UNIVERSITY OF NAIROBI, KENYA

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

CENTER FOR BIOTECHNOLOGY & BIOINFORMATICS

METAGENOMIC ANALYSIS OF GROUNDNUT (*Arachis hypogaea* L.) RNA VIRUSES IN GEM AND MATAYOS SUB-COUNTIES, WESTERN KENYA

DENNIS OBONYO OTIENO

I56/83660/2015

(**B.Sc. Hons. Agricultural Biotechnology**)

A thesis submitted to the University of Nairobi in partial fulfillment of the requirement for the degree of Master of Science in Biotechnology.

NOVEMBER, 2019

DECLARATION

I hereby declare that, I undertaken the study under supervision, and except for the specific references which have been duly and deservedly acknowledged, this is my original work and has not been submitted either partly or wholly for examination and for the award of degree in any other institution of learning.

_____

Dennis Obonyo

(Student)

I56/83660/2015

University of Nairobi

Signature: _____ Date: _____

This work has been submitted for examination with our approval to be presented to the board of postgraduate studies as per the University of Nairobi regulations.

_____

Dr. Rachel Ikawa, PhD

Centre for Biotechnology and Bioinformatics

University of Nairobi

Signature: _____ Date: _____

_____

Dr. Damaris Achieng' Odeny, PhD

Theme Leader Biotechnology (ESA) Genomics and Trait Discovery Genetic Gains

ICRISAT-ESA, Nairobi, Kenya

Signature: _____ Date: _____

_____

Prof. George Ouma, PhD

Professor of Horticulture and Crop Sciences

Institute for Climate and Adaptation

University of Nairobi

Signature: _____ Date: _____

DEDICATION

I dedicate this thesis to my dear family, friends and my mentors; the people I have drawn a lot of inspiration from over the years. But most importantly, the following individual deserves it all. My late father, James Otieno “Rateng’”, my look-alike and the embodiment of the person I have become. My friend and gracious mother, Josinter Otieno; the only lady of a humble means with a character I know. She made me to fundamentally accept we were never poor because of our temporal financial struggles. To my dearest sister Jane Anyango; my second mother, a lady with boundless inner strength and sheer determination. I am forever indebted to you all.

ACKNOWLEDGMENTS

I express my heartfelt thankfulness to the NICHE Project under the stewardship of the project manager, Professor Julius Ochuodho and his entire team for awarding me the scholarship to undertake my postgraduate studies and this research work.

I express my deepest appreciation to the International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Nairobi through Dr. Damaris Odeny for offering additional funding by linking the study with FEED THE FUTURE INITIATIVE and allowing me to use the ICRISAT-Genomics laboratory for my work. I am more thankful for the opportunity to be part of her dedicated team comprising Samuel Manthi, Teddy Amuge and Velma Okaron. Samuel was my aid from the first day at the ICRISAT laboratory. I am forever indebted to Dr. Damaris Odeny for all the insights, the commitments for being a stickler for the best and for holding my hand from the beginning to the end.

I am profoundly grateful to my able supervisors, Professor George Ouma, Dr. Rachel Ikawa and once more Dr. Damaris Odeny, who has been the cog of this team. All their valuable support, criticism, suggestions, patience and immense knowledge and skills made this research work a success. I just want to thank all for everything!

I am thankful to the members of the Department of Biotechnology, the University of Eldoret led by the Head of Department and all my colleagues for their understanding and patience throughout my study. I can never find more worthy colleagues than them.

A worthy mention goes to my friends and colleagues in the field, both in Gem and Busia. Notably, David Mola and Ibrahim Omondi, who were there with me for sample collection and chauffeuring me through the villages in Gem and Matayos. Leah Munala her team

gave me valuable links with the farmers. Not forgetting all the farmers whose warm welcome expressed much love.

I am very fortunate to have a family, which has always been there for me. I am grateful to my mother and siblings for being part of the lovely home I always missed so much while away. May all of them be blessed!

Lastly, much gratitude goes to the Center for Biotechnology & Bioinformatics (CEBIB), the University of Nairobi headed by the Director, Dr. George Obiero and the staff and colleagues at CEBIB. A special mention goes to my two wonderful friends Stephen Owino and Accadius Lunayo, with whom we have shared a lot throughout this journey.

ABSTRACT

Groundnut or peanut (*Arachis hypogaea* L.) is one of the most valuable legume grown throughout the tropical and subtropical regions in almost 100 countries on six continents. Viruses are the most important pathogens causing several diseases among cultivated legumes including groundnut. The most abundant category of viruses causing most economic losses is the RNA viruses although only a few viruses in this diverse group have been studied. Previous studies have in most cases dealt with single species but not diverse species simultaneously. These previous approaches have relied on availability of sequence information and could not detect variety of groundnut viruses. In this study, NGS for metagenomics analysis was used to detect the presence and relative abundances of RNA viruses in symptomatic and asymptomatic groundnut leaf samples collected from farmers' fields in Western Kenya. Total RNA was extracted from groundnut leaves collected from the fields and sequenced on the Illumina HiSeq 2500 platform. Partial genomes of GRV, satRNA, PCSV, CPPV1 and CPPV2 were detected. Co-infections were established between two sets, GRV- satRNA and CPPV1 - CPPV2. This is the first report of CPPV 1 and CPPV 2 viruses outside Burkina Faso and on groundnut as a host. Overall, PCSV was the most prevalent virus detected at 50% of all samples. Gem Sub-county in Siaya reported higher incidence and severity than Matayos Sub-county but with similar viral distribution. In conclusion, the results establish a universal platform for simultaneous detection of several groundnut RNA viruses. The use of viral metagenomics diagnostic procedures offers capability of generic use in groundnut seed system development, exchange and breeding programmes.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS, ACRONYMS AND SYMBOLS

BAM: Binary Alignment Map file

cDNA: Complementary Deoxyribonucleic Acid

CEFA: Committee for European on Training and Agriculture

CGMMV: Cucumber Green Mottle Mosaic Virus

CMV: Cucumber Mosaic Virus

CPPV 1: Cowpea Polerovirus 1

CPPV 2: Cowpea Polerovirus 2

dsRNA: Double-stranded RNA

ELISA: Enzyme-Linked Immunoabsorbent Assay

FAO: Food and Agriculture Organization

GBNV: Groundnut Necrosis Virus

GPS: Global Positioning System

GRD: Groundnut Rosette Disease

GRSV: Groundnut Ringspot Virus

ICRAF: World Agroforestry Center

INSDC: The International Nucleotide Database Collaboration

IPCV: Indian Peanut Clump Virus

MEGA: Molecular Evolutionary Genetics Analysis

MOA: Ministry of Agriculture

nt: Nucleotides

NGS: Next Generation Sequencing

PCSV: Pepper Chlorotic Spot Virus

PCV: Peanut Clump Virus

PDI: Percentage of Disease Incidence

PDS: Percentage of Disease Severity

PeMoV: Peanut Mottle Virus

PRSV-W: Papaya Ringspot Virus-W

PStV: Peanut Stripe Virus

qPCR: Quantitative Polymerase Chain Reaction

rDNA: Ribosomal Deoxyribonucleic Acid

RNA: Ribonucleic Acid

RT- PCR: Real-time Polymerase Chain Reaction

SAM: Sequence Alignment Maps file

SatRNA: Groundnut rosette satellite RNA

SqMV: Squash Mosaic Virus

SSA: Sub-Saharan Africa

SSP: Single Superphosphate

ssRNA: Single-stranded Ribonucleic Acid

TAS- ELISA: Triple Antibody Sandwich Enzyme-Linked Immunosorbent Assay

TSP: Triple Super Phosphate

TSWV: Tomato Spotted Wilt Virus

TSV: Tobacco Strike Virus

WMV-2: Watermelon Mosaic Virus-2

ZYMV: Zucchini Yellow Mosaic Virus

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background to the study

Groundnut or peanut (*Arachis hypogaea* L.) is an allotetraploid (AABB; 2n = 4x = 40) (Moretzsohn et al., 2013; Nigam et al., 2013) of the sub-tribe *Stylosanthinae*, tribe *Aeschynomenea*, and Leguminosae family. The cultivated groundnut has its center of origin in Southern Bolivia/Northwest region of South America and is adapted to fairly dry tropics (Committee for European on Training and Agriculture , 2011; Wangai et al., 2001).

Groundnut is the 6th most valuable vegetable oilseed crop and 13th most important global crop, which is grown throughout the tropics and subtropical regions in almost 100 countries on six continents between latitudes 40^{o}N and 40^{o}S (Naidu et al., 1999; FAO, 2003; Okello et al., 2010; CEFA, 2011). It is cultivated on 26.4 million hectares in the world over with a total production of 37.1 million metric tonnes and an average production of 1.4 metric tonnes per hectare.

Developing countries constitute 97% of the global area and 94% of the global production of groundnut, which is mainly in Africa and Asia (Ntare et al., 2008). Groundnut is a key source of dietary nutrients to many persons in developing countries with a substantial amount of proteins (23-34%), oil (44-56%), dietary fiber, minerals and vitamins (Naidu et al., 1999; Settaluri et al., 2012). In Western Kenya, groundnut is both a principal source of protein and a major source of cash income to smallholder growers who are preponderantly women (Naidu et al., 1999; Mukoye et al., 2015; Roossinck, 2015). Groundnut is largely self-pollinated but can also be cross-pollinated in areas where pollinators' activity is available (Lim and Gumpil, 1984).

However, groundnut yield is continuously on the decline with farmers obtaining less than 50% of the potential output in Western Kenya (Koech et al., 2007; Mukoye et al., 2015). This drop in yield is attributed to both abiotic and biotic factors such as fungi, bacteria, nematodes, and viruses. Moreover, viruses are responsible for 47% of all emergent phytopathogens cases of infectious diseases reported (Anderson et al., 2004). This makes viruses the most economically important biotic constraint. Plant viruses and viroids have been shown to cause huge losses in many different crops in terms of quantity and quality of produce (Mukoye et al., 2015; Sastry and Zitter, 2014). Viruses are ubiquitous, abundant and exist in great diversity as biological entities on earth. Their relationship with plants can be symbiotic, mutualistic or pathogenic, hence impacting both negatively and positively (Roossinck, 2015b). Riboviruses (RNA viruses) are of special concern because of their extreme adaptive nature to diverse environments resulting from their high mutation rates and formation of quasi-species (Krausslich, 2009; Roossinck, 2012). All the economically important groundnut viruses are ssRNA viruses, half of them are transmitted by the same insect vector, aphids (Sreenivasulu et al., 2008).

Groundnut rosette disease complex (GRD) is endemic and the most devastating disease of groundnut in Africa, with up to 100% production loss (CEFA, 2011; Naidu et al., 1999). A few studies have been conducted on other groundnut viruses in Western Kenya such as Cucumber mosaic virus (Mukoye et al. 2015), the second most economically important groundnut virus and Peanut mottle potyvirus, first identified in the USA. No study in Kenya has been reported of tackling all the groundnut viruses at once and thus the status of most of these viruses is not known. Appropriate and accurate viral detection methods are a fundamental aspect of viral exclusion efforts, and sensitive, reliable assays and efficient use of resources would be an ideal strategic pipeline (Mollov and Malapi-Wight, 2016).

