgccggaa   540

tccggagacg atgcggagtg gatggagcaa gcatacactc cagatgcaga attgttcgtc   600

aacgagccca ctgatgaccc cattccaacc acagactgca agagacctat ccaacccaca   660

tggtctgtgg atgtctactc gaaacaagtt gactcagact gggggctcag tgattcaaca   720

aacgtcacgg tacgttcggc tagcagttcg cttctccctg ctggccgcgg gggcatattt   780

ttcatgtgtc ctcgcccgga aggcattgaa agagtatgtc tacaggtggg agctaaaatc   840

cgatcaggaa ctcccaactc gggaacagtc acaaatctcc caggtggaac aggctatgga   900

acagaatgga gcgatgatgg ttatgaaact caatggtccg acggccccta ttctatccct   960

tcaggactgt ctgactaaac acgttctggt tccagagatt gtggacgaag aaggcacagt   1020

gacacaaaaa gagcagtcca catttctcat ccgtgagttt ggccctgccg tcagaaacct   1080

agtgactctc gcaaaactag aattcggcgg gattccaaag aaaaccactg ccaacgaact   1140

cacggtttgg cggttcctgg tacggaagtg tgagcaagca aatatgaatc caactgactc   1200

ccgtactgcc atttcgatgg ccctgccata tgtgtttatg ccgtgccgta ctgatgttgg   1260

aagagcttca attccactgc gcgatgaatc cattgagatc tgtcgccaat atcgcgcaca   1320

atttgtggag gagacacccc tacggcgagt attcaacaac ccactgagcg gaaaagcatg   1380

gcgtaattgg gttaggcacc tgggtgggct ggatgaccca gccctattcc aggagctgaa   1440

ataggggtgt cttgaagagt ggcttggggt gcagtctcgg cgcacacgtg cgcggcaccc   1500

caaactccgt agccacttta aagacacgag acacaaaacc cgcagagtgt tccggattgc   1560

aggtttgggc aacctgtatg aatttggggt ccacaacaac tctgcggtga acctggaacg   1620

tggactgatg gaacgagtat tttatgtgaa aaatgataag ggggaactcg tttcatgccc   1680

tgagccaatt agtgggatat tttggaaaaa tttgaaggga ttcagaaata gcattgttca   1740

ccatgtggga cacagacatc cggtgtcacg agagacgttt ctcgcttact atacagggcc   1800

aaagcgcacc atgtatgaga aagcggtgaa ttccctctac gaaatgcctg tctcgtatga   1860

cgatgctaaa ctgaaaacat ttgtcaaagc agaaaagatc aatctcacta agaaagctga   1920

tcctgtacca cgtgtgatcc aaccccgtgc tccccggtac aatgtggaac tgggtaggta   1980

tttgagaccg gttgagcatc caatctacca tgccatagac aagatttggg gtggtccgac   2040

tatcatgaaa ggttactcgg ttgaacagat tggaagacac atcgaaaacg cgtttcgatc   2100

gttcaccgat ccagtagcca ttggatttga tgcaagtcgg tttgatcaac acgtgagtgt   2160

tgaagctcta cgctgggaac atagcgtata cacaaggatt tatggctatc cagagcttct   2220

cacacaattg ctgcggtggc agattcacaa tcgtggaaca gcatacgctt cagacggcgc   2280

gttcaactac caggtggatg gtaaacgcat gagtggagac atgaacacat cattgggcaa   2340

ttgcatcctg gccacagcaa tcacacatga cttcgtgacc aagctgggga tcccagccag   2400

attaatcaac aatggagatg acaacgtcct gatctgtccg gcagtggagg taggcagagt   2460

cagacaggaa ctgtacagac attggttgaa ctacggattt gaagtcatat cagaagaacc   2520

tgtatacatc cttgaacagg ttgagttttg ccagatgcga ccagtctttg acggtacaca   2580

atacaccatg atgagagacc cacgcaccac catgtccaag gatgcatatg ctgttacccc   2640

tttttaacacg ccaaccgcag caaggaggtg gatgagagca gttggggaat gcggactcag   2700

ccttacagga ggcctccccg taaagcagga atactatgct gctctcgtga aacacggact   2760

agatccgaaa aacattaaac agggaaaaga ctttgacagt gggctatatt atcttagcaa   2820

actatccaat cgcaaatggc aggaggtcca agaaagcgct cgttactcat tctggcttgc   2880

cttcgggtac acaccagatg aacaacgagc gttagaggag tatttagat cttggactcc   2940

tacgtttgaa tggtccacaa cgggtatttt ggcagaaatc cccgaatgtc ttcttctcaa   3000

acacaatccc ctcccaccga cgtagcccgc agacaacaga ccggcactgc cgtgcgatcc   3060

gaagataacc atcgcaatca gttggaaaac attgctgttg gaaaactcac caaatctgaa   3120

ggagcacccg cccaacagaa cgtgattatt gcaaaagagg tggttatcaa taaccacttc   3180

aatttcaact gagctggagt gtgtgtgtgt agaatcgtcc tggcctcagt ggttaaggaa   3240

tttaatcttg gggattttga tctcctcgat tctgttcatc ttgacgaaaa cacaggacac   3300

cgttgccgtt tatacgagc catcggtgta ctcgatcgat caaacacaaa aattccaaaa   3360

gatcgacata cacaatggcg gcaagtagcc ggtctacccg aggtagaaag cagcgcggac   3420

gtagcgtgga agcaaaatcc agagctattc gagccaaccc gcctgtccct cgacccaacc   3480

cgcagcgaaa ccgtccccca cctgcgggaa caacctgctc catgtctgaa attctgcttg   3540

cagtgtcagc aacaactgct gaccaaattc tcgagattcc agtgtgcgca gggattgact   3600

tcccagctgg aacgccaccc cgatacattg gggcggccaa gtggctggca gcacaatcac   3660

agatgtggaa cacaattgtg ttcaactctg tgcgcatcac ttgggaaaca ttcacagcag   3720

121

acaccactag cggatacatc tcaatggcat tcctctctga ttacatgcta tcaataccca   3780

ctggggtgga ggatgttgcc aggatcgtgc cctcagctac aatagctctg aagaacagag   3840

ggccgtccat tgttatgcca caaaaccgca ctgtgttcag gtgtatacag gctggtcagt   3900

ttgctgcgtt gggcagcgcg gctgacaagc aaatgtattc tcccggacga ttcattgtgg   3960

ctatccccaa agctagtgca acacaagcgg taggccaaat caaaatctcg tactccgtgt   4020

cctaccgtgg agcagcaatc ctacaacctg ccctggttcc aggcccaggg ctggcaaatc   4080

attgagcata aggtgagccg gcatgaggtt gcaagaccgg aacaaccagt ccttctggca   4140

gagtcctgcc aatccaaagt gtctgggagc ctaaatcgta tactagtagt ttgcgtgatg   4200

accatgactg gagagtgggc ggcggctgcc aaccgcagac tgggcgtata tagtaagcct   4260

tgacccaccc atggaaaaca cgtatttacg aacggtacga catggtaact ggatacactt   4320

aaccctgggg caagtagatg ctaggaaact agcatcgggc cgcccacgag ggtttctgaa   4380

ctcaacggag tagaacgggg gggggtaatg cccctctctt ccggccctaa   4430

Sequence Listings 4. 1:Borabu MCMV isolate

gcggcaacag accccaacgc gctagacacg accttacagt tgactcacgt cttgcatcct   60

gtgagagctt tagcgtggga atttgcccct gactggcaat caggtttcat gccctctccg   120

tgcacatatg gcgaccctac cttcaattca cgcattttgg aagctgtggt ggccgatgtt   180

ctcggaggaa cggaagatga cggtggtcca agccttgagg aatggtttga cgcgcaaact   240

ctttcagatt acacgaattg cgcaacagat cctccgatgg ccaccgcaca tacgcgagag   300

aatgacatca agtcttggac ggagcttagt gaaaactttc cagaccttgt taggtatccg   360

gaaagtcttg tcgaaactct gcttgacacg gaacatgaat gtggacattt ttacgatgct   420

cctgattcct ttcagatttc tgtcacagct atgttcacag atcgttgcaa gtgccaggca   480

tgtgatccag actttcaagc acggtctctc tcgcggggctc tgcttggccc gttgccggaa   540

tccggagacg atgcggagtg gatggagcaa gcatacactc cagatgcaga attgttcgtc   600

aacgagccca ctgatgaccc cattccaacc acagactgca agagacctat ccaacccaca   660

tggtctgtgg atgtttactc gaaacaagtt gactcagact gggggctcag tgattcaaca   720

aacgtcacgg tacgttcggc tagcagttcg cttctccctg ctggccgcgg gggcatattt   780

ttcatgtgtc ctcgccccgga aggcattgaa agagtatgtc tacaggtggg agctaaaatc   840

cgatcaggaa ctcccaactc gggaacagtc acaaatctcc caggtggaac aggctatgga   900

acagaatgga gcgatgatgg ttatgaaact caatggtccg acggcccccta ttctatccct   960

122

tcaggactgt ctgactaaac acgttctggt tccagagatt gtggacgaag aaggcacagt　1020

gacacaaaaa gagcagtcca catttctcat ccgtgagttt ggccctgccg tcagaaacct　1080

agtgactctc gcaaaactag aattcggcgg gattccaaag aaaaccactg ccaacgaact　1140

cacggtttgg cggttcctgg tacggaagtg tgagcaagca aatatgaatc caactgactc　1200

ccgtactgcc atttcgatgg ccctgccata tgtgtttatg ccgtgccgta ctgatgttgg　1260

aagagcttca attccactgc gcgatgaatc cattgagatc tgtcgccaat atcgcgcaca　1320

atttgtggag gagacacccc tacggcgagt attcaacaac ccactgagcg gaaaagcatg　1380

gcgtaattgg gttaggcacc tgggtgggct ggatgaccca gccctattcc aggagctgaa　1440

ataggggtgt cttgaagagt ggcttggggt gcagtctcgg cgcacacgtg cgcggcaccc　1500

caaactccgt agccacttta aagacacgag acataaaacc cgcagagtgt tccggattgc　1560

aggtttgggc aacctgtatg aatttggggt ccacaacaac tctgcggtga acctggaacg　1620

tggactgatg gaacgagtat tttatgtgaa aaatgataag ggggaactcg tttcatgccc　1680

tgagccaatt agtgggatat tttggaaaaa tttgaaggga ttcagaaata gcattgttca　1740

ccatgtggga cacagacatc cggtgtcacg agagacgttt ctcgcttact atacagggcc　1800

aaagcgcacc atgtatgaga aagcggtgaa ttccctctac gaaatgcctg tctcgtatga　1860

cgatgctaaa ctgaaaacat ttgtcaaagc agaaaagatc aatctcacta agaaagctga　1920

tcctgtacca cgtgtgatcc aaccccgtgc tccccggtac aatgtggaac tgggtaggta　1980

tttgagaccg gttgagcatc caatctacca tgccatagac aagatttggg gtggtccgac　2040

tatcatgaaa ggttactcgg ttgaacagat tggaagacac atcgaaaacg cgtttcgatc　2100

gttcaccgat ccagtagcca ttggatttga tgcaagtcgg tttgatcaac acgtgagtgt　2160

tgaagctcta cgctgggaac atagcgtata cacaaggatt tatggctatc cagagcttct　2220

cacacaattg ctgcggtggc agattcacaa tcgtggaaca gcatacgctt cagacggcgc　2280

gttcaactac caggtggatg gtaaacgcat gagtggagac atgaacacat cattgggcaa　2340

ttgcatcctg gccacagcaa tcacacatga cttcgtgacc aagctgggga tcccagccag　2400

attaatcaac aatggagatg acaacgtcct gatctgtccg gcagtggagg taggcagagt　2460

cagacaggaa ctgtacagac attggttgaa ctacggattt gaagtcatat cagaagaacc　2520

tgtatacatc cttgaacagg ttgagttttg ccagatgcga ccagtctttg acggtacaca　2580

atacaccatg atgagagacc cacgcaccac catgtccaag gatgcatatg ctgttacccc　2640

ttttaacacg ccaaccgcag caaggaggtg gatgagagca gttggggaat gcggactcag　2700

ccttacagga ggcctccccg taaagcagga atactatgct gctctcgtga aacacggact　2760

agatccgaaa aacattaaac agggaaaaga ctttgacagt gggctatatt atcttagcaa　2820

actatccaat cgcaaatggc aggaggtcca agaaagcgct cgttactcat tctggcttgc　2880

cttcgggtac acaccagatg aacaacgagc gttagaggag tattttagat cttggactcc   2940

tacgtttgaa tggtccacaa cgggtatttt ggcagaaatc cccgaatgtc ttcttctcaa   3000

acacaatccc ctcccaccga cgtagcccgc agacaacaga ccggcactgc cgtgcgatcc   3060

gaagataacc atcgcaatca gttggaaaac attgctgttg gaaaactcac caaatctgaa   3120

ggagcacccg cccaacagaa cgtgattatt gcaaaagagg tggttatcaa taaccacttc   3180

aatttcaact gagctggagt gtgtgtgtgt agaatcgtcc tggcctcagt ggttaaggaa   3240

tttaatcttg gggattttga tctcctcgat tctgttcatc ttgacgaaaa cacaggacac   3300

cgttgccgtt tatcacgagc catcggtgta ctcgatcgat caaacacaaa aattccaaaa   3360

gatcgacata cacaatggcg gcaagtagcc ggtctacccg aggtagaaag cagcgcggac   3420

gtagcgtgga agcaaaatcc agagctattc gagccaaccc gcctgtccct cgacccaacc   3480

cgcagcgaaa ccgtccccca cctgcgggaa caacctgctc catgtctgaa attctgcttg   3540

cagtgtcagc aacaactgct gaccaaattc tcgagattcc agtgtgcgca gggattgact   3600

tcccagctgg aacgccaccc cgatacattg gggcggccaa gtggctggca gcacaatcac   3660

agatgtggaa cacaattgtg ttcaactctg tgcgcatcac ttgggaaaca ttcacagcag   3720

acaccactag cggatacatc tcaatggcat tcctctctga ttacatgcta tcaatacccca   3780

ctggggtgga ggatgttgcc aggatcgtgc cctcagctac aatagctctg aagaacagag   3840

ggccgtccat tgttatgcca caaaaccgca ctgtgttcag gtgtatacag gctggtcagt   3900

ttgctgcgtt gggcagcgcg gctgacaagc aaatgtattc tcccggacga ttcattgtgg   3960

ctatccccaa agctagtgca acacaagcgg taggccaaat caaaatctcg tactccgtgt   4020

cctaccgtgg agcagcaatc ctacaacctg ccctggttcc aggcccaggg ctggcaaatc   4080

attgagcata aggtgagccg gcatgaggtt gcaagaccgg aacaaccagt ccttctggca   4140

gagtcctgcc aatccaaagt gtctgggagc ctaaatcgta tactagtagt ttgcgtgatg   4200

accatgactg gagagtgggc ggcggctgcc aaccgcagac tgggcgtata tagtaagcct   4260

tgacccaccc atggaaaaca cgtatttacg aacggtacga catggtaact ggatacactt   4320

aaccctgggg caagtagatg ctaggaaact agcatcgggc cgcccacgag ggtttctgaa   4380

ctcaacggag tagaacgggg gggggtaatg cccctctctt ccggccctaa ta   4432

Sequence Listings 4. 2:Kericho MCMV Isolates

gcggcaacag accccaacgc gctagacacg accttacagt tgactcacgt cttgcatcct   60

gtgagagctt tagcgtggga atttgcccct gactggcaat caggtttcat gccctctccg   120

tgcacatatg gcgaccctac cttcaattca cgcattttgg aagctgtggt ggccgatgtt   180

ctcggaggaa cggaagatga cggtggtcca agccttgagg aatggtttga cgcgcaaact   240

cttttcagatt acacgaattg cgcaacagat cctccgatgg ccaccgcaca tacgcgagag   300

aatgacatca agtcttggac ggagcttagt gaaaactttc cagaccttgt taggtatccg   360

gaaagtcttg tcgaaactct gcttgacacg gaacatgaat gtggacattt ttacgatgct   420

cctgattcct ttcagatttc tgtcacagct atgttcacag atcgttgcaa gtgccaggca   480

tgtgatccag actttcaagc acggtctctc tcgcgggctc tgcttggccc gttgccggaa   540

tccggagacg atgcggagtg gatggagcaa gcatacactc cagatgcaga attgttcgtc   600

aacgagccca ctgatgaccc cattccaacc acagactgca agagacctat ccaacccaca   660

tggtctgtgg atgtctactc gaaacaagtt gactcagact gggggctcag tgattcaaca   720

aacgtcacgg tacgttcggc tagcagttcg cttctccctg ctggccgcgg gggcatattt   780

ttcatgtgtc ctcgcccgga aggcattgaa agagtatgtc tacaggtggg agctaaaatc   840

cgatcaggaa ctcccaactc gggaacagtc acaaatctcc caggtggaac aggctatgga   900

acagaatgga gcgatgatgg ttatgaaact caatggtccg acggcccta ttctatccct   960

tcaggactgt ctgactaaac acgttctggt tccagagatt gtggacgaag aaggcacagt   1020

gacacaaaaa gagcagtcca catttctcat ccgtgagttt ggccctgccg tcagaaacct   1080

agtgactctc gcaaaactag aattcggcgg gattccaaag aaaaccactg ccaacgaact   1140

cacggtttgg cggttcctgg tacggaagtg tgagcaagca aatatgaatc caactgactc   1200

ccgtactgcc atttcgatgg ccctgccata tgtgtttatg ccgtgccgta ctgatgttgg   1260

aagagcttca attccactgc gcgatgaatc cattgagatc tgtcgccaat atcgcgcaca   1320

atttgtggag gagacacccc tacggcgagt attcaacaac ccactgagcg gaaaagcatg   1380

gcgtaattgg gttaggcacc tgggtgggct ggatgaccca gccctattcc aggagctgaa   1440

ataggggtgt cttgaagagt ggcttggggt gcagtctcgg cgcacacgtg cgcggcaccc   1500

caaactccgt agccacttta aagacacgag acacaaaacc cgcagagtgt tccggattgc   1560

aggtttgggc aacctgtatg aatttggggt ccacaacaac tctgcggtga acctggaacg   1620