The groundnut viruses detection methods used currently include serological assays, virus-sensitive indicators, bioassays and nucleic acid detection assays (i.e., nucleic acid hybridization and polymerase chain reaction) (Sreenivasulu et al., 1991; Sreenivasulu et al., 2008).Though these techniques are excellent at checking for known viruses, they have the restrained capability for characterizing new emerging viruses. This is because the assays are based on known sequences or reference information (CEFA, 2011).

A few years ago, high throughput next-generation sequencing (NGS) technologies were adopted for virus detection, i.e. viral metagenomics, which entails the purification and sequencing of an uncultured environmental sample of viruses. It provides an avenue for identifying the genome composition of a sample and thus unearthing different categories of viruses (Roossinck, 2015; Soueidan et al., 2015). Up to 50 new viruses have been identified through viral metagenomics, 36 of which were classified in new families while 14 were matched in the respective genera (Wu et al., 2015). The characterization of new viruses is of economic importance (Kreuze et al., 2009; Gutiérrez et al., 2016; Wu et al., 2015) has also been made possible through viral metagenomics.

Up to 31 viruses found in 14 genera have been reported to naturally infect groundnut world over. Nineteen of these were first extracted from groundnut; while the remaining isolated first from other hosts but are common in groundnut. Eight (8) viruses, namely, Cucumber mosaic virus (CMV), Groundnut bud necrosis virus (GBNV), Groundnut rosette viruses (GRD), Peanut clump virus (PCV), Peanut mottle virus (PeMoV), Peanut stripe virus (PStV), Tobacco streak virus (TSV), Indian peanut clump virus (IPCV) and Tomato spotted wilt virus (TSWV), are of most economic importance to groundnuts and lead to great yield losses both regionally and globally (Sreenivasulu et al., 2008).

Mixed infections with two or more viruses has been reported in groundnuts, for example, Peanut stripe virus and Peanut mosaic virus (Kuhn et al., 1973), or that of groundnut rosette virus and leaf spot disease (Okello et al., 2014). While infection with some viruses gives clear symptoms, many viral infections are asymptomatic. This has led to confusion in disease identification (Thottappilly and Rossel, 1997; Liu et al., 2011; Okello et al., 2014). Different studies focusing on high throughput sequencing analysis of RNA only or both DNA and RNA viruses have identified new viruses besides known ones that could be assigned to different families while others distantly related to known genus (Al Rwahnih et al., 2009; Kreuze et al., 2009; Gutiérrez et al., 2016). A detailed description of all the RNA viruses presently infecting groundnut in Western Kenya is lacking. This study reports a metagenomic analysis of *Arachis hypogaea* RNA virome using RNA sequence-based viral metagenomics. The specific objectives of this research were: (i) to assess the relative abundance and distribution/incidence of groundnut (*Arachis hypogaea* L.) viruses in Gem and Matayos Sub-counties Western, Kenya and, (ii) to determine the diversity of groundnut RNA viruses in Gem and Matayos Sub-counties in Western, Kenya.

1.2 Problem statement

Groundnut is a principal source of protein and a source ofsmall holder farmers' income in Western Kenya. However, groundnut production has been on the decline with farmers obtaining less than 50% of their potential output. This fall in production has been attributed to a number of abiotic and biotic factors, of which viruses are the most significant (Mukoye et al., 2015). Viruses are highly diverse and have exploited nearly all possible host genomes since they have evolved over the course of millennia with the host-plants (Okello et al., 2010). Riboviruses (RNA viruses) are of particular importance because of their extreme adaptive nature to diverse environments; resulting from their high mutation

rate and formation of quasi-species (Krausslich, 2011; Roossinck, 2012). Almost all the economically important groundnut viruses are ssRNA viruses (Sreenivasulu et al., 2008). These viruses impact a lot on yield losses. For instance, Peanut mottle virus can cause 30% loss in yield (Ghanekar, 1980), Bud necrosis virus infection in India has been reported to cause up to 50% yield losses (Ghanekar et al., 1979) and Peanut clump virus causing up to 60% loss in yields (Ghanekar, 1980). Groundnut rosette disease is the most devastating viral infections and can occur at very high levels resulting in a 100% loss in production. Additionally, average field incidence infection by chlorotic and green rosette is 40% within the fields in Western Kenya (Naidu et al., 1999; Wangai et al., 2001; CEFA, 2011). It is also the most familiar groundnut virus to farmers given its vibrant symptoms but a mixed infection in the field has been a cause of confusion in disease identification (Thottappilly and Rossel, 1997).

Furthermore, only three groundnut viruses have been surveyed in Western Kenya namely, Groundnut rosette complex, Cucumber mosaic virus; the second most important and Peanut mottle virus (Mukoye et al ., 2015). The current usable diagnostic tools are designed based on sequence-specific assays, and therefore, cannot identify different categories of groundnut viruses without prior knowledge of the targeted viruses (Ng, 2010). Additionally, most of these methods are less feasible for on-field pathogen detection and early disease warning utility. This is because they cannot be used for real-time detections (Fang and Ramasamy, 2015). Lack of a universal assay for the identification and characterization of all types of viruses has been and still is one of the biggest challenges in the monitoring and surveillance of plant virus status.

1.3 Rationale/justification of the study

The studies of relationships between viruses and their hosts have been biased towards the economically potent pathogenic aspect for years, that is despite viruses exhibiting mutualistic or symbiotic relationships with plants (Roossinck et al., 2015) such as groundnuts. Viruses being one of the most significant pathogens cause a reduction in groundnut yields and quality (Naidu et al., 1999; de Breuil et al., 2012). This can lead to reduced food production and nutritional security, low volume of local and regional trade, and thus loss of cash income augmenting the dire situation of high poverty levels in Western Kenya.

However, viral metagenomics comes handy as a viable platform to profile all the groundnut viruses. It is relatively a new powerful technique that is unbiased in sequencing multiple viruses from environmental samples, eliminating the need for costly and laborious downstream processes (Adams et al., 2009). In as much as viruses mutate rapidly, RNA-Seq based viral metagenomics has the capacity to provide real-time sequences and updated viral status unrivaled by current or other viral diagnostic techniques used before.

The study endeavoured to establish a platform for simultaneous identification of groundnut RNA viruses through metagenomics techniques. The present study provided the needed real time information that led a reduction of lengthy phyto-sanitation and virus screening procedures. The expected output enabled rapid screening and selection of groundnut genotypes that would be resistant to RNA viruses and the introgression of resistance into high yielding farmer-preferred varieties (FPVs) (Singh and Nigam, 2016). The results provided a rapid, reliable and robust viral detection method and, the creation of capacity for groundnut RNA viruses' management.

1.4 Research question

What are the categories of groundnut RNA viruses in Gem and Matayos Sub-counties, Western Kenya and can they simultaneously be detected and identified using viral metagenomics?

1.5 Study hypothesis

1. There are abundant groundnut RNA viruses that are widely distributed in Gem and Matayos Sub-counties in Western Kenya.
2. There are several unidentified groundnut RNA viruses as a result of emergence and re-emergence in Gem and Matayos Sub-counties in Western Kenya.

1.6 General objective

To determine the distribution, the composition and the diversity of groundnut RNA viruses in Gem and Matayos Sub-counties in Western, Kenya using high-throughput sequencing technologies.

1.6.1 Specific objectives

1. To determine the relative abundance, incidence/severity, distribution and prevalence of groundnut (*Arachis hypogaea* L.) RNA viruses in Gem and Matayos Sub-counties, Western Kenya.
2. To simultaneously detect and identify groundnut RNA viruses composition in Gem and Matayos Sub-counties, Western Kenya.

CHAPTER TWO

LITERATURE REVIEW

2.1 Origin, taxonomy and distribution of groundnut (or peanut)

Groundnut (or peanut, *Arachis hypogaea* L.) is a grain legume belonging to the family *Leguminosae* and the genus *Arachis* originating in South America region covering Brazil, Peru, Bolivia, Argentina and Paraguay. The native genus *Arachis* comprises of 80 described species including the cultivated species, *Arachis hypogaea* L., which is further assembled into nine (9) sections pertaining to geographical distribution, morphology and cross-compatibility relationships (Krapovickasi and Gregory, 1994; Valls and Simpson, 2005; Bertioli et al., 2011).

2.2 The genetics of groundnuts

The majority of species belonging to the genus *Arachis* have diploid genome (2n = 2x = 20) while three (3) species have aneuploidy genomes (2n = 2x =18) and two other species, *Arachis hypogaea* and *Arachis monticola,* are allotetraploid (2n = 4x = 40) combining two sub-genomes, AABB (Bertioli et al., 2011; Seijo et al., 2007).*Arachis* sections have five types of genomes (A, B, D, F and K) (Moretzsohn et al., 2013). Most species indicate an A type of genome, characterized by A-chromosome pair (Seijo et al., 2004). The remaining diploid species of haploid status, n =10 have "B" type genome of metacentric chromosomes with the exception of *A. glandulifera* (Robledo and Seijo, 2010), having a D type of genome with six sub-metacentric chromosomal pairs (Robledo and Seijo, 2008). The F and K genome types were the latest addition based on rDNA and heterochromatin presence (Robledo and Seijo, 2010).

The cultivated species, *Arachis hypogaea,* anallotetraploid (Seijo et al., 2007; Bertioli et al., 2011), has a probable genome origin from a single hybridization of the species, *A. duranensis* (AA genome) and *A. ipaensis* (BB genome) followed by a chromosomal duplication (Seijo et al., 2007). It has two subspecies namely *hypogaea* and *fastigiata,* which are further divided into botanical varieties; *hypogaea* which has var. *hypogaea* and var. *hirsute*; *fastigiata* is divided into var. *fastigiata*, var. *vulgaris*, var. *peruviana* and var. *aequatoriana* (Krapovickasi and Gregory, 1994; Ferguson et al., 2004). Only the subspecies *hypogaea,*var. *hypogaea*, *fastigiata,*var. *fastigiata* and var. *vulgaris* are the most cultivated in the major groundnut cultivation regions (Ferguson et al., 2004).

2.3 Groundnut production in the world

Currently, groundnut is widely distributed and cultivated as a major legume of the world, across 118 countries in the tropical and subtropical regions of the world (Abate et al., 2012).

Groundnut cultivation worldwide covers more than 22.6 million hectares and giving a production of about 36.4 million metric tonnes, averaging 1600 kg per hectares of yield globally and 1000 kg per ha in sub-Saharan Africa. India boasts the largest area under production but China is the highest producer of groundnut per unit area (Abate et al., 2012). South Asia accounts for 31% of world production total with India taking about 83% of it. Although the average production area has declined, yield and production have increased in South Asia with the fastest growth recorded in Myanmar.