tggactgatg gaacgagtat tttatgtgaa aaatgataag ggggaactcg tttcatgccc   1680

tgagccaatt agtgggatat tttggaaaaa tttgaaggga ttcagaaata gcattgttca   1740

ccatgtggga cacagacatc cggtgtcacg agagacgttt ctcgcttact atacagggcc   1800

aaagcgcacc atgtatgaga aagcggtgaa ttccctctac gaaatgcctg tctcgtatga   1860

cgatgctaaa ctgaaaacat ttgtcaaagc agaaaagatc aatctcacta agaaagctga   1920

tcctgtacca cgtgtgatcc aacccccgtgc tccccggtac aatgtggaac tgggtaggta   1980

tttgagaccg gttgagcatc caatctacca tgccatagac aagatttggg gtggtccgac   2040

tatcatgaaa ggttactcgg ttgaacagat tggaagacac atcgaaaacg cgtttcgatc   2100

gttcaccgat ccagtagcca ttggatttga tgcaagtcgg tttgatcaac acgtgagtgt   2160

tgaagctcta cgctgggaac atagcgtata cacaaggatt tatggctatc cagagcttct   2220

cacacaattg ctgcggtggc agattcacaa tcgtggaaca gcatacgctt cagacggcgc   2280

gttcaactac caggtggatg gtaaacgcat gagtggagac atgaacacat cattgggcaa   2340

ttgcatcctg gccacagcaa tcacacatga cttcgtgacc aagctgggga tcccagccag   2400

attaatcaac aatggagatg acaacgtcct gatctgtccg gcagtggagg taggcagagt   2460

cagacaggaa ctgtacagac attggttgaa ctacggattt gaagtcatat cagaagaacc   2520

tgtatacatc cttgaacagg ttgagttttg ccagatgcga ccagtctttg acggtacaca   2580

atacaccatg atgagagacc cacgcaccac catgtccaag gatgcatatg ctgttacccc   2640

ttttaacacg ccaaccgcag caaggaggtg gatgagagca gttggggaat gcggactcag   2700

ccttacagga ggcctccccg taaagcagga atactatgct gctctcgtga aacacggact   2760

agatccgaaa aacattaaac agggaaaaga ctttgacagt gggctatatt atcttagcaa   2820

actatccaat cgcaaatggc aggaggtcca agaaagcgct cgttactcat tctggcttgc   2880

cttcgggtac acaccagatg aacaacgagc gttagaggag tattttagat cttggactcc   2940

tacgtttgaa tggtccacaa cgggtatttt ggcagaaatc cccgaatgtc ttcttctcaa   3000

acacaatccc ctcccaccga cgtagcccgc agacaacaga ccggcactgc cgtgcgatcc   3060

gaagataacc atcgcaatca gttggaaaac attgctgttg gaaaactcac caaatctgaa   3120

ggagcacccg cccaacagaa cgtgattatt gcaaaagagg tggttatcaa taaccacttc   3180

aatttcaact gagctggagt gtgtgtgtgt agaatcgtcc tggcctcagt ggttaaggaa   3240

tttaatcttg gggattttga tctcctcgat tctgttcatc ttgacgaaaa cacaggacac   3300

cgttgccgtt tatcacgagc catcggtgta ctcgatcgat caaacacaaa aattccaaaa   3360

gatcgacata cacaatggcg gcaagtagcc ggtctacccg aggtagaaag cagcgcggac   3420

gtagcgtgga agcaaaatcc agagctattc gagccaaccc gcctgtccct cgacccaacc   3480

cgcagcgaaa ccgtcccca cctgcgggaa caacctgctc catgtctgaa attctgcttg   3540

cagtgtcagc aacaactgct gaccaaattc tcgagattcc agtgtgcgca gggattgact   3600

tcccagctgg aacgccaccc cgatacattg gggcggccaa gtggctggca gcacaatcac   3660

agatgtggaa cacaattgtg ttcaactctg tgcgcatcac ttgggaaaca ttcacagcag   3720

acaccactag cggatacatc tcaatggcat tcctctctga ttacatgcta tcaataccca   3780

ctgggggtgga ggatgttgcc aggatcgtgc cctcagctac aatagctctg aagaacagag   3840

ggccgtccat tgttatgcca caaaaccgca ctgtgttcag gtgtatacag gctggtcagt   3900

ttgctgcgtt gggcagcgcg gctgacaagc aaatgtattc tcccggacga ttcattgtgg   3960

ctatccccaa agctagtgca acacaagcgg taggccaaat caaaatctcg tactccgtgt   4020

cctaccgtgg agcagcaatc ctacaacctg ccctggttcc aggcccaggg ctggcaaatc   4080

attgagcata aggtgagccg gcatgaggtt gcaagaccgg aacaaccagt ccttctggca   4140

gagtcctgcc aatccaaagt gtctgggagc ctaaatcgta tactagtagt ttgcgtgatg   4200

accatgactg gagagtgggc ggcggctgcc aaccgcagac tgggcgtata tagtaagcct   4260

tgacccaccc atggaaaaca cgtatttacg aacggtacga catggtaact ggatacactt   4320

aaccctgggg caagtagatg ctaggaaact agcatcgggc cgcccacgag ggtttctgaa   4380

ctcaacggag tagaacgggg gggggtaatg cccctctctt ccggccctaa g   4431

Sequence Listings 4. 3:3: Bomet MCMV isolate

aacacaacta gctattaaga aaaacaacaa aactcaacaa cacacaacaa aacacaacca   60

aacaacacca agttttcttt gctcagattg tagtgaacgg ctcgcaagaa acggttcttc   120

gagatcactc tctgacttct ttctctctca cacttgcatt caagcggaat ggcgggatcg   180

tggactcacg tgtcttacaa gtggcagcca aatgtcaaca atgaccgtga tgtaagaaag   240

gtaatggaaa tgtttgcagc aaaacatcaa cattatacag aggagcagcg acttgcacat   300

aacagcaagt tgctaaggaa agcgtatgtt gtggacgttg aaccagtgaa gccagcaccg   360

gagcctataa ggcgtaaagt gtgggtggaa aaattcgacc acaacccaac cgaagacctg   420

gtgtatccgc gccttgtcac ggttaaaaag gcagcagaaa tgaagcccgt gaacacctct   480

ataaataaac ttataaggga tgttctggac atatcaaaag ggagcagcct taaacttgag   540

ctaattggta aacgccagaa gtgcagaaca caactagcta ttaagaaata caacaataag   600

gactacctcc attgcaggac acgtcatgaa gacaacatgt tcaagaggag ggacgtggca   660

atcggcattg aatggatccc aaccattgaa gcgattgctc gatgctacag cacagtaaac   720

aaacaggaga tgcaaagtct ctacaaaggc agtagtggct taacattcat gcaaaatgat   780

gaattattca ttgttagagg cagaatgaat ggcgagttag ttaatagcct ggaagagaac   840

aagaacgtgc tcgacatcga tcactacgct gatccacaag cgaacgactt ttggaaaggg   900

tacactgatg catacgttgc gaatcgcagt atctccacta cacataccgt gcatactcca   960

acaatcaact ggtagagtg tggaaagaga atggcattac ttgaaatttt attccactca   1020

actttttaaga ttacatgcaa gacatgcaat attgatgact tggagttatc agacgatgaa   1080

127

tttggagcaa aacttttcaa gaatctacaa cgtatcgagg aacagcaacg cgaatatctc   1140

gctaaagatc aaaagctgcg tcgcatgatt caattcatca aagaaaggtg taatccaaag   1200

ttctctcatc taccattgct ttggcaagtt gcagaaacaa taggacatta tacggacaat   1260

cagtcaaagc aaataattga tatcagtgaa gcactcatca aagtgaacac attaacacct   1320

gatgatgcag tcaaggcgag cgtagcactg cttgaggtgg cacgatggta taaaaatagg   1380

aaagaatcac ttaaaacaga cacattagat tcgttccgaa acaagatatc cccaaagagc   1440

acgataaaca cagcattgat gtgtgataat caactagaca agaatgcaaa ttttgtctgg   1500

ggaaacaggg agtatcatgc aaagcgcttc ttctcaaatt actttgaagc agtagatcct   1560

actgacgcat acgagaagca tgtaacccgt ttcaatccaa atggtcaacg gaaattatca   1620

attggtaaac tagtaattcc gttggatttc cagaaaataa gagactcatt cgttggtttg   1680

gcaataaaca aacagccact tacaaaggct tgtgtgagca aaatcgatgg tggttatgta   1740

tacccatgct gttgtgtcac aacagaattt ggaaaaccag catactctga gataatacca   1800

ccaacaaagg gccatataac aattggaaat tctgtagacc caaagatcgt agacctgcca   1860

aatacaacac caccaagtat gtacattgcc aaagacggat attgttacat caacattttc   1920

ttagcagcaa tgataaatgt caatgaagac tcagcaaagg actatacaaa attcctcagg   1980

gatgaattag tgagcgactt ggaaagtggc cgaaacttaa agatgttgca acagcgtgct   2040

atgcgttatc agttatgttt ccagagatta agaacgctga attgcctcca atactcgttg   2100