Up to 44 countries are groundnut producers in sub-Saharan Africa, which provide 40% of the world total production. Production is expected to increase from 10.4 to 13 million metric tonnes by 2020 in sub-Saharan Africa (SSA) with the largest bulk of it coming from

Nigeria, Sudan and Senegal. The ten major producers of groundnut in SSA are Sudan, Senegal, Nigeria, Ghana, Chad, DRC, Tanzania, Guinea, Burkina Faso and Mali (Abate et al., 2012). The average productivity has grown by 1.3% and the fastest growth rate has been recorded in Cameroon (Abate et al., 2012).

2.4 Groundnut production in Kenya

Groundnut in Kenya is cultivated mainly by small-scale farmers and the production area is much smaller compared with other common crops. Kenya has the potential for high production of groundnut but depends on import from the Southern African region. Groundnut production is mainly concentrated in warm, humid areas, particularly along the coastal and lake regions - Western and Nyanza provinces (Masira, 2017). Groundnut is also grown in other regions such as the Rift Valley, especially in Trans Nzoia, Uasin Gishu, Nandi and West Pokot Counties. Most groundnut producing areas lie between altitudes of 1000 - 1500 meters above sea level with a mean temperature of 21-24 ^{o}C. Groundnut can also be cultivated at an altitude lower than 400 meters and a mean temperature of 24- 27 ^{o}C at the coast (Ministry of Agriculture, 2004a).

Homa Bay County is the leading groundnut growing area in the Western region and Kenya at large by production value, while Busia and Siaya Counties occupy fifth and sixth positions (Onyuka, 2016). Busia and Siaya Counties experience bimodal rainfall pattern between January-June and July-December with the first season receiving most rainfall, an average of 1775 mm. Temperature range of 21-25 ^{O}C is experienced in both counties (Masira, 2017; MoALF, 2016). The common groundnut genotypes grown in the two Sub-counties include ICGV-99568, ICGV-90704; Homa bay grows Valencia Red, ICGV-83708, ICGV-90704, ICGV-12988, ICGV-12991, J24 and CG-7, all of which have been

introduced by ICRISAT except in Homa Bay where Valencia Red is predominant. Only a small fraction of farmers plant local genotypes such as Uganda local red because they are less resistant to viral diseases (i.e., groundnut rosette disease) although they low yielding (Masira, 2017).

2.5 Groundnut agronomic practices

Groundnut production in Kenya is by small-scale farms and mostly intercropped with maize and sorghum in low input rain-dependent conditions. It is grown once or twice in a year depending on rainfall availability. Planting can be during the long rains (March-May) as well as short rains (late August-November (Ministry of Agriculture, 2004b). Groundnut viruses such as rosette are more common during short rains, which is followed by mid and end of season drought. These stressed conditions make viral infections severe (Okello et al., 2014). In recent past, efforts to achieve effective management have been focused on improving agronomic practices to delay the onset and spread of groundnut viruses' vectors and diseases, and on the breeding of plant host resistance (Naidu et al., 1999). Groundnut production in Kenya is labor intensive, right from planting to harvesting and shelling. This is due to the lack of appropriate mechanization technologies to be employed by farmers (MOA, 2004a; MOA, 2004b).

2.5.1 Groundnut growth requirements

Soil: Groundnuts grow well in soils that are deep, loose, well-drained and without compaction i.e., sandy soils. This makes it easy for pegs and roots penetration and water infiltration. Similarly, loamy, sandy loams or sandy soils are all suitable for groundnut cultivation (Directorate Plant Production, 2010; Desmae and Sones). The soil pH should be

about 5.5 to 7.0 while saline soils are not tolerable to groundnuts (Desmae and Sones, 2017).

Water: This depends on the variety of groundnut cultivated. In general, groundnut requires rainfall of between 250-1,000 mm during its growth period; 250-400 for extremely early maturing 250-400 mm; 300-500 mm for early maturing; and 500-1,000 mm for late maturing varieties. Rainfall of above 1,000 mm requires ridging to avoid water logging in groundnut (Desmae and Sones, 2017).

Temperature: Groundnut grows best in an optimum temperature range of 25-30 °C, while beyond 35 °C is above the threshold. For example, germination is delayed under lower temperatures exposing the seeds to soil pathogens attack and at below 17 °C, growth almost stops. In addition, cooler night temperatures may delay harvesting.

Altitude: Groundnut grows well in altitude of 1,500 meters and below with a likely desirable temperature for its growth (Okello et al., 2013; Desmae and Sones, 2017).

Land preparation: Groundnuts need deep well-prepared seedbeds free of hardpan or compact layers (CEFA, 2011; Desmae and Sones, 2017). Thus, groundnut suited to conventional tillage as compared to conservation agriculture.

Land preparation follows harvesting of the previous crop where the stalks and other crop residues are cut into small pieces and incorporated back into the topsoil, using a tractor or a hand hoe or any other plow. This process allows decay of crop residues to decay adequately and to prevent root rot diseases. Deep plowing is encouraged to break the hardpan, kill weed seeds by covering them deeper into the soil, and leaving the soil for easier roots and pegs to penetration and for easy harvesting of pods (CEFA, 2011; Okello et al., 2013; Desmae and Sones, 2017).

Groundnut can be cultivated either on ridges or flatbeds, but ridges are more recommended if drainage is poor. Additionally, groundnut cultivated on ridges tends to have higher yields because the soil becomes looser enabling better rooting and pod formation. Also, box ridges can be used with spacing at 75 cm apart, allowing double rows planting along at a row spacing of 30 cm. This is vital for rapid cover and smothering of weeds.

For a place with an issue of water scarcity, tied ridges are used to conserve moisture by reducing surface runoff and enhancing infiltration. Desirably, phosphorus (P) fertilizer; SSP or TSP is better applied before planting. Lime is usually applied to those soils that are acidic (Desmae and Sones, 2017).

Crop rotation: Like other legumes, groundnut should not be cultivated on the same plot for successive seasons. Rotation is encouraged, with groundnut cultivated every 2 to 5 seasons (Desmae and Sones, 2017). This is recommended for the following reasons: to avoid pests and diseases build up, i.e., insects, nematodes, leaf spots and white mold; avoiding soil nutrients depletion and improving organic matter; improving soil physical structure and reducing loss of humus as a result of loss that occurs at harvest time (CEFA, 2011; Okello et al., 2013; Desmae and Sones, 2017).The best crops to form a part of a rotation system with groundnut are sorghum, millet and maize. Others may include sweet potato, sunflower, and cassava (CEFA, 2011; Desmae and Sones, 2017).

Intercrops: Groundnut is known to be tolerant of shading, therefore, conveniently grown as an intercrop with crops such as castor beans, pigeon pea, bananas, cereals, cotton and sugarcane and with permanent crops such as rubber, cocoa and coconut palms. Groundnut intercrop with sorghum, maize or millet, gives a higher total profit than groundnut alone.

The priority of either groundnut or cereal will determine the choice plant pattern for intercrop by the farmer. The relative importance of the two crops is also taken into consideration (Desmae and Sones, 2017).

Inorganic fertilizer application: Groundnut responds to fertilizers with the major nutrients such as phosphorus (P), nitrogen (N), magnesium (Mg), potassium (K), Calcium (Ca) and other relevant elements (Okello et al., 2013). However, the fertilizers are costly to small-scale farmers though they are required for optimum yield for commercial groundnut production. Groundnut benefits from residual nutrients from the previous cropping season when it follows heavy feeders such as maize in a rotational system. Phosphorus is the essential nutrient required at early stages of growth. Due to its deficiency in most groundnut producing areas in Africa, it is supplemented by the application of inorganic form at the rate of 20 kg per hectare before planting (Desmae and Sones, 2017).

Time of planting: Early planting is required as soon as the rains begin in season since delay in planting groundnuts can reduce both the quality and yield of pods/grains (Sogut et al., 2016; Desmae and Sones, 2017). Therefore, planting should occur within 14 days of the start of the rainy season. However, planting immediately after heavy rain should be discouraged because the seeds can absorb too much water and rot, and a warm adequately moist soil is ideal for planting (Desmae and Sones, 2017). Groundnut sowing normally takes in the months of April-May for first season cropping system (Sogut et al., 2016). Groundnut seeds should be planted at a depth of six centimeters either in a furrow or a hole. Then filling and compressing the furrows or holes with soil to enable adequate contact between the seeds and soil. Very shallow or deep planting leads to poor

germination. Two seeds per hole are recommended but a seed per hole is preferred because it lowers seed rate (Desmae and Sones, 2017).

Late planting leads to delayed 50% anthesis and pod development and lower dry matter content. However, higher protein has been recorded in late-planted groundnuts. Overall, late sowing thus leads to a short growing season (Desmae and Sones, 2017), which affect harvesting dates, i.e. ten days delay after rains begin (Canavar and Kaynak, 2016).

Spacing: Spacing recommendation varies with the growth habit of the variety cultivated and nature of the plant cultivation. For example, the Spanish varieties require small intra-row spacing than the Virginia varieties which are runners and semi-spreading. Generally, the Spanish varieties should be planted with an inter-row spacing of 30-45 cm and intra-row spacing of 7.5-10 cm; the Valencia varieties are planted at an inter-row spacing of 60 cm and 10-15 intra-rows. The recommended row spacing can go as high as 90 cm, depending on variety (Okello et al., 2013; Desmae and Sones, 2017). The intra-row spacing of 10 cm is recommended as it gives a higher yield of up to 40% as compared to a spacing of 40 cm for rain-fed groundnut production. But close spacing can substantially reduce groundnut yield as a result of interplant competition (Naim et al., 2011).

Weeding: Weeds out-compete groundnuts, severely reducing yields if groundnut is not weeded enough (Desmae and Sones, 2017). Furthermore, weeds can affect groundnut by reducing shoot length, hence, plant height by up to 70% (Naim et al., 2011). Weed management, therefore, is important when groundnut is young, i.e., during the first 42 days, and as the pods set. Normally, 2-3 weeding regimes may be required. If early weeding and correct plant spacing are provided, groundnut adequately covers the soil and smother weeds effectively (CEFA, 2011; Okello et al., 2013; Desmae & Sones, 2017).

Other practices that also help to manage weeds in groundnuts include adequate field preparation, a deep turning of the soil, and crop rotation. Some farmers may use both pre- and -post emergence herbicides which are costly to many small-scale farmers. Noteworthy, weeding in groundnut is done with care for the following reasons: reduction in yield and increase in disease risk by covering of the plant with the soil; for avoidance of destruction of flowers while walking between plants**;** pulling weeds by hand to avoid disturbing soil at the base of plant at pegging stage (Desmae and Sones, 2017).

Earthing-up: This practice is carried at an early stage of growth cycle and has been found to have the potential of increasing pathogen and pest attack. Besides, reducing flowering and pod development at the lower nodes, reducing the total yield. During flowering, it may lead to damage of delicate flower buds in turn affecting peg formation.