accacgaaag caagtcaatg cacgtgattg attcatacgg ctcattgagc gttggattcc   2160

acatccttaa agcaagcact gttggccagc tcataaaatt tcaatacgag tcaatggata   2220

gtgagatgcg cgagtatata gtgggaggta ctttaacgac ccaaacattt aacaaactcc   2280

ttacatcttt agctaaaaat atgttcaaac cagatcaaat taagcagatg attgaggagg   2340

aacccttcct actaatgatg gcaatcgcgt caccaactat gcttatagcg ctatacaata   2400

attgctatat agagcaggct atgacgtatt ggatcgttaa aaaccaagga gtggcagcga   2460

tattcgcaca actggaagca ctagcaaaga agacctcaca agctgaattg ctagtccaac   2520

aaatgcaaat acttgaaaag gcgtctaatc aattgaggct tgcagtcacg ggtttaaacc   2580

atgtcgatcc agccaagcgt ctgctctggt cgcatctaga agcaatgaca acacgatcag   2640

agatgaacaa ggaattaatt gctgaaggct atgccctgta tgatgagcgc ctgtatgcac   2700

ttatggaaaa aagttacgta gatcaattaa accaatcatg gcagagtta tcattctgtg   2760

gaaaattttc agcaatatgg cgtgtgttca aagtcaagaa atattacaag ccgtctttaa   2820

ccgtgagaaa aagcgtagat ttaggcgctg tgtacaatat atcagctacg catctaatat   2880

cagatttagc gcagagaagt cgcgatcgag ccagctctat tttaaccaaa ctccgcaacg   2940

gtttttatga taaggtagag aaggctagaa ctcgagcaat taaaactgtt tattggttca   3000

128

tacctgacat atttagactt atgcatattt tcatagttct tagtttatta acaactgtgg   3060

caaacactat tattgtgact atgaatgatt acaaaaagct gaaaaagcaa caaagagagg   3120

atgagtatga agctgaaatt aacgaggttc gaaaaattca cgctaatctg atgaaagagc   3180

ataatgacaa tctaacatgt gaacagttta tcgaacatat gcgccagacg catccacgat   3240

taattgaagc cacattggag ttgacacaca caggcgtcat tcatgaagga aagagcaacc   3300

tagaaaccaa tctcgagcaa gctatggcag tgggcacttt actcacaatg atgttagacc   3360

cacagaagag tgatgcagta tacaaagtgt tgaataagat gagaactgta attagcacat   3420

ttgagcagaa tgtcccattt ccttcaatca actttacaaa tattttaaca ccatcagtgg   3480

cgcaacaaag tgtagatgtt gatgaacctc taactttaag tactgacaag aatctaacaa   3540

tagattttga tactaatcaa gatttaccag cggacacgtt tagcaatgat gttactttcg   3600

aagattggtg ggcaaatcaa ataagcaaca atagaacagt tccacattat cggcttggtg   3660

gcaaatttgt agaattcaca agagaaaatg cagcacacgt tagcattgag ctcgcacact   3720

caaatcttga gaaggagttt ctactcagag gagctgttgg ctcaggaaaa tccactggtc   3780

ttccatatca tcttagtatg cgtggtaagg tgcttttaat agaaccaaca agaccactag   3840

ccgagaatgt ctgcagacag cttcaaggac caccatttaa tgtcagcccg acactacaaa   3900

tgcgtggttt aagctctttt ggttgcacgc caattacgat aatgacatca ggtttcgcac   3960

tgcacatgta tgcaaataat ccggataaga tctctgatta cgattttata atatttgatg   4020

agtgtcatat catggaagcg ccagccatgg cgttctactg tttgttgaaa gaatatgaat   4080

atagaggcaa aatcattaag gtgtccgcaa caccaccagg acgagagtgt gagttctcta   4140

ctcagcatcc agtcgatatc catgtgtgcg agaatctcac tcagcaacaa ttcgttagag   4200

agcttggttc cggatcaaac gttgacgcaa caaagtatgg aaacaacata cttgtgtatg   4260

ttgcaagtta caacgacgtc gactcattag cacatgcttt aactgaatta cactattcag   4320

tcataaaggt tgatgggaga acaatgaaac agaacactac agggatagtg acaaatggca   4380

cctctcaaaa gaagtgtttt gtcgttgcta cgaatatcat tgagaatggt gtcacattag   4440

acgttgatgt tgtggttgac ttcggactca aagtaacagc tgaattagat gttgacaaca   4500

gagctatttt atacaagcgt gttagcattt catatggtga gcgaattcaa aggctagggc   4560

gtgttggaag aaacaaaccc ggaacagtcg ttcgaattgg aaaaaccatg aagggtttgc   4620

aagaaatccc agccatgata gcgacagaag cagcattcat gtgctttgca tatggactaa   4680

aggtcataac acacaatgtt tcaacaacac acttagcaaa gtgcacagtc aagcaagcgc   4740

gaacaatgat gcaatttgag atttcaccat ttgtaatggc agaactagtc aaatttgatg   4800

gctcaatgca cccacaaatt catgaaacac ttgtgaaata taagcttaga gattctgtaa   4860

taatgctcag acctaacgca attcccaaag ttaattttca caactggctg acagcacgtg   4920

attacaacag gataggttgt tcactagagc tagaggatca tgtcaaaatt ccatattaca   4980

tacgaggcgt ccctgacaag ttatacggga aattgtatga tatcatttta caatatagtc   5040

caaccagttg ttacggaaga ctatcaagtg cttgtgctgg gaaggtagca tacaccttgc   5100

gaacggaccc ttgttcgcta ccacgaacaa tagcaataat caatgcatta atcacagagg   5160

agtatgctaa gagagatcat taccgtaata tgatagcaaa tccttcgtct tcgcatgcat   5220

tctcactaaa tggattagtc tccatgattg cttcaagata catgaaagac cacacaaaag   5280

agaatataga caaacttatt aaggtgcggg atcaattact cgaatttcaa ggcatgggta   5340

tgcaatttca agatccatca gaactaatgg atattggcgc cttgaacaca gttatacatc   5400

aaggaatgga tgcaacagct gcatgcattg gtctccaagg acgatggaat gcatcactca   5460

tacaacgtga tctcatgatt gcaggaggag tattcattgg aggaatatta atgatgtgga   5520

gcttgttcac taagtgggga acgaccaatg tatcacacca aggaaagaac aagcgtagtc   5580

gacaaaaatt gaagttcaag caagctagag acactaaata cgcatatgat gtgacaggat   5640

cggaagaaac tcttggtgaa aactttggaa cagcttatac aaagaagggc aaaggaaaag   5700

gaaccaaagt tggtcttggg gtgaagcagc ataaatttca catgatgtat ggttttgatc   5760

ctcaagagta caacctaatt cgttttgtcg accctcttac aggtgcaaca ttagatgaac   5820

aaatccatgc tgatattcgc ttagtgcaag aacactttga cataatccga gaagaggcag   5880

tcgcaaatga cacaattgag cgacaacaca tatacggaaa tcctggtctc caagcattct   5940

tcatacagaa tggatcagct aatgcattaa gagttgattt aacaccacat tcacctttgc   6000

gtgttatgac gaacaacaac atagcaggat ttccagaata tgagggcaca ttacgacaaa   6060

ctggcacagc tctcactgta cctgtgaatc aagtaccagc agcaaatgag acaggagttg   6120

cccacgaatc taaatctatg atggctggat tgggcgatta cacacctatt tctcagcagc   6180

tttgtctcgt acaaaacgat tctgaaggag tcaaacgtaa tgtgtatgca attggatatg   6240

gatcatattt aatatcaccg gcgcatcttt tcaagtataa caatggtgaa atcacaatta   6300

aatcctcaag agggctgtat aaaattagaa attcagtcga tgtcaaattg cacccgattg   6360

cacagaggga catggtcata attcaacttc caaaagactt cccaccgttc ccaatgcgac   6420

ttaagttctc aactccgtca agagatgtgc gtgtgtgctt agttggaatc aactttcaac   6480

agaatcatac cacgtgcata atatccgaaa gcagtgtgac agcacccaaa ggaaatggtg   6540

atttctggaa acattggatt tcaactgttg atgggcaatg tgggctaccg ttagttgacg   6600

ttaagaataa acacattgtc ggaattcaca gcctggcctc aacaagtgga aatacgaact   6660

ttttcgttgc aatgcctgag aacttcaatg aatatatatc taatctcgtg caaacaaata   6720

agtgggaaaa gggatggcat tacaacccaa atcttatttc atggtgtggt ctaaacttag   6780

ttgattcagc acctaaagga ttgtttaaaa catcaaaact tgttgaagat ttggatatga   6840

gcgttgaaga acaatgcaag gtgacagaga catggttgac ggaacacatc caggataatc   6900

tacaggtcgt tgcaaaatgt ccaggccaac ttgtaacaaa gcatgtcgtt aaaggcccat   6960

gtccgcattt tcaactgtat ttatccacac atgatgaggc aaagttgtac ttttcacctt   7020

tgcttggaaa gtatgacaag agtaggttga acagggcagc atttatcaaa gatctttcaa   7080

agtacgcaaa gccgatttat attggagaga tcaattatga aatctttgat aaggcagttg   7140

atcgagtcat aagcatcctc agaagtgtag gaatgctaca gtgtacatac gtgacggacg   7200

aagaagaaat tttcaattcg ttaaatatga acgcagccgt aggtgcactc tacacaggaa   7260

agaagaaaga ctatttcaaa gatttctcga acgatgacaa agccgaaatc attatgcgtt   7320

catgtgagcg catctacaat ggacaattgg gtgtatggaa cggttcactc aaagctgaaa   7380

tacgaccaat agagaaaacc atactaaaca agacacgcac tttcacagca gcgccattag   7440

aaactctact tggtgggaaa gtatgtgtgg acgatttcaa caatcaattt tactcacatc   7500

atcttgaagg cccatggacc gtaggaatta caaagtttta tggaggatgg aaccgacttt   7560

tggagaaatt gccagaagga tggatttatt gcgatgcaga tggatcccag tttgacagct   7620

cgctaactcc atatcttatt aatgctgtat tgcacattcg cttacatttc atggaagaat   7680

gggagttggg agctcagatg ttgcgaaatt tatacacaga gattgtttat acgccaatcg   7740

caacgcctga tgggtctgtc atcaagaaat tcaaaggaaa taatagtggg caaccatcta   7800

cagtcgttga caacacgctt atggttatta tagcatttaa ttacgcaatg ttatcaagtg   7860

gcattcctga agacaaaatt gacgactgct gtagaatgtt tgcaaacggt gacgacttac   7920

tcttggcagt gcatccggat tatgaatata tattggacgg atttcaaaat cattttggaa   7980

accttggcct taactttgag ttcacatcga ggacaaagga caaatcagag ttatggttca   8040

tgtcaacaca aggagtcaag tgtgaaggta tctacatacc aaaactcgaa agagaaagaa   8100

tagtcgcaat ccttgaatgg gaccgatcga acttgcctga gcatcgtctt gaagctatct   8160

gtgcagccat ggttgaagca tggggttacc cagatttagt tcatgaaatt cgaaagtttt   8220

acgcgtggct tcttgaaatg caacccttcg cgaacctggc aaaggaaggc atggcgccat   8280

acatagcaga aacagcactc cgcaacctct accttggaac aggcatcaaa gaagaagaaa   8340

ttgaaaaata ttttagacag tttgtcaagg atcttcctgg atacgtagaa gattacaatg   8400

aagaagttat tcatcaatct ggtcaagttg acgcagggag acagggcggt agcggcgctc   8460

aaggaggcac gccaccagca ggaagtggag gcactggatc tggcactcaa ggcaatgggg   8520

gtcagacggg atcccaagga agtagtggtc aacaagggtc cggtgggggc actggtcaag   8580

gagcagctgg aaacaacggc ggaggtcaga caggaggctc tagtgggaca gctggtcaga   8640

gagataagga cgttgacgca ggctcggctg gaaagatatc cgtaccaagc ttaaagccat   8700

gtcaaagaaa atgcgcttgc caaaggcaaa aggaaaagag tcttgcattt ggactttttg   8760

131

ttgacataca agccacaaca gcaggacata tcgaacacaa gagcaaccaa ggaagagttt   8820

gatagatggt acgatgccat aaagaaagag tacgagattg atgacacaca aatgacagtt   8880

gtcatgagcg gtcttatggt ctggtgcatc ga   8912

Sequence Listings 4. 4:Bomet SCMV isolate

tttttttttt ttttttttgt ctcttaccag gagactcgca acacaagtgg caaagccaca   60

cccagtgcag tccatctcat ggctcgacgc tagtacacgc tacaccagaa cacactgggt   120

ccaaccctgt atgctgcaga ctggttcacg ctttgaatac atacgtaata aaatagtaaa   180

gtcgaggtaa aacctcactg agatactata tattagtata tattataggt actgcaaaca   240

gggtttccag gagactagtg gtgctgctgc actcccaaca gagagtgcat attgcgacta   300

acgtcgccag ctgtgtgtct ctctgtattc tcctgggtct cgccgacatt tccgtccaga   360

ccgaacagtc gtgtgtttga accacgaact gctgcggctt tcatctgcat gtgggcttcc   420

ttagctctcg caggtgtgcg tgaagtcatt tcatagaaat caaatgcata ccgtgctaag   480

ctatagtcgg tgagattgcg ctgaagtccg tatcttggca tgtatcgctc agtagagttt   540

ctgtactcta tgtacgcttc agctgcatca ctaaaatgat gcataatttg tcggaaagtt   600

ggagatgcat tttcaataac tggtttgagt ggaaaaactc tttgttcatc tccatccatc   660

attgtccaat ttccgtttat gtttggtgag caaccatttt cgatgcacca gaccatcaga   720

ccactcatga cgactgtcat ttgtgtatca tcgatctcgt actccttctt tatggcgtcg   780

taccatctat cgaactcttc cttagttgct cttgtgttcg atatgtcctg ctgttgtggc   840

ttgtatgtca acaaaaagtc caaatgcaag acatcttttc cttttgcctt tggcaagcgc   900

attttctttg acatggcttt aagctttggt acggatatct ttccagccga gcctgcgtca   960

acgtccttat ctctctgacc agctgtccca ctagagcctc ctgtctgacc tccgccgttg   1020

tttccagctg ctccttgacc agtgcccca ccggaccctt gttgaccact acttccttgg   1080

gatcccgtct gacccccatt gccttgagtg ccagatccag tgcctccact tcctgctggt   1140

ggcgtgcctc cttgagcgcc gctaccgccc tgtctccctg cgtcaacttg accagattga   1200

tgaataactt cttcattgta atcttctatg tatccaggga gatccttaac aaattgctta   1260

aaatattttt caatttcttc ttccttaatg cctgttccaa gataaaggtt acggagcgct   1320

gtttctgcta tgtatggcgc catgccttct tttgctagat ttgcgaaagg ttgcatttct   1380

aaaagccacg cataaaattt ccgaatttca tgaacaagat ctgagtaacc ccatgcttca   1440

accatggctg cacaaatggc ttccaatcta tgctcaggta aatttgatcg atcccattcg   1500

agtattgcga ctattctttc tttctcgagt ttaggtatgt agattccttc acatttgatt   1560

cctcttgtag acatgaacca cagttcggat ttatctcgtg ttcgtgatgt aaactcaaaa   1620

tttaatccga gatttccgaa gtgatcctga aatccgtcta aaatgtattc aaaatcagga   1680

tgtactgcta ggagcaggtc atcaccattt gcaaacatcc tacagcagtt atcaatctct   1740

tcttctctaa tgccgcttga tagcatggca tagttgaaag ctattataac catcaatgta   1800

ttgtcaacta ctgtagaagg ttgtccacta ttgtttcctt tgaatttctt cacgatagat   1860

ccatctggtg ttgcaattgg tgtataaaca atctcagtat aaaggttttt aagcatttgc   1920

gctcctatgt tccaatcttc catgaactgt aatcgaatat tcaacactgc attaatgaga   1980

tatggtgtta atgaactatc aaattgggat ccgtcagcgt cgcaataaac ccatccttcc   2040

ggcaacttct caagtaaacg attccaacct ccatagaatt ttgttatccc aacagtccat   2100

gggccttcta aatggtgtga ataaaattga ttattaaaat cgtccacaca cacttttcct   2160

ccaagcaaag tttccaatgg agctgctgta aaggttcgag tcttatttaa catggttttc   2220

tctattggtc tgatttcggc tttaagtgac ccattccaca cgccaagttg tccattgtaa   2280

atgcgttcac acgatctcat tacgatttct tccttgtctt catttgaaaa actttcaaag   2340

taatctttct tctttcctgt gtacaatgct ccgacagctg cgtttaggtt gagtgattta   2400

aaaatttctt cttcgtctgt aacataaacg cattgttgca ttccaacatt tctaagaata   2460

ttaatgactc gttgtacagc tctatcaaag atatcataat tgatttctcc aatataaatt   2520

ggttttgcat actttgatat gtctttgata aatgccgctc tattgagtct gctcttatca   2580

tattttccaa gcagaggtgc gaaatattct ttggcatcat tatgtgttga caaatataac   2640

tggaagtgcg ggcattgacc tttgacgaca tgcttagtga cgagttggcc tggacatttc   2700

gcgaccactt gcaaattatc ttgtaattgc tctgtgagcc atgtttcagt aactttgcac   2760

tgctcctcaa cactagtatc caaatcttca actagttttg acgttttaaa caaacccttt   2820

ggagcagagt caactaagtt taatccacac caagatatga gattcggatt atagtgccat   2880

cccttttccc atttgtttgt ttgcacaagt ccattaatgt attcattaaa gttctcaggc   2940

atggcaacaa agaaattggt gttcccactt gtagatgcaa gactatgaat ccaacaatg   3000

tgtttgttct tagtatctac caatggtagt ccacattgac catcgactgt tgatatccaa   3060