However, earthing-up done following final weeding and gypsum application enhances peg formation and pod development through soil compaction at the root zone. This is especially useful for varieties that develop aerial pegs, which would otherwise be unproductive without earthing-up. It also assists the late formed pegs and pod formation. Although it encourages sprouting of early maturing with no dormancy while waiting for late pods to form or weakly attached pegs remaining in the soil during harvesting. In cases of intense and heavy rainfall, the pods may be exposed to the risk of pests and direct sun damage due to exposure of topsoil erosion (Desmae and Sones, 2017).

Irrigation: Groundnut is considered a relatively drought tolerant with capacity to do well in the semi-arid tropics, however, some varieties are more tolerant to water stress (Desmae

and Sones, 2017). In case water scarcity is a major issue, small-scale farmers should choose locally available drought-tolerant varieties. Groundnut needs enough moisture at its critical stages of growth: the first 14 days, from planting to the emergence of seedling, and at the peak of anthesis, pegging and pod development to achieve high yield. If an irrigation system is needed and available, light but frequent watering may be applied during the dry spell. Sprinkler irrigation is more efficient and beneficial to groundnut, while flooding method of irrigation due to excessive watering leads to water wastage and destruction by an operator. However, sprinkler irrigation encourages weed development and uses more water as compared to drip irrigation. Therefore, irrigation methods that waste water are discouraged in groundnut cultivation (CEFA, 2011; Okello et al., 2013; Desmae and Sones, 2017).

2.5.2 The relationship between agronomical practices and groundnut viruses

Management considerations in groundnuts have been influenced by the many viruses that infect it, and the actions taken in the form of re-adjusting the groundnut agronomic practices. These practices influence the incidence and severity of groundnut viruses at any given locality (Wright et al., 2016). The following are some of the consistent factors that have exhibited the relationship between the agronomic practices and the groundnut viruses.

Time of planting: Planting date has been an important aspect of managing groundnut viruses because it impacts both on the incidence and the severity of the viruses. For instance, early sowing at the beginning of the rainy season, the high intensity of rain adversely affects the vector population and prevent the spread of viral diseases such as rosette (Farrell, 1976) by limiting their contact with the host. However, rain coupled with warm and humid conditions canfavour the development and spread of vector, which in turn

increase disease development. Viral incidence has been noted to increase with changes in the growth cycle of the plant. Where the plants become prone to disease attack with aging as sexual generations of disease vectors such as aphids are produced at this stage. This increases viral infection due to the upsurge in vector numbers (Kone et al., 2017), i.e., late sowing supports growing stages of groundnut aphid, *Aphis craccivora* (Booker, 1963; Farrell, 1976).

Temperature: Deviation of temperature from optimum affects viruses in specific ways. Averagely high daily temperature can result in weakened symptom expression (Naim et al., 2011). Further, elevated temperature leads to a faster growth rate of disease symptoms developing on groundnut and initial accumulation of viruses such as Peanut stunt virus. Exponential increase in virus population at high temperature causes a reduction in an accumulated protein involved in carbohydrate metabolism, protein metabolism and photosynthesis (Renaut et al., 2015).

Rainfall: Collectively rainfall and temperature influence the dynamism of infection of both vectors and the viruses concern. It is evident that rainfall quantity and distribution influence the incidence of viruses, i.e. Indian peanut clump virus and fungus vector, *Polymyxa graminis* in case of natural viral transmission (Shoba et al., 2002).

Intercrop: Intercrop has been one of the cultural practices used to manage viruses in groundnut. Intercrop can reduce incidences of viruses such as Peanut bud necrosis virus by up to more than 69% and this is best exhibited by sorghum, pigeonpea, maize and bajra. In essence, intercrop act as a barrier to the movement of thrips, aphids and other insect vectors from one place another (Kenchanagoudar et al., 2005).

Weeding: The presence of volunteer plants and weeds within and around groundnut fields favour the occurrence of groundnut viruses'. Geographical distribution of viruses such as Groundnut ringspot virus is related to the vector population and the host reservoir which are mostly weeds (Naim et al., 2011). Peanut mottle virus has been recovered from weed host which act as their major reservoir (Demski, 1975).

2.5.3 Constraints of groundnut production in Kenya

Groundnut production constraints are myriad but pest and disease and low productivity are the major ones taking about 28.5% of all challenges. Other major challenges include drought, unavailability of healthy and improved varieties, inappropriate cultural practices, less stable government policies promoting bureaucratic procurement of inputs and poor market infrastructure (Onyuka, 2016).Drought is of a major significance because most of the groundnut cultivation is through rain-fed conditions. It can occur at any stage of the groundnut growth and development and thus affecting not only the physiology of the crop but also predisposing the groundnut pods to infection by *A. flavus,* an aflatoxin pathogen (Singh and Nigam, 2016). Other factors include unsuitable soil pH and temperature. These factors are common in America, Asia and Africa continents occurring in various combinations. Low soil pH causes nutrient fixation and calcium deficiency, which is an important limiting factor in groundnut production mostly in highly weathered soils in the tropics. In addition, low pH hampers nitrogen-fixing bacteria that aid the biological fixation of nitrogen by groundnut. Nitrogen and other major nutrients are important for pod filling and yields in groundnut (Chirwa et al., 2017).

Several biotic factors cause considerable yield losses in groundnut, these include viruses, insect pests, bacteria, fungi, and nematode (Kokalis-Burelle et al., 1997). Groundnut rosette is the most important constraint to groundnut production of all the groundnut

diseases (Coulibaly et al., 2017; Laing et al., 2018). The disease is endemic to sub-Saharan Africa and Madagascar (Subrahmanyam et al., 1997; Naidu et al., 1998). Rosette disease outbreaks are sporadic and unpredictable but when the disease occurs in epidemic proportions, yield losses are high. Rosette-plants produce significantly lower kernel yields (34-90%) depending on the severity of the disease (Naidu et al., 1999).

In 1975, groundnut rosette disease destruction was estimated to be 0.7 million hectares of groundnut worth the US $ 250 million in Nigeria (Naidu et al., 1998). In the same year losses up to about the US, $ 5 million was experienced in Eastern Zambia as a result of a rosette disease (Subrahmanyam et al., 1997). Waliyar et al., (2007), estimated yield loss due to rosette disease at about US $ 156 million per annum.

2.5.4 Groundnut viruses

Groundnut diseases caused by viruses are a major production constraint in all groundnut growing areas in the world (Sreenivasulu et al., 1991). Groundnut viruses are extremely important with significant impact on the quality and quantity of production (Sreenivasulu, et al., 1991; Sreenivasulu et al., 2008). Due to the frequent cultivation of other legumes in groundnut production areas, and the shared insect vectors, viruses spread from one crop to another. This has made epidemiological considerations and management of a single virus complex (Sreenivasulu et al., 1991). Groundnut is naturally infected by about 31 viruses belonging to 14 genera, a number which is unprecedentedly high compared with most close members of the same family (Sreenivasulu et al., 2008). There are eight (8) of these viruses which are economically important to groundnut today, three of which have already been reported in Kenya (Ryland, 1957; Bock, 1973; Mukoye et al., 2015). Some of these studies suggest the presence of asymptomatic and novel viruses in Western Kenya.

2.5.5 Identification of groundnut viruses

All the eight economically important groundnut viruses have ssRNA genome organization; three with linear, positive sense, monopartite ssRNA; a linear, positive sense bipartite ssRNA; two, linear, positive sense, tripartite ssRNA; and two, linear envelope & pseudo-circular ambisense tripartite ssRNA genomes (Sreenivasulu et al., 1991; Naidu et al., 1999; Sreenivasulu et al., 2008). The taxonomy of groundnut viruses is basically pegged on the physiochemical or intrinsic properties of the virus particles such as morphology and size, number and size of virus nucleic acid, number and size of virus polypeptide, and the genomic and biochemical mechanism of virus replication (Sreenivasulu et al., 1991). Several kinds of literature provide experimental procedures and guidelines for identification and characterization of viruses generally but laboratory facilities and scientific expertise are few or lacking in many groundnut production areas. This, therefore, necessitates a reliance on a combination of chemical, biological and serological properties of the viruses for identification and detection (Sreenivasulu et al., 1991).

2.5.6 Symptomatology of groundnut viruses

The viral attack on groundnut manifest in several ways, viz., concentric ring spots, stunting, various patterns of chlorosis, necrosis, stripe banding, mottling, mosaic patterns, and yellow blotches (Sreenivasulu et al., 2008). Variability in disease symptoms; type and intensity of the symptoms, depends on environmental conditions, diversity among the causal agents, differences in genotype responses, mixed infection with other viruses and variable nutrition (Sreenivasulu et al., 1991; Sreenivasulu et al., 2008).

Furthermore, different symptoms in the same host are frequently caused by strains of a virus, while specific strains of different viruses may cause similar symptoms in the same

host at some stage in disease development (Sreenivasulu et al., 1991). For instance, symptoms of bud necrosis caused by groundnut bud necrosis virus and spotted wilt caused by tomato spotted wilt virus are similar in manifestation; chlorosis and necrosis caused peanut mottle virus to mimic those of tomato spotted wilt virus while stunting and green patches of leaflets in peanut clump virus are often confused with those symptoms of groundnut rosette (Sreenivasulu et al., 2008).

Notwithstanding, symptoms are important in detecting and identification of the presence of virus infections and specific strains of viruses (Green, 1971). When used together with host range studies, they can be used to identify the virus, especially when a diagnostic host has been ascertained (Alemu, 2015). This is true for many plant viruses naturally infecting peanuts, and have already observed symptoms (Sreenivasulu et al., 1991). Preliminary identification studies of known groundnut viruses require selection of up to 8 to 12 host range plant species that are frequently infected by the familiar viruses for the reliability and for quick identification of the groundnut viruses using a diagnostic host. A host range key used for the tentative diagnosis of certain viruses' naturally infecting groundnut (Sreenivasulu et al., 1991), but not conclusive and comprehensive in the identification of all groundnut viruses at once.

2.5.7 Transmission of groundnut viruses

Groundnut viruses' transmission can be through seed, soil-based fungi, insect vectors and mechanical inoculation. These are the most common methods of transmission (Kokalis-Burelle et al., 1997). Insect vectors are the most important method of virus transmission. Most of these viruses are transmitted by several aphids and thrips species. For example, Cucumber mosaic virus, Peanut mottle virus, Peanut stunt virus, Peanut stripe and

Groundnut rosette viruses are transmitted by aphids. The first three viruses are non-persistently transmitted by various aphid species (Green, 1971; Sreenivasulu et al., 2008). Groundnut rosette complex; groundnut rosette assistor virus (GRAV) is also transmitted by the aphid, *Aphis craccivora*in a persistent and circulative way, while groundnut rosette virus (GRV), is only transmitted by the same aphid vector persistently in a mixed infection with GRAV (Green, 1971; Misari et al., 1988; Waliyar et al., 2007).