tgcttccaaa aatctccatt cccctttggt gctgttacgc tgctttctga tacgatacaa   3120

gtgctatagt tctgttggaa attgactcct actaagcaaa ctcgcatctc tcgtgatggt   3180

tgtgtgaact tcaaacgcat tgggaacggc gggaaatcct ttgggagttg gattatgacc   3240

atgtctctgt gtgcaatcgg atgtaatttt aaatccacag agttgcgcat tttatacaat   3300

cctcttgatg atctaattgt aatttcgcca ttgttatatt taaataagtg cgctggtgaa   3360

atgagatacg agccatatcc aattgaaaat acattccgtt ttactccatc tgagtcattt   3420

133

tgcactaaac acaattgttg cgatattggc gtgtaatcac ccagtccatt catcatagat   3480

tttgattcgt gtgcaactcc tgcttcattc gcgatcggga cttgaccaat gggtatggtt   3540

atggctgttc cagtctgacg aagtgtgcct tcatattctg ggaatcctgc tatattatta   3600

cctgtgacaa ctcgtgtagg tgaatgtggc gttaaatcaa cccttagggc gtttgctgac   3660

ccattttgta tgaagaatgc ttgtagtcca ggattgccgt agatctgctg tctttcaatt   3720

gtgtcgttag ctactgcctc ctcacgaatt tcagcgaaat gctcttgaat taagcgtatg   3780

tcggcatgaa tttgctcatc aagagttgct cctgtgagtg gatcgacaaa tcgaattagg   3840

ttgtactctt gaggatcaaa accgtacatc atatgaaatt tgtgttgttt cacaccgagt   3900

ccaactttag ttccttttcc tttacctttc tttgtatagg ctgtcccgaa attctcgcca   3960

aggcactctt ctgatcctgt gatatcatat gcatatttgt tatctcttgc ttctttgaat   4020

cggagttttt gtctactacg tttgttcttt ccttgatgtg agacatttgt gttgctccat   4080

ttagtgaata agctccacat catcaaaatg cctccgataa aaactcctcc tgcgattaag   4140

aggtcacgtt gtatgagcga agcattccat cgtccttgta atccgataca agctgcagtt   4200

gcatccatcc cttggtgaat gactgtattg agagccccaa tttccatgag ttctgacgga   4260

tcttgaaatt gcattccggt gccttgaaac tcaagtaatt gatcacgtac tctgatgagt   4320

ttgtcgatat tctcttttgt atgatccttc atatatctag ttgcgatcat agataccaac   4380

ccattgagtg agaatgcatg tgatgaagaa gggtttgaaa tcatgtttcg gtagtgatct   4440

ctcttcgcat attcctccgt gattaaagca ttgattattg ctattgttct tggaagtgaa   4500

cacggatctg ttcgtaaagt gtatgctact ttacctgcac acgcacttga tagtctacca   4560

taacaactag ttggactata ctgtaaaata atatcatata gctttccata caacttatcg   4620

ggaactcccc taatatagta cggaatttta acatgatctt caagttctaa tgagcaacct   4680

attctattat aatcccgggc tgtaagccaa ttgtgtaaat taacttttgg aattgcattc   4740

ggcctaagca ttatgacaga gtctctaagc ttgtatttaa ctaaagcctc atgtatttgt   4800

ggatgcattg aaccatcaaa cttaacgagc tcagccatga caaatggtga taattcaaat   4860

tgcatcatgg ttcttgcttg tttaactgtg cactttgcaa gatgggtcgt tgaaacatta   4920

tgggtgatga ctttgaggcc atatgcgaaa cacatgaaag ctgcttctgt tgcgatcatt   4980

gctggaatct cttgcaaacc tttcatggtt tttccaatac gaataactgt gccaggttta   5040

tttcttccaa cacgacccag tcgttgtatg cgttcaccat atgatatact cacgcgttta   5100

tacaatattg ccctgttgtc aacatccaaa tcagctgata ctttcagtcc aaagtcaaca   5160

acaacatcga tatctagtgt gacgccattt tcaattatat ttgttgcgac aacaaaacac   5220

ttcttttgtg aggtaccgtt tgtggtgatt cctgttgtgt tttgcttcat tgttcggcca   5280

tccactttaa taacggaata tttgagttca actaatgcat gcgacaatga atcgacgtca   5340

ttataacttg ctacataaac taagatatta tttccgtact tcgtagcatc tgcggttgaa   5400

ccggttccga gttccataac aaattgttgc tgagttaggt tttcacaaac atggatgtct   5460

actggatgtt gtgttgtgaa ttcgcactcc cttcctggtg gcgtggctga taccttgatg   5520

atttttcctc ggtattcgta ttctttgagt aagcaataaa aggccattgc tggtgcttcc   5580

attatatgac attcatcaaa gattataaaa tcgtattcag atattttatc tggattattt   5640

gcgtacatgt gcaatgcaaa acctgatgtc atgatcgtga ttggagtaca cccgaaagaa   5700

ctcagtccac gcatttgaag agttggactt acattaaatg gtggtccttg taattgccta   5760

cacacgttct cggctagtgg ccttgtaggc tctagtaaaa gcacttttcc acgcatgcta   5820

agatggtatg gtaacccagt agattttcct gagccgactg ctcctctaag taagaactct   5880

ttctcaatgt ttgagtgtgc aagttcgatg ctcgtgtggg ctgcgttttc tcgtgtgaat   5940

tcaacaaact ttcccccaag tcggtagtgt ggcactgttc tattgttgct taattgattt   6000

gaccaccaat cttcaaatgt cacatcatta ctgaatgtat cggcaggtaa atcttgattc   6060

gtgtcaaagt ctattgttaa atttttctca gtgctaagtg tcaatggctc atcaacatct   6120

acactctgtt gtgtcactgg aggtgtcaag atgttggaaa agtttactga agggaatggg   6180

acgttttgtt caattgtact aattactgtc cgcattttgt tcaacacctt atagacagca   6240

tcgcttttct gtggatcaag tatcatggtt atcaaggtgc caactgccat tgcttgttcc   6300

aaaattggttt cgagattgga ttttccttca tgtatgacac ctgtgtgggt taagtccagt   6360

gttgcttcaa ctaatcgcgg atggttttgg cgcatatatt caacaaattg ttcacatgtc   6420

aaaattgtcct tccgctcttc catcaaggta gaatggattt tgcgaacttc gttaatttct   6480

gcttcatatt cgtcttctct ttgttgtttc ttcaatttct tgtagtcatt catagttact   6540

atgatggtgt tcgctatggt agttaataaa ctcaaaacta tgaatatatg catgagtcta   6600

aatatatcgg gtataaacca ataaaccgtt tttatagtgc gtattctagc tttctctaat   6660

ttatcataga aaccgttgcg gagtttggtt aaaatagagc tgacttgatc ttgacttttc   6720

cgcactaaat ctgatattag atgcgtagct gatatattgt atacagcgcc taaatctacg   6780

ctttttctca cggttaaaga cggtttgtaa tacttcctga ctttgaacac atgccatatt   6840

gctgaaaatt ttccacagta tgataattct gcccatgact ggtttaattg atctacgtaa   6900

ctttttttcca tcagggtata tagacgctcg tcgtatagtg catatccctc agctatcaac   6960

tccttgttca tttctgaccg tgttgtcatt gcttcgaggt gtgaccacaa aagtcgcttt   7020

gctgggtcga catggctaag tcctgaaact gctaatctta gttgactaga tgctttttca   7080

agtatctgca tttgtagaac tagtaattca gcctgggatg ttttcttggc taatgcttcg   7140

agttgtgcga atatggctgc aactccttga ttcttcacga tccagtatgt catagcttgc   7200

tcaatgtaac aattattata cagtgctatc aataccgttg gagatgcgat tgccatcata   7260

135

agcaagaagg gttcctcctc aattatctgc ttaatacgct ctggtttgaa catgtttttc   7320

gtgagcatct taagaagtgt attgaatgtc tgttgcgtaa gagttcctcc tactatgtat   7380

tcgcgcattt cgctatccat agattcatat tgaaatttaa ttaattgacc aatcgtgctt   7440

gctttcaata tgtgaaatcc aacgcttagt gaaccatatg agtcgatgac gtgcatagat   7500

ttattttcat ggtcaactaa gattggaggt agctcagcat cttaatttc tggaaacatt   7560

acagataatg cataacacgc tgttgctacg tctttaagcc ttggccactt tccaagacgc   7620

tcaactagtt catctctcaa gaatttcgta taatccttgg cagattcttc attgacgttg   7680

atcatggctg ctagaaagat attgatataa caatacccgt ccttagcaat gtacatgctg   7740

ggtggtgttg tgtttggtag gtccacaatc tttggatcaa ttgaattgcc tattgttatg   7800

tggccttttg ttggaggtat tatctcagag tatgctggtt taccaaattc tgtcgtgacg   7860

cagcaacacg ggtatatata tcctccttcg atcttactaa cacaacattt gtctagcggt   7920

tgtctgttga tcggaagtcc aacaaatgat tctctaatct tttgaaagtc tagtggaata   7980

actaactttc ctatcgataa ctttcgttga ccgttagggt tgaaccgtgt aacgtgcttt   8040

tcatatgcgt ctgtgggatc cactgcttca aaatagtttg caaagaatcg tttggcgtga   8100

tattccctat taccccatac aaaatttgca tttttatcca attgattgtc acacattaaa   8160

gctgcattta ttgtactctt tggtgatatt ttatttctaa aagattccaa tgagtcagtt   8220

ttgagtgatt ccttacgatt cttataccac cgcgacactt cgagtaatgc tgcgcttgct   8280

ttcatagcat catcaggagt taaagtatta actttgatga gcgcttcgct gatatccatt   8340

atttgctttg actgattatc agtatagtgt cctattgttt ctgcaacttg ccacaacatt   8400

ggtaaatgtg agaattttgg gttgcatctt tctttgataa attgtatcat tctggataac   8460

ttttgatcct tggcaaggta ctctcgttgc ttctcttcga tgcgttgcaa atttttgtag   8520

agtttagctc caaattcatc atctgataat tcaagatcat ccatgttaca tgttttacat   8580

gtaatcttga atgtggagtg gaataaatatc tcaagtagag ccattcgttt tccacactct   8640

tctagattga ttgtcggagt gtgctctgta tgggtggtag aaatactacg attatccacg   8700

taagcatttg tatatcctct ccaaaagtca ttagcttggg gatctgcgta gtgttcaata   8760

tccataaccc ggtctgtttc acataggcta ttgactagtt caccattcat tctccctcta   8820

acaacgaaca gttcatcatt ttgaatgaat gtgagaccac tgctgcctct ataaaggctc   8880

tgcaattcct cttggttcat tgtactatag caacgagcaa tggcttcgat tgttggaagc   8940

catttgtgtc cgatggctat gtccctcctc ttaaacaaat tgttttcatg gtttgttttg   9000

caatggagaa agtccttgcc attgaacctt ttgattgata actgtgtttt agctctttta   9060

cgcttatcaa ttagttcaac tttaaggccg ctaaccttcg agatctctaa aatctctctg   9120

gttagcttgc ttacagatgc acttttgatg ctcgtatcca ctttgttgat aggtctttca   9180

aatcgtgaat atactaagtg atcaacagga ttgtgatcac atttctcaac ccacatttgt   9240

tgagttgcct gcttctgttt cggttttgca ggctcaacgc tcacaacact tgtcttcctc   9300

aataatttca tattgtgagc aagtcgctgt tcttctgagt aatgttgatg ttttgctgca   9360

aacatttcca ttactctctt aatgtgacga tcattgttga catctggctg ccacttgtat   9420

gtcacatggg tccacgttcc cgccatttct ttcgaaagaa gttgagagat gaggagaaat   9480

cagagagtga tctcgaagaa ccgtttcgtt cgagccgttc actacaatct gagcgcaagt   9540

aaattggttt tgcttggttg tgttttgttg tgttgtgttg agttttgtgt tttt   9594

Sequence Listings 4. 5:Kericho SCMV isolates

tgcactatac gacgagcgtc tatatactct gatggaaaaa agttacgtag atcaattaaa   60

ccagtcatgg gcagaattat catactgtgg aaaattttca gcaatatggc gtgtgttcaa   120

agtcaggaaa tattacaagc cgtctttaac cgtgagaaaa agcgtagatt taggcgctgt   180

gtacaatata tcagctacgc atctaatatc agatttagcg cagagaagtc gcgatcgagc   240

cagctctatt ttaaccaaac tccgcaacgg tttttatgat aagttagaga aggctagaac   300

tcgagcaatt aaaactgttt attggttcat acctgacata tttagactta tgcatatttt   360

catagttctt agtttattaa caactgtggc aaacactatt attgtgacta tgaatgatta   420

caaaaagctg aaaaagcaac aaagagagga tgagtatgaa gctgaaatta acgaggttcg   480

aaaaattcac gctaatctga tgaaagagca taatgacaat ctaacatgtg aacagtttat   540

cgaacatatg cgccagacgc atccacgatt aattgaagcc acattggagt tgacacacac   600

aggcgtcatt catgaaggaa agagcaacct agaaaccaat ctcgagcaag ctatggcagt   660

gggcacttta ctcacaatga tgttagaccc acagaagagt gatgcagtat acaaagtgtt   720

gaataagatg agaactgtaa ttagcacatt tgagcagaat gtcccatttc cttcaatcaa   780

ctttacaaat attttaacac catcagtggc gcaacaaagt gtagatgttg atgaacctct   840

aactttaagt actgacaaga atctaacaat agattttgat actaatcaag atttaccagc   900

ggacacgttt agcaatgatg ttactttcga agattggtgg gcaaatcaaa taagcaacaa   960

tagaacagtt ccacattatc ggcttggtgg caaatttgta gaattcacaa gagaaaatgc   1020

agcacatgtt agcattgagc tcgcacactc aaatcttgag aaggagtttc tactcagagg   1080

agctgttggc tcaggaaaat ccactggtct tccatatcat cttagtatgc gtggtaaggt   1140

gcttttaata gaaccaacaa gaccactagc cgagaatgtc tgcagacagc ttcaaggacc   1200

accattcaat gtcagcccga cactacaaat gcgtggttta agctcttttg gttgcacgcc   1260

aattacgata atgacatcag gtttcgcact gcacatgtat gcaaataatc cggataagat   1320

ctctgattac gattttataa tatttgatga gtgtcatatc atggaagcgc cagccatggc   1380

gttctactgt ttgttgaaag aatatgaata tagaggcaaa atcattaagg tgtccgcaac   1440

accaccagga cgagagtgtg agttctctac tcagcatcca gtcgatatcc atgtgtgcga   1500

aaatctcact cagcaacaat tcgttagaga gcttggttcc ggatcaaacg ttgacgcaac   1560

aaagtatgga aacaacatac ttgtgtatgt tgcaagctac aacgacgtcg actcattagc   1620

gcatgcttta actgaattac actattcagt cataaaggtt gatgggagaa caatgaaaca   1680

gaacactaca gggatagtga caaatggcac ctctcaaaag aagtgttttg tcgttgctac   1740

gaatatcatt gagaatggtg tcacattaga cgttgatgtt gtggtcgact tcggactcaa   1800

agtaacagct gaattagatg ttgacaacag agctatttta tacaagcgtg ttagcatttc   1860

atatggtgag cgaattcaaa ggctagggcg tgttggaaga aacaaacccg gaacagtcgt   1920

tcgaattgga aaaaccatga agggtttgca agaaatccca gccatgatag cgacagaagc   1980

agcattcatg tgctttgcat atggactaaa ggtcataaca cacaatgttt caacaacaca   2040

cttagcaaag tgcacagtca agcaagcgcg aacaatgatg caatttgagc tttcaccatt   2100

tgtaatggca gaactagtca aatttgatgg ctcaatgcac ccacaaatcc atgaagcact   2160

tgtgaaatat aagcttagag attctgtaat aatgctcaga cctaacgcaa ttcccaaagt   2220

taattttcac aactggctga cagcacgtga ttacaacagg ataggttgtt cactagagct   2280

agaggatcat gtcaaaattc catattacat acgaggcgtc cctgacaagt tatacgggaa   2340

attgtatgat atcattttac aatatagtcc aaccagttgt tacggaagac tatcaagtgc   2400

ttgtgctggg aaggtagcat acaccttgcg aacggaccct tgttcgctac cacgaacaat   2460

agcaataatc aatgcattaa tcacagagga gtatgctaag agagatcatt accgtaatat   2520

gatagcaaat ccttcgtctt cgcatgcatt ctcactaaat ggattagtct ccatgattgc   2580

ttcaagatac atgaaagacc acacaaaaga gaatatagac aaacttatta aggtgcggga   2640

tcaactactc gaatttcaag gcatgggtat gcaatttcaa gatccatcag aactaatgga   2700

tattggcgcc ttgaacacag ttatacatca aggaatggat gcaacagctg catgcattgg   2760

tctccaagga cgatggaatg catcactcat acaacgtgat ctcatgattg caggaggagt   2820

attcattgga ggaatattaa tgatgtggag cttgttcact aagtggggaa cgaccaatgt   2880

atcacaccaa ggaaagaaca agcgtagccg acaaaaattg aagttcaagc aagctagaga   2940

cactaaatac gcatatgatg tgacaggatc ggaagaaact cttggtgaaa actttggaac   3000

agcttataca aagaagggca aaggaaaagg aaccaaagtt ggtcttgggg tgaagcagca   3060

taaatttcac atgatgtatg gttttgatcc tcaagagtac aacctaattc gttttgtcga   3120

ccctcttaca ggtgcaacat tagatgaaca aatccatgct gatattcgct tagtgcaaga   3180

acactttgac ataatccgag aagaggcagt cgcaaatgac acaattgagc gacaacacat   3240

atacggaaat cctggtctcc aagcattctt catacagaat ggatcagcta atgcattaag   3300

agttgattta acaccacatt cacctttgcg tgttatgacg aacaacaaca tagcaggatt   3360