Groundnut bud necrosis virus, Tobacco streak virus and Tomato spotted wilt virus are vectored by thrips (Green, 1971; Sreenivasulu et al., 2008; Webster et al., 2015). Peanut clump virus is soil-borne vectored by a fungus, *Polymixa graminis* (Sreenivasulu et al., 2008) while Peanut mottle virus, Peanut stunt virus, Peanut clump virus, Peanut stripe virus and Cucumber mosaic virus are all transmitted through seed apart from insect vectors (Green, 1971).

2.5.8 Management of groundnut viruses

There are spelled out management strategies, which have formed an integrated system encompassing growing of tolerant or resistant genotypes and adoption of cultural practices that resist virus spread in an effort to manage the viruses (Sreenivasulu et al., 1991; Sreenivasulu et al., 2008).

Groundnut rosette disease complex is the most important groundnut virus in sub-Saharan Africa (Naidu et al., 1999; Waliyar et al., 2007). Most farmers do nothing when Groundnut rosette complex is spotted, while some do spray. An activity deemed expensive for most farmers. However, some level of management has been achieved by adhering to good agronomic practices and growing resistant varieties of groundnut (Okello et al., 2014). Like all viruses, there is no treatment in case of infection. With this in mind, a long-term strategy has been to develop resistant genotypes against groundnut rosette and other viruses (Olorunju et al., 1991; Naidu et al., 1999). It has been identified that both green and

chlorotic rosette are controlled by two recessive genes (Nigam and Bock 1990). Based on this premise, groundnut genotypes resistant to various viruses have been developed and released (Waliyar et al., 2007). However, most of the developed genotypes are long maturing, small-seeded and low yielding as compared to farmer preferred genotype (Ntare and Olorunju, 2001). Additionally, only a few resistant groundnut genotypes have been developed in Africa (Waliyar et al., 2007). This provides an avenue for the identification of resistant genotypes among the available genotypes cultivated currently, and possible introgression into the farmers preferred genotypes.

2.6 Genomes, abundance, composition and distribution of groundnut viruses

Groundnut viruses like other biological organisms have either DNA or RNA genomes. The functions required for the viruses to complete their life cycles and for interaction with their surroundings are encoded within the genome. Viruses generally have the smallest genomes of all the biological organisms but show the greatest variation in the nature of their genomes. These unique features can be revealed by the sequencing process. Moreover, both RNA and DNA viral genomes can be either double-stranded or single-stranded. Additionally, the genomes can be monopartite or multipartite; having several segments. However, nearly all the double-stranded DNA sequenced so far are monopartite and only a few of the single-stranded DNA genomes are multipartite. Conversely, multipartite genomes are a commonplace when it comes to RNA viruses and especially single-stranded negative RNA genomes.

RNA viruses use RNA-dependent RNA polymerase which is encoded in the positive ssRNA genome for replication while for the negative ssRNA genomes must be copied and the copy later used as a transcript. Some RNA viruses like the retroviruses use reverse

transcriptase to aid this process. The RNA viruses are mostly smaller than DNA viruses. The fragility of both single-stranded RNA and DNA viruses explains the small nature. RNA viruses have small size genomes as a result of susceptibility to mutation (Mcclean, 2004).

The abundance of viruses has been found to be determined by environmental variables and among other factors. Viral abundance a cross-location and time relates to environmental factors, but majorly phosphorus and nitrogen concentrations and bacterial abundance as well. Essentially, mineral elements' concentrations affect virus production and in turn the abundance. As Illustrated by the stoichiometry of the viral particle, phosphorus and nitrogen form key components in virus replication (Suttle et al., 2017).

Plants are naturally infected by two or more viruses, these different interactions among viruses can be grouped into antagonism, synergism, cross-protection, replacement, mutual suppression, helper-dependence and a mixture of antagonistic and synergistic interactions (Zeng et al.,2007). These interactions usually lead to unpredictable epidemiological and biological manifestations in a mixed infection. The effects of the mechanisms of these interactions in multiple infections of the viruses are not yet deciphered.

Some research has managed to identify less synergistic interactions as compared to antagonistic effects in multiple plant viral infection. A mixed infection of some known groundnut viruses has been reported, i.e. Cucumber mosaic virus mixed infection with Watermelon mosaic virus-2 (WMV-2), Zucchini yellow mosaic virus (ZYMV), Squash mosaic virus (SqMV), Cucumber green mottle mosaic virus (CGMMV), Papaya ringspot virus-W (PRSV-W) (Nontajak et al., 2014). Also reported is a mixed infection of Peanut

mottle virus (PeMov) with Cucumber mosaic virus (CMV) or Groundnut ringspot virus (GRSV), which form part of the viral genera potyviruses, cucumoviruses, and tospoviruses (Breuil et al., 2008). An infection of multiple viruses or single virus will show similar symptoms during the early stages of viral infection (Nontajak et al., 2014). This makes reliance on symptomology only to identify a specific viral infection in groundnut not feasible (Okello et al., 2014). In addition, both CMV (Mukoye et al., 2015) and rosette complex virus (Waliyar et al., 2007) have been identified in Western Kenya, making the possibility of these interactions likely.

In Kenya, little has been documented on groundnut viruses. Only three viruses have been reported so far, these are groundnut rosette virus, Peanut stunt virus and Cucumber mosaic virus (Mukoye et al., 2015; Waliyar. et al., 2007). Most studies have focused on groundnut rosette virus (Naidu et al., 1998; Naidu et al., 1999; Waliyar. et al., 2007). Field studies have identified both green and chlorotic rosette symptoms with field incidence ranging from 24-40%. The highest incidence was recorded in Homa Bay and Kendu Bay, while the Rift valley region has an average incidence of 30%. In Busia and Kisumu extending up to Gem, Siaya county indicating a low field severity and incidence of 24% which is attributed to late field infection (Waliyar. et al., 2007).

2.7 Viral diagnostic methods

Plant viruses of all dimensions pose a major threat to crop production in many places in the world, but our comprehension of viral diversity is very little. Moreover, monitoring of plant viruses has been skewed towards those with vibrant symptoms and those of economic importance to crops. More is needed to be discovered about the asymptomatic and emergent novel viruses (Ng, 2010). Nevertheless, a few studies have been conducted so far

in Western Kenya and the entire country at large using the existing groundnut viruses' diagnostic tools (Mukoye, et al., 2015).

2.7.1 Traditional viral diagnostic methods

These methods have been used since the start of virology in the 1880s. For instance, field diagnosis of plant viral infections, which are based on vibrant symptoms have a limitation in that they can result in misleading or difficult diagnosis of viruses or pathogens in case of array of strains for the same virus infection and induction of a variety of symptoms as a result of environmental influence such as temperature (Dietzgen et al., 2001; Culbreath et al., 2003; Adams et al., 2009; Boonham et al., 2014). Therefore, some viral infections may go unnoticed. In comparison, bioassays involving sap inoculation from suspected virus-infected plants to diagnostic hosts, e.g., *Chenopodium amaranticolor* used in case of Groundnut rosette complex viruses is time-consuming and unsuitable for testing many samples in case of persistent sap inoculation. But they are an important first tentative viral investigation (Sreenivasulu et al., 2008; Boonham et al., 2014).

Moreover, use of transmitting viral vectors to indicator hosts are least suitable in routine or even large-scale applications, while the use of electron microscopy is unsuitable for large-scale testing and faces complications due to confusion between viral particle and host-membrane.

2.7.2 Current viral diagnostic methods

2.7.2.1 Serologically – based phytodiagnostics

The introduction of the ELISA technique in 1970s revolutionized phytodiagnostics as it became the most successfully established plant viral detection tool (Mathews, 2010). The key to this was a highlevel of reproducibility and repeatability converse to previous traditional diagnostics. Its ease of establishment, accessibility and robustness even with

inexperienced and unfamiliar users made it a laboratory-based test of choice. The traditional methods required great amounts of antisera and a limited format, therefore, only a few samples could be tested (Sreenivasulu et al., 2008). In contrast, not only did ELISA simplify investigations to get conclusive results but also the time needed to do so. More advancement was further brought about by the use of monoclonal antibodies improving assay specificity and sensitivity through adjustment of the test format, making it the most versatile assay for viruses that are simple and sensitive (IITA, 1971). A specific technique called triple antibody sandwich ELISA (TAS- ELISA) is still in use today for Groundnut rosette disease (D'Arcy et al., 1989; Naidu et al., 1999).

A serological-based multiplex version of ELISA expansion with immunological detection capability called Microsphere immune assay (MIA) is offering a high-throughput for multiple viruses, but is still limited to targeting only known viruses. Immunological blot and ELISA have been optimized for routine and large-scale detection systems. All the same, ELISA is not meeting all challenges despite its widespread use because it requires high-quality antisera for sensitive and specific viral binding antigen, antisera purification and production is lengthy and requires expertise and inability to give absolute resolution for different viral strains (Boonham et al., 2014).

2.7.2.2 Nucleic acid-based phytodiagnostics

The use of molecular-based assays has made the limitations of earlier diagnostics be at a touching distance and hence, revolutionizedphytodiagnostics further. Molecular-based methods focus on DNA/ RNA viral detection or nucleic acid hybridization techniques. These methods have given way to conventional PCR, quantitative PCR (q-PCR), real-time PCR (RT-PCR) (Vandesompele, 2009) and Loop Amplified mediated polymorphism (LAMP) (Thekisoe et al., 2009), which have inherent compatibility and flexibility as

compared to ELISA techniques. RT- PCR has achieved both high levels of sensitivity and specificity and has eliminated the problem of post- PCR contamination common with conventional PCR (Dorak, 2006). RT-PCR is also more sensitive than ELISA because it can detect viruses at pictogram level compared with a milligram required by the latter (Boonham et al., 2014; Sastry and Zitter, 2014).

Nucleic acid-based methods have been multiplexed to simultaneous detection of multiple viruses in a single assay, reducing the cost, time andlabour- input required for single conventional detection techniques. Multiplex PCR has been successful in detecting multiple groundnut viruses in seeds (Dietzgen et al., 2001). The caveat is a limited range of multiple –colour signals that can be detected, thus only a narrow range of assays can accurately be detected in a single tube. Microarrays techniques are platforms for simultaneous detection and identification of multiple viruses, with good sensitivity and specificity using known probes from either variable or conserved regions of the targeted viruses (Pearson and Wei, 2007). Although not suitable for high-throughput testing, are better placed for screening new and unusual viruses. Summarily, all the nucleic acid-based techniques involve universal, degenerate or pre-designed primers and require prior knowledge for diagnostics to be conducted, hence cannot detect novel and new uncharacterized viruses (Pearson and Wei, 2007; Boonham et al., 2014).

2.7.3 Viral metagenomics and Next-generation sequencing

Next generation sequencing (NGS) is high throughput sequencing technologies which became available not long ago and are still in continuous development and improvement. Next generation sequencing has gained wide usage in projects such as metagenomics, RNA sequencing, whole genome sequencing and small RNA discovery, its common denominator being high-throughput data generation. The improved sampling of diverse

environments and the advancement in the development of NGS techniques have accelerated the rate of new metagenomes discoveries. Positive results have been yielded with several studies, where diverse plant viruses have been recognized using a range of sequencing platforms either starting with DNA or RNA (Adams et al., 2009; Al Rwahnih et al., 2009; Kreuze et al., 2016; Gutiérrez et al., 2016).

Surprisingly, the focus on the major challenges associated with metagenomics studies has shifted from data generation to analysis. The current challenges are in the forms of a large volume of data, heterogeneity of genome, incomplete sequences and sparse metadata (Costa et al., 2010; Argonne and Meyer, 2016). Metagenomics has been about the analysis of sample drawn directly from the environment having an unknown mixture of uncultured life forms, i.e. virome (Rampelli et al., 2016). Metagenomics geared towards the analysis of virome only is referred to as viral metagenomics. Viral diagnostics is one of the most successful applications of metagenomics and with amazing achievements in the discovery and characterization of new viruses and novel viral pathogens (Barzon et al., 2011). Additionally, the rapid development of NGS techniques has drastically reduced the time and cost needed for a pathogen such as a virus identification by metagenomics (Wu et al., 2015; Jain et al 2016).

RNA sequencing has been combined with metagenomics techniques to form RNA-Seq - based metagenomics, with many merits over culture-based methods and molecular tests(Schlaberg, 2016). RNA-Seq is the most sophisticated of the NGS studies; used for transcriptome studies which can be essential when comparing samples for disease-related projects. The attribute is not easily achievable by the earlier extensively used hybridization-based techniques. In contrast, RNA-Seq is not limited to the detection of

known transcripts but allows the identification, characterization and quantification of new splice isoforms to determine the correct annotation.

Despite the complexity, RNA-Seq is becoming a technique of choice with an edge above most hybridization-based techniques (Costa et al., 2010). Several studies involving RNA-Seq have adopted different upstream processes to achieve maximum yield for the target reads (Ozsolak et al., 2009). For example, less optimization required as compared to Sanger-based sequencing which allows errors introduced from cloning DNA to reach sequencing (Costa et al., 2010; Jain et al., 2016). To curb the problem, high fidelity enzyme is introduced, but errors occurring early in amplification are still possible to cause incorrect sequencing. The proposed massively parallel sequencing of RNA molecules, directly without prior synthesis of cDNA or the need of ligation/ amplification steps can potentially bypass all these errors (Marston et al., 2013; Jain et al., 2016), i.e., Nanopore MinION sequencing platform can identify viruses in under six hours from sample collection, reading individual bases based on induced characteristic changes of current caused by each base (Jain et al., 2016). But Illumina Hiseq platforms are preferred by a majority of researchers because of the low raw error rate of less than 0.4% and low false positive in comparison to Ion Torrent of greater than 1% (Quail et al., 2012).

In addition, the first generation of Illumina-based RNA-Seq used random hexamer primers to reverse transcribe poly (A) selected mRNA, but this has been faulted for not retaining the information contained on the DNA strand that is expressed. Although enrichment of poly-A RNA for deep sequencing has led to the discovery of new viruses, it also deletes reads from many families of RNA viruses that do not synthesize poly-A RNA such as cucumber mosaic virus and tobacco mosaic virus which is of interested in this study. Therefore, use of selective RNase H-based digestion for the depletion of the undesired

RNA-including poly (rA) carrier and ribosomal RNA from the viral RNA samples is to increase the proportion of the desired RNA (Matranga et al., 2016).

2.8 Bioinformatics

2.8.1 Analysis of RNA sequence (RNA-Seq) data

The power of RNA sequencing is anchored on the two aspects, gene quantification and identification of transcript being combined in a single high-throughput sequence assay. This has led to extensive adoption of RNA-Seq and the spread beyond genomics, becoming a standard toolkit in the life science research community. A variety of RNA-Seq protocols and analyses have been availed but challenging to new users from appreciating all the steps involved in conducting the process.

Presently, there is no optimal pipeline for the many different applications and analysis scenarios in RNA-Seq. Each RNA-Seq experimental incidence can possibly have a different optimal method for transcript quantification and differential expression analysis. Planning of experiments and the adoption of different strategies by researchers is based on the organism under study and the research set goals. For instance, if a reference genome is available, the organism's transcript can be identified by mapping the RNA-Seq reads to the genome. In contrary, an organism without reference sequence genome, identification and quantification can be achieved by the de novo assembly of reads into contigs and afterward map the contigs onto the transcriptome (Conesa et al., 2016).

Moreover, for a study which focuses on profiling the taxonomic and phylogenetic composition of viral communities, it is pivotal to understand both the host virome interplay and other relationships role in plant biology (Rampelli et al., 2016). Without overemphasizing the evidence of the importance of the interplay among virome,

microbiome, and other relationships, the available techniques for virome characterization usually underestimate the quantity and diversity of viruses in the samples (Castrignano and Nagasse-sugahara, 2015). For example, the viral isolation methods based on filtering procedures miss giant viruses (Hall et al., 2014) while viral characterization is also difficult because of lack of a single common gene to all viral genomes. This rules out techniques analogous to that of bacterial profiling using rDNA (Virgin, 2015).

The viral taxonomic composition of metagenomics samples' RNA-Seq and short-gun sequencing can be estimated by the detection and assignment of the viral reads to the corresponding viral taxa. The current most advanced experimental procedures for manipulating metagenomics samples can only possibly extract and isolate the enveloped viruses (Duhaime and Sullivan, 2012) at later stages, to characterize the viral metagenomes by assembly or reads mapping techniques. Standalone software, easy to control modules within pipelines and pipelines such as viromeScan, a tool that requires only a few minutes to accurately profile viral communities from thousands of metagenomic reads has been used successfully in other studies. It uses sequenced samples with no upstream processing and directly assigning the reads to the right taxa without missing giant viruses because of filtering processes. Therefore, the potential to profile all the viruses within the microbiome, i.e., approximating the relative abundance of viruses by filtering out the reads of undesired genomes and mapping the remaining viral sequences on hierarchical virus databases.

Moreover, standalone software has a number of trade-offs; firstly, specifically dedicated to the analysis of viromes, providing an edge over most general software. Secondly, analysis is fast because of the user's option to choose from several localized and inbuilt reference databases. Thirdly, an exhibition of high levels of superiority- stages of filtration;

screening and selection of candidate sequence reads, and finally mapping of those reads. Fourthly, a resolution up to the level of species with a few errors and the ability to preserve all the information confined to the input files (Rampelli et al., 2016). However, like most currently used pipelines, it is blind to those viruses without a reference sequence genome in the databases (Palacios et al., 2008; Popgeorgiev et al., 2013; Rampelli et al., 2016). ViromeScan works through screening and selecting of candidate viral reads while the unmapped reads are excluded from the proceeding input file (Rampelli et al., 2016). With all the above-enumerated merits, standalone software is still not always user-friendly in terms of installation. This potentially limits their use in any study. To achieve a more exhaustive viral profiling, parallel analysis of the unmapped reads maybe employed concurrently. The unmapped reads file can be retrieved for use in the downstream analysis as described in (https://www.biostars.org/p/89123/). Afterwards, de novo assembly of unmapped reads can be conducted using Megahit (Li et al., 2016; Sadakane and Lam, 2016) or Velvet software because of its low error rate and large sequence coverage as compared to other assemblers available (Zerbino and Birney, 2008) while mapping with Bowtie2; a memory efficient and ultrafast tool suitable for long reference sequences (Ben Langmead and Salzberg, 2013). The open reading frames (ORFs) identification for the viral sequences can be achieved using MegaBLAST and Glimmer version. 3 in a cheaper, faster, and with improved sensitivity and specificity (Delcher, 2006; Morgulis et al., 2008).

Furthermore, phylogenetic reconstruction of amino sequences or nucleotides is inferred for the viral contigs using the Maximum Likelihood method incorporating the LG+G+I substitution model (Le and Gascuel, 2001). Maximum likelihood analyses can be done using nucleotide sequences calculated with the HKY+G+I model (Hasegawa, 1985). MUSCLE is used here for multiple sequence alignments purposefully because of its high

accuracy and high throughput in the case using large datasets (Edgar et al., 2004). Finally, evolutionary analyses are conducted in MEGA using 1000 bootstrap replicates to portray the most optimal relationships among the different viral sequences as per our methods (Tamura et al., 2013). SeaView version 4 becomes also an alternative to MEGA because of its user friendly nature having a multiplatform for most of the algorithms required for building a phylogenetic tree. PhyML module within SeaView version 4 is used to construct a maximum-likelihood phylogenetic tree diagram (Gascuel and Gouy, 2017).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Description of the study area

The suspected virus infected groundnut leaf samples were collected from groundnut fields across two (2) sub-counties: Matayos and Gem in Busia and Siaya Counties respectively. These are some of the places groundnuts are commonly grown but with minimal research work undertaken. These two counties are bordered by Vihiga and Kakamega to North East, Bungoma to the North, Kisumu to the South East and Homa Bay to the South and Lake Victoria to the West. The two Counties occur between latitudes 0^o 28'N to 0^o 42' N and longitudes 33^o 58'E to 34^o 42'E with an average altitude of 1270 meters and covering an area of about 3664 km^2. The average annual rainfall is 1632 mm and the highest amount of rainfall is received from the month of March to April, while the driest month is January (CLIMATE-DATA.ORG; https//en.climate-data.org/location/11165-8/). Importantly, the agro-ecological zones and farming systems tend to overlap between these two Counties and Sub-counties.

The farming system in Gem and Matayos Sub-counties is mainly subsistence. This tends to be intensive subsistence and commonly polyculture farming system. Maize has been reported as the most commonly grown crop, which often occurs in a mixed stand with beans. Other crops cultivated in this region include bananas, sweet potatoes, cassava, millet, sorghum, kales, pumpkin, groundnuts and pineapple. Maize can also be found planted under established long maturing crops such as bananas and pineapple. Livestock farming is also common with most farmers keeping a few numbers of a different kind of livestock. They benefit mostly from the crop residues after harvesting (Nduru and Bein, 2011). In Busia County, nuts and oil crops such as cashew nut, coconut, rapeseed,

macadamia nut, sunflower and *simsim*, in addition to peanuts are major cash crops and support more than 300,000 households (Masira, 2017).

3.2 Sampling strategy

Sampling was conducted between January and July in these two Sub-counties according to published protocols (Ndunguru et al., 2009). Unbiased sampling points (groundnut plants) were randomly selected along the transect lines walking in a double Z-pattern for all the 24 farmers' fields sampled. The leaves were chosen from the lower, middle and upper regions of the groundnut plant (Ndunguru et al., 2009). A bulk sample was obtained from each sampling site. This gave a total of 24 samples, of which, 21 of the samples were from symptomatic sampling sites while 3 were from asymptomatic sampling sites. The samples were kept in duplicates; storage in RNAlater® solution (Ambion, Life Technologies, CA, US) and in wet-tissue during transportation to the ICRISAT laboratory, Nairobi for RNA extraction. Global Positioning Systems (GPS) readings were taken at each sampling point using a GPS receiver and the coordinates used for generation of the map using the Tableau Public v.2018.3 software as shown in the **Figure 3.1**. In this map some coordinates have overlapped appearing as single points. A picture of one of the visited sampling sites is provided in the **Figure 3.2** below.