tccagaatat gagggcacat tacgacaaac tggcacagct ctcactgtac ctgtgaatca   3420

agtaccagca gcaaatgaga caggagttgc ccacgaatcc aaatctatga tggctggatt   3480

gggcgattac acacctattt ctcagcagct ctgtctcgta caaaacgatt ctgaaggagt   3540

caaacgtaat gtgtatgcaa ttggatatgg atcatattta atatcaccgg cgcatctttt   3600

caagtataac aatggtgaaa tcacaattaa atcctcaaga gggctgtata aaattagaaa   3660

ttcagtcgat gtcaaattgc acccgattgc acagagggac atggtcataa ttcaacttcc   3720

aaaagacttc ccaccgttcc caatgcgact taagttctca actccgtcaa gagatgtgcg   3780

tgtgtgctta gttggaatca actttcaaca gaatcatacc acgtgcataa tatccgaaag   3840

cagtgtgaca gcacccaaag gaaatggtga tttctggaaa cattggattt caactgttga   3900

tgggcaatgt gggctaccgt tagttgacgt taagaataaa cacattgtcg gaattcacag   3960

cctggcctca acaagtggaa atacgaactt tttcgttgca atgcctgaga acttcaatga   4020

atatatatct aatctcgtgc aaacgaataa gtgggaaaag ggatggcatt acaacccaaa   4080

tcttatttca tggtgtggtc taaacctagt tgattcagca cctaaaggat tgtttaaaac   4140

atcaaaactt gttgaagatt tggatatgag cgttgaagaa caatgcaagg tgacagagac   4200

atggttgacg gaacacatcc aggataatct acaggtcgtt gcaaaatgtc caggccaact   4260

tgtaacaaag catgtcgtta aaggcccatg tccgcatttt caactgtatt tatccacaca   4320

tgatgaggca aagttgtact tttcacctttt gcttggaaag tatgacaaga gtaggttgaa   4380

cagggcagca tttatcaaag atctttcaaa gtacgcaaag ccgatttata ttggagagat   4440

caattatgaa atctttgata aggcagttga tcgagtcata agcatcctca gaagtgtagg   4500

aatgctacag tgtacatacg tgacggacga agaagaaatt ttcaattcgt taaatatgaa   4560

cgcagccgta ggtgcactct acacaggaaa gaagaaagac tatttcaaag atttctcgaa   4620

cgatgacaaa gccgaaatca tcatgcgttc atgtgagcgc atctacaatg gacaattggg   4680

tgtatggaac ggttcactca aagctgaaat acgaccaata gagaaaacca tactaaacaa   4740

gacacgcact ttcacagcag cgccattaga aactctactt ggtgggaaag tatgtgtgga   4800

cgatttcaac aatcaatttt attcacatca tcttgaaggc ccatggaccg taggaatcac   4860

aaagtttttat ggaggatgga accgactttt ggagaaattg ccagaaggat ggatttattg   4920

cgatgcagat ggatcccagt ttgacagctc gctaactcca tatcttatta atgctgtatt   4980

gcacattcgc ttacatttca tggaagaatg ggagttggga gctcagatgt tgcgaaattt   5040

atacacagag attgtttata cgccaatcgc aacgcctgat gggtctgtca tcaagaaatt   5100

caaaggaaat aatagtgggc aaccatctac agtcgttgac aacacgctta tggttattat   5160

agcatttaat tacgcaatgt tatcaagtgg cattcctgaa gacaaaattg acgactgctg   5220

tagaatgttt gcaaacggtg acgacttact cttggcagtg catccggatt acgaatatat   5280

attggacgga tttcaaaatc attttggaaa ccttggcctt aactttgagt tcacatcgag   5340

gacaaaggac aaatcagagt tatggttcat gtcaacacaa ggagtcaagt gtgaaggtat   5400

ctacatacca aaactcgaaa gggaaagaat agtcgcaatc cttgaatggg accgatcgaa   5460

cttgcctgag catcgtcttg aagctatctg tgcagccatg gttgaagcat ggggttaccc   5520

agatttagtt catgaaattc gaaagtttta cgcgtggctt cttgaaatgc aacccttcgc   5580

gaacctggca aaggaaggca tggcgccata catagcagaa acagcactcc gcaacctcta   5640

ccttggaaca ggcatcaaag aagaagaaat tgaaaaatat tttagacagt ttgtcaagga   5700

tcttcctgga tacgtagaag attacaatga agaagttatt catcaatctg gtcaagttga   5760

cgcagggaga cagggcggta gcggcgctca aggaggcacg ccaccagcag gaagtggagg   5820

cactggatct ggcactcaag gcaatggggg tcagacggga tcccaaggaa gtagtggtca   5880

acaagggtcc ggtgggggca ctggtcaagg agcagctgga aacaacggcg gaggtcagac   5940

aggaggctct agtgggacag ctggtcagag agataaggac gttgacgcag gctcggctgg   6000

aaagatatcc gtaccaaagc ttaaagccat gtcaaagaaa atgcgcttgc caaaggcaaa   6060

aggaaaagat gtcttgcatt tggacttttt gttgacatac aagccacaac agcaggacat   6120

atcgaacaca agagcaacta aggaagagtt cgatagatgg tacgacgcca taaagaagga   6180

gtacgagatc gatgatacac aaatgacagt cgtcatgagt ggtctgatgg tctggtgcat   6240

cgaaaatggt tgctcaccaa acataaacgg aaattggacg atgatggatg gggatgaaca   6300

aagagttttc ccactaaaac cagttatcga aaacgcatct ccaacttttc gacaagttat   6360

gcatcatttc agtgatgcag ctgaagcgta tatagaatac agaaattcta ctgagcgata   6420

catgccaaga tatggacttc agcgaaatct caccgactat agcttagcgc ggtatgcttt   6480

tgatttctat gaaatgactt cacgcacacc agctagagct aaggaagccc acatgcagat   6540

gaaagccgca gcagttcgtg gttcaaacac acgactgttc ggtttggacg gaaatgtcgg   6600

cgagactcag gagaatacag agagacacac agctggcgac gttagtcgca atatgcactc   6660

tctgttggga gtgcagcaac accactagtc tcctggaaac cctgtttgca gtacctataa   6720

tatatactaa tatatagtac gttggtgagg ctttgcctcg gttttactat cttattatgt   6780

atgtatttac agcgtgaacc agtctgcagc atacagggtt ggacccagtg tgttctggtg   6840

tagcgtgtac tagcgtcgag ccatgagacg gactgcactg ggtgtggcta tgccacttgt   6900

gttgcgagtt tcctggtaag agac   6924

Sequence Listings 4. 6: Borabu SCMV isolate
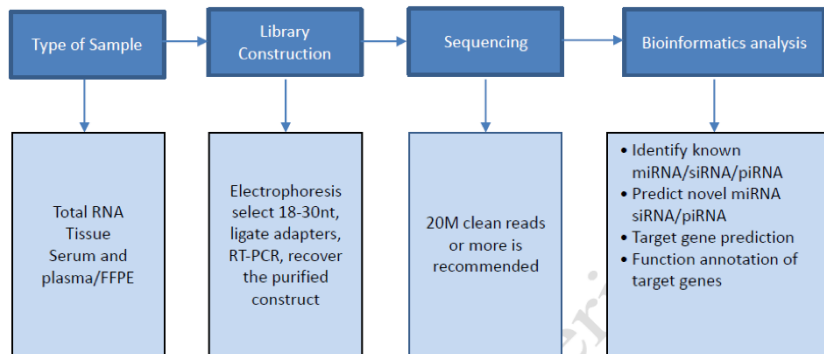
5 Research pipeline

Figure 1 Standard small RNA research pipeline

5.1 Sequencing pipeline

5.1.1 Libary construction: 18-30nt

5.1.2 Sequencing: SE50

5.1.3 The library construction process of small RNA sequencing

1) Filter Small RNA screening

   Separate RNA segment by size on PAGE gel, extract fragments between 18 and 30nt .

2) 3'Adaptor ligation

   A 5-adenylated, 3-blocked single-stranded DNA adapter is ligated to 3' of the RNA extracted in step 1).

3) Reverse transcription primer annealing

   RT primers are added to the mixture in step 2), hybridizing to the ligated 3' adapters in RNA and dissociative 3' adapters.

4) 5'Adaptor ligation

   5' adapters are linked to the 5' end of ligation product in step 3). 5' adapters are preferentially attached to the single-stranded molecules instead of the hybrid products of RT primers and 3' adapters, drastically easing the self-linking problem.

5) One strand cDNA synthesis

   A reverse strand cDNA was synthesized with RT primer (in step 3) through reverse transcription extension

6) PCR Amplification

   cDNA is amplified by high-sensitivity polymerase, and the cDNAs ligated with 3 'adapter and 5' adapter were enriched to increase the library output.

7) Library fragment selection

Appendix 4. 12:Sequencing steps: BGI sample preparation

PAGE electrophoresis is used to separate 100-120bp length PCR products, effectively removing primer dimers and other by-products.

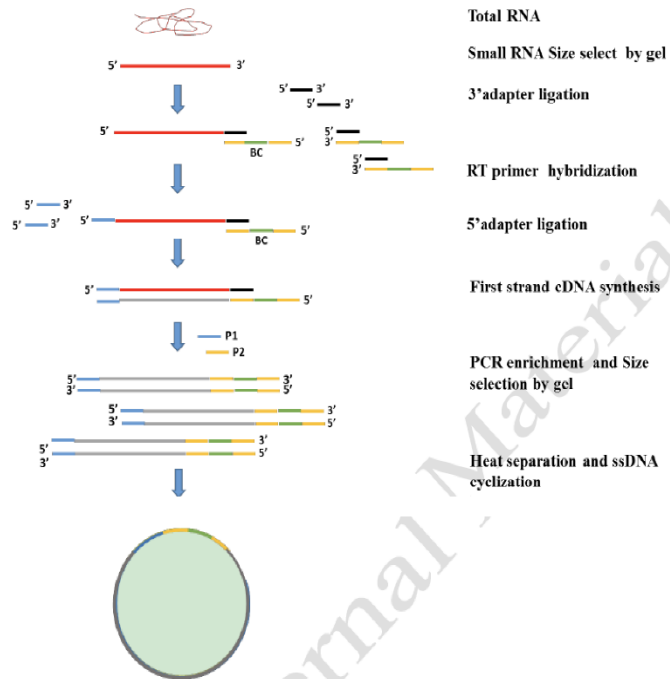8) Library quantification and pooling cyclization

Figure 2 BGISEQ-500 small RNA sequencing library construction process

Appendix 4. 13:Sequencing steps: BGI

144