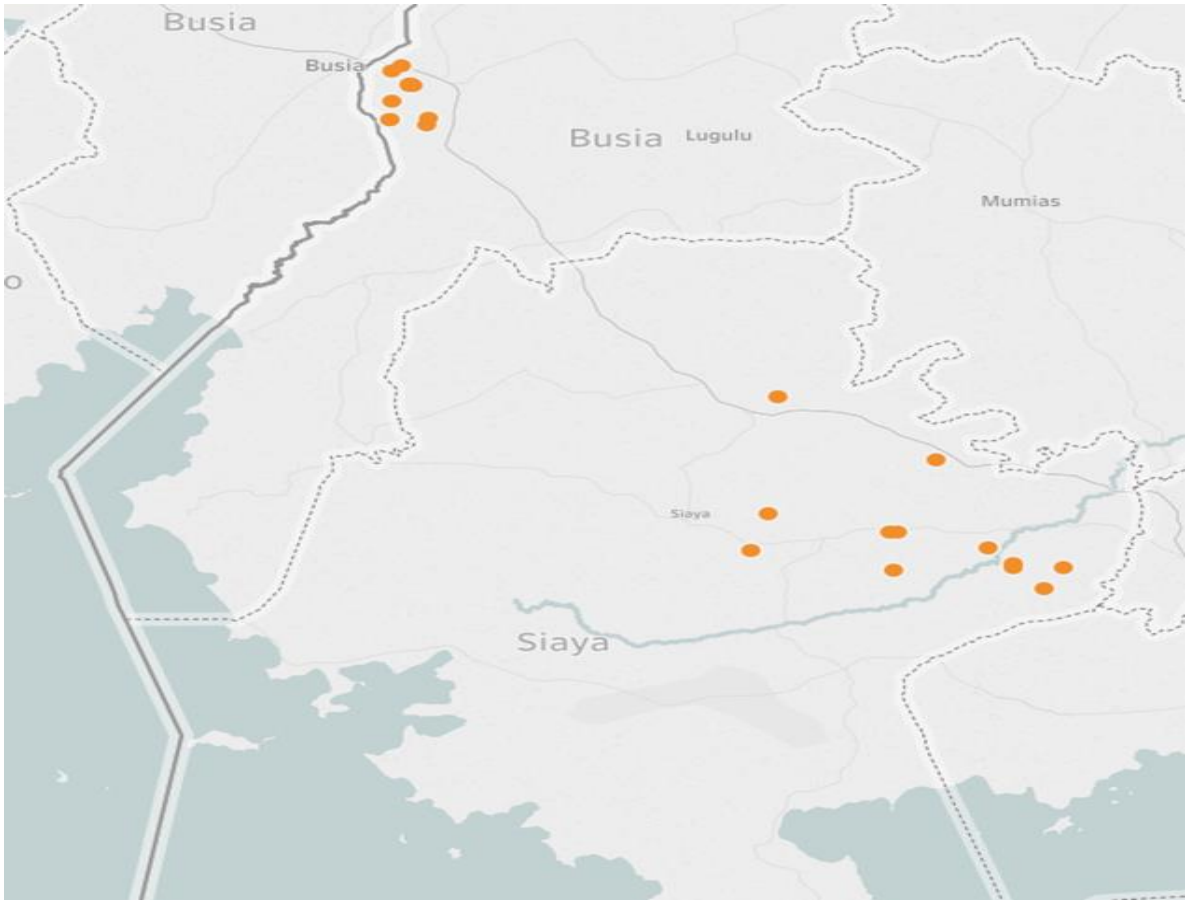


Figure 3.1: Map of sampling sites in Matayos Sub-county, Busia County and Gem Sub-county, Siaya County drawn using Tableau Public v.2018.3 software.

Figure 3.2: A picture taken at one of sampling sites in Gem Sub-county, Siaya County.

The severity of groundnut viruses at each sampling point was evaluated on a scale of 0-5; 0- representing no symptom, (1-3) – mild to moderate symptoms, and (4-5) severe to very severe symptoms. In addition, groundnut viruses' incidence was represented as a percentage (%) of visually infected plants counted over the total plant sampled in each field (Domola et al., 2008; Ndunguru et al., 2009) and severity as percentage of disease severity as described by Getachew et al., (2014).

$$\text{Viral disease incidence (\%) (PDI)} = \frac{\text{Number of infected plant counted X 100}}{\text{Total number of units assessed}}$$

Viral Disease Severity (%) (PDS) = nxv /5N x100;

Where,

(n)= Number of plants in each rating category, (v) = Numerical values of symptoms category.

(N)= Total number of plants, (5) = Maximum numerical value of symptom category.

3.3 Groundnut leaf materials, total RNA isolation, purification and quantification

The groundnut leaf samples were collected from the sampling sites in the two Sub-counties after 45 days following planting. Total RNA was extracted from all the 24 groundnut leaf samples kept in wet-tissue and RNAlater® solution (Ambion, Life Technologies, CA, US) using Direct-zol MiniPrep kit (ZYMO Research), in conjunction with TRI Reagent Solution (Ambion, Life Technologies, CA, USA). About 100 mg fresh sample was homogenized in liquid nitrogen and then mixed with three times volume of TRI Reagent in 2 ml microcentrifuge tubes. This was then followed by phase separation, DNase I digestion, RNA precipitation, RNA washing, RNA elution and storage as described in the protocol by ZYMO Research Inc. The RNA quality was evaluated with an agarose gel

electrophoresis and the concentration measured with Qubit® RNA BR Assay Kit (Life Technologies, USA).

3.4 First strand cDNA preparation, viral enrichment and metagenomics sequencing

To obtain metagenomics cDNA, first strand cDNA was synthesized using the Thermo Scientific RevertAid First-strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc., USA) as per the manufacturer's instructions. About 100-500 ng of total RNA sample and 2-4 µL of random hexamer primer was used in either 20 µL or 40 µL volumes. The reverse transcription was conducted using a thermal cycler at 25 ºC for 5 min, followed by 45 ºC for 60 min and then terminated by heating at 70 ºC for 15 min. This was conducted largely as described by the manufacturer's instructions.

To enrich the groundnut viruses' RNA present, a selective depletion of rRNA using a thermostable RNase H was incorporated in the protocol. This was done by adding 1 µL of RNase H to each reaction tube followed by 20 minutes incubation at 37 ºC.

The first strand cDNA samples were sent to Xcelris Labs Limited, Gujarat, India for metagenomics cDNA library preparation. The concentration and final size of each library were confirmed using a Bioanalyzer (Agilent Technologies, CA; USA) and Qubit. The metagenomics sequencing was performed at Xcelris Labs Limited, Gujarat; India on Illumina HiSeq 2500 platform using 2x150 cycle reagent kit ((Illumina, Inc); Jain et al., 2016; Kamitani et al., 2016).

3.5 Analysis of metagenomics sequences

No single pipeline can be used to comprehensively analyze metagenomics sequence data. And this study was not an exception. Therefore, this study integrated several individual software and pipelines to achieve the extensive analysis of the aspects of all viruses present in the groundnut viruses' sequences.

3.5.1 Quality control

Paired-end reads in the fastq format were generated from the HiSeq 2500 sequencing platform (Illumina, Inc.). The quality of the raw reads was visualized with FASTQC v.0.11.5 then followed by Illumina universal adapter removal, read trimming and filtering with Trimmomatic v0.36 (Bolger et al., 2014) to ensure a default minimum length of 36 bps, maximum length of 150 bps, and a sliding window that cuts a read once the average quality in a window size of 4 falls below a Phred score of 30.

3.5.2 Host subtraction

The groundnut specific reads were removed by the end-to-end alignment mode using bowtie2 (Ben Langmead and Salzberg, 2013). The unmapped were retrieved for further analysis while the reads were mapped to the cultivated peanut genome (*Arachis hypogaea* cv. Tifrunner genome) discarded. This was an effort to reduce the overwhelming number of host reads that were present.

3.5.3 Reads error correction and metagenomics assembly

The host-free and contamination free reads were error corrected using the inbuilt module within the metaSPAdes assembler (Nurk et al., 2017). The reads were then assembled into contigs using the metaSPAdes (Nurk et al., 2017) and the megaHIT assembler (Li et al., 2016). The quality of assembly using the two assemblers was assessed and visualized using the MetaQUAST (Mikheenko et al., 2016), it compares the statistics of the assemblies software outputs and provides a basis for deciding on the best assembly outcome based on the parameters used.

3.5.4 Taxonomic classification

Clean reads and the assembled contigs were classified into known viral species using Kraken (Wood and Salzberg, 2014) classification module and the downloaded databases within MetaWRAP pipeline (Uritskiy et al., 2018). Kraken (Wood and Salzberg, 2014) uses exact k-mers of the sequences or contigs for accurate classification up to the species level and comparable to the fastest BLAST program (Altschul et al., 1990), but faster than Megablast.

3.5.5 Binning and refinement of metagenomics dataset

Binning of contigs following the assembly of all the samples was done using the binning module within the MetaWRAP pipeline and the obtained bins refined using the binning refinement in the same pipeline. The appropriate MetaWRAP commands were adjusted and simple scripts created to run these modules (Uritskiy et al., 2018).

3.5.6 Viral metagenomics annotation

To further reduce the level of bacterial contamination if at all still present following the reads assembly into contigs, viral contigs were selected using Virsorter (Roux et al., 2015), by predicting virus-like reads only. The selected contigs were annotated using BLASTn and seaview (Gascuel and Gouy, 2017). This followed the online alignment of the contigs using MAFFT (Katoh et al., 2002) with its default settings. Estimations of maximum likelihood phylogenetic trees were produced from the aligned contigs using PhyML (Guindon et al., 2010) implemented within SeaView using K2 substitution model and 1000 bootstrap replicates (Gascuel and Gouy, 2017). Finally, a summary of the entire groundnut viral metagenomics workflow drawn is provided in the **Figure 3.3** below.

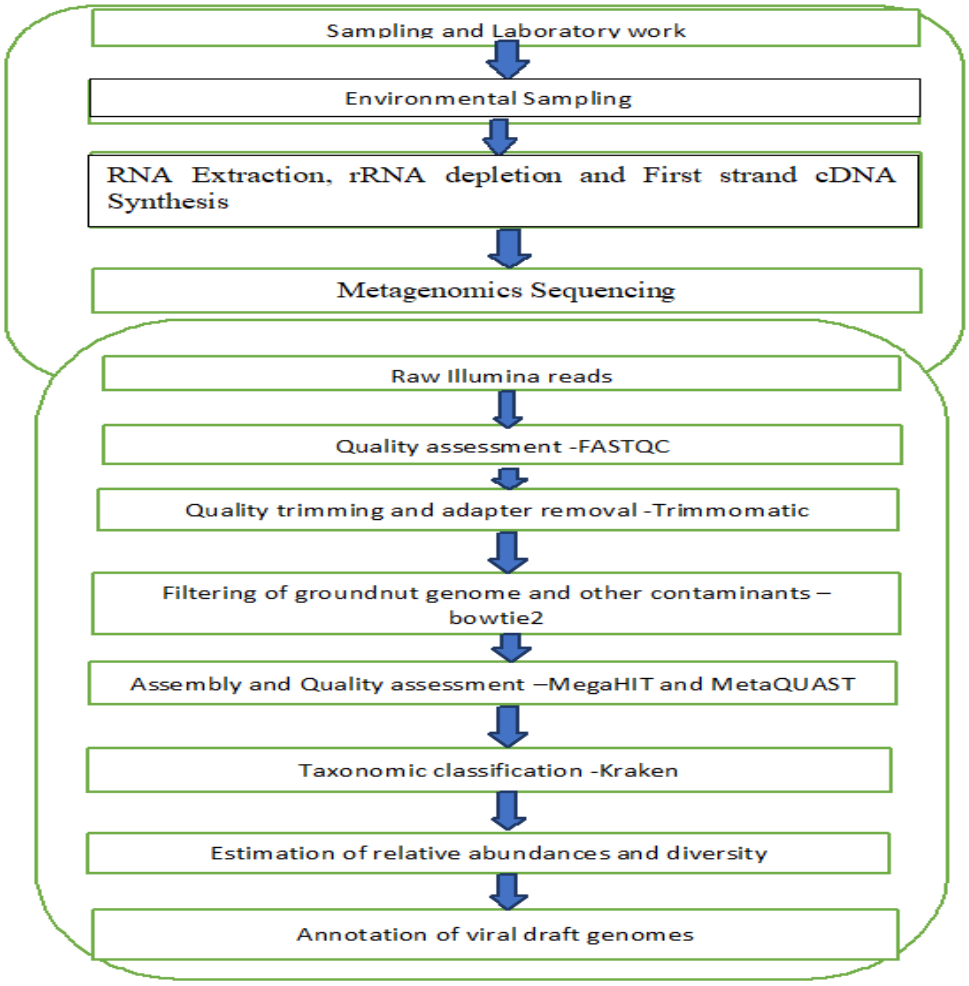


Figure 3.3: A schematic workflow of groundnut viral metagenomics analysis drawn at the end of the analysis.

CHAPTER FOUR

RESULTS

4.1 Groundnut RNA viruses' incidence and severity, relative abundance, distribution and prevalence in Gem and Matayos Sub-counties, Western Kenya

4.1.1 Incidence and severity of groundnut RNA viruses

The assessment of viral incidence and severity was done in both Matayos Sub-county in Busia County and Gem Sub-county in Siaya County. This involved calculation of the incidence and severity of plants assessed for a sampling site and their respective means converted as a percentage. This was in reference to a study by Getachew et al., (2014). In the present study, both viral incidences and severity were higher in Gem Sub-county, Siaya County in comparison to Matayos Sub-county, Busia County. This is represented in **Table 4.1** below. Three samples at the start of the study were asymptomatic because they had no visible symptoms for viral infections.

Table 4.1: Percentage viral disease incidence (PDI) and percentage viral disease severity (PDS) in Gem Sub-county, Siaya County and Matayos Sub-county, Busia County

Location/Region					
Gem Sub-county, Siaya County			**Matayos Sub-county, Busia County**		
Sampling site/Sample ID	PDI % Incidence	PDS % Severity	**Sampling** site/Sample ID	PDI % Incidence	PDS % Severity
G1	80	64	B1	40	16
G7	80	64	B2	50	25
G8	70	49	B3	40	16
G12	80	64	B4	40	16

G13	90	81	B5	60	36
G23	60	36	B7	30	09
G24	100	100	B8	70	49
G25	70	49	B9	00	00
G27	50	25	B10	60	36
G28	40	16	B11	40	16
G30	00	00	B15	40	16
G33	00	00	B16	50	25
Mean =	60	45.67	Mean =	43	21.67

The viral disease incidence and severity from **Table 4.1** above were separately represented graphically as shown in **Figure 4.1 and 4.2** below.

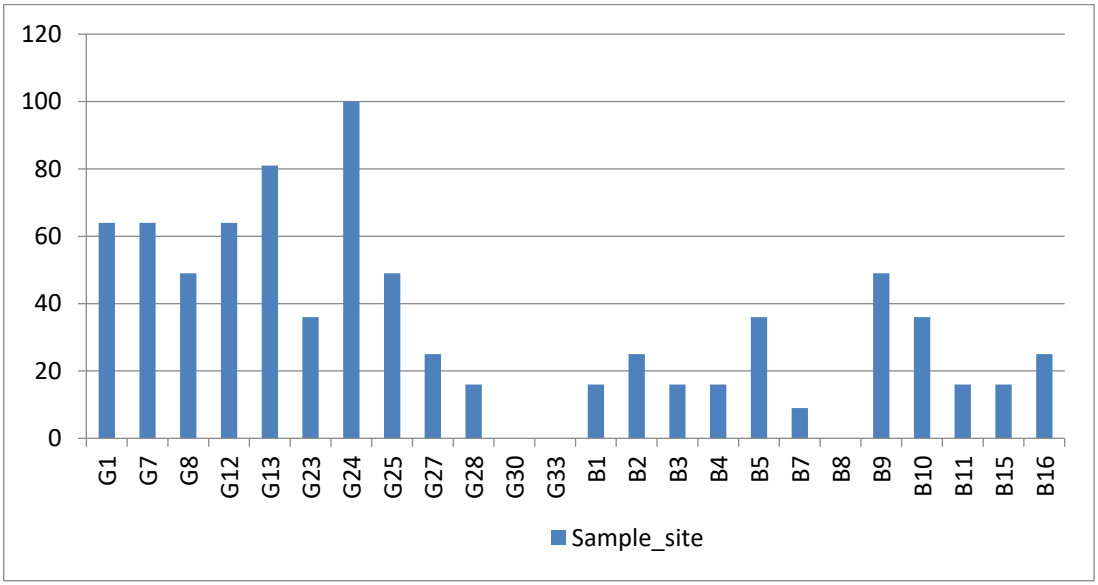


Figure 4.1: Viral disease incidence for the sampling sites in Gem and Matayos Sub-counties in Western Kenya.

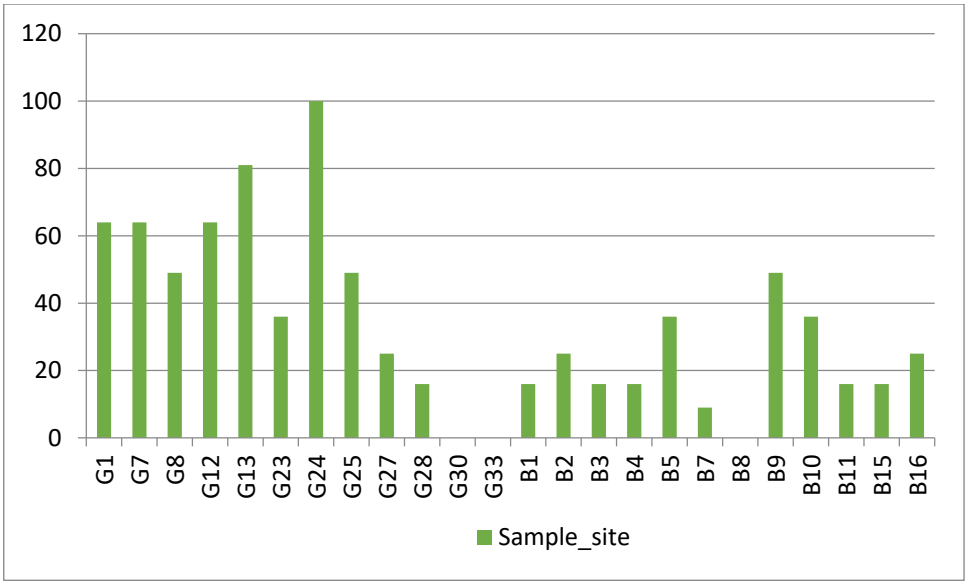


Figure 4.2: Viral disease severity for sampling sites in Gem and Matayos Sub-counties in Western Kenya.

4.1.2 Relative abundance, distribution and prevalence of groundnut RNA viruses

PCSV was the most prevalent virus across the two Sub-counties of Gem and Busia Counties. It was detected in 12 samples representing 50% of all the samples. CPPV 2 was the second most prevalent virus at 29% of total samples. GRV, satRNA and CPPV 1 were all present in 6 samples only amounting to 25% prevalence. PeMoV was the least prevalent because it was only present in one sample (**Table 4.3**).

A total of 6 different viruses were detected in the two Sub-counties, five of the detected viruses were present a cross the two regions except PeMoV which was only present in one sampling site in Matayos Sub-county, Busia County. CPPV 1 was equally present in 3/12 of the samples in Matayos Sub-county and 3/12 in Gem Sub-county; CPPV 2 presence was detected in 4/12 of the samples in Matayos Sub-county and 3/12 in Gem Sub-county; PCSV was the most abundant, present in 5/12 of the samples in Matayos Sub-county and 7/12 of samples in Gem Sub-county; GRV was found in one sample in 1/12 in Matayos Sub-county and 5/12 of those in Gem; and satRNA was detected in 2/12 samples in Matayos sub-county and 4/12 of the samples in Gem Sub-county (**Table 4 and Figure**

4.7). This showed that the five viruses are widely distributed across Matayos and Gem Sub-counties.

No virus was detected in seven samples. Four samples had a combination of 5 of the viruses. One of the samples analyzed had a combination of 3 viruses detected. Four samples again had a combination of two viruses each and lastly, eight samples had a single virus detected. PCSV was present in all combinations witnessed. Both CPPV 1 and CPPV 2 were present in combinations of 5, 3 and 2 and more than 90% found together. GRV and satRNA were in combinations of 5, 2, and 1and more than 50% present together. PeMoV was found singly and only in one sample (**Table 4** and **Figure 4.6**).

Table 4.2: The relative abundances of the six viruses identified in Gem and Matayos Sub-counties, Western Kenya

Gem Sub-county, Siaya County				**Matayos Sub-county, Busia County**			
Site ID	Detected virus	% Relative abundance	Relative abundance Log_{10}	Site ID	Detected virus	% Relative abundance	Relative abundance Log_{10}
G1	PCSV	0.006466%	-4.1894				
	satRNA	0.041129%	-3.3858	B1	-		-
G7	-	-	-	B2	PCSV	0.003162%	-4.5000
G8	PCSV	0.005204%	-4.2836	B3	-		-
G12	satRNA	0.124298%	-2.9055	B4	satRNA	0.001349%	-4.8697

	GRV	0.020843%	-3.6810				
	PCSV	0.014322%	-3.8440				
	satRNA	0.509198%	-2.2931				
G13	GRV	0.228001%	-2.6420	B5	PeMov	0.001235%	-4.9082
	CPPV1	0.029154%	-3.5353				
	CPPV2	0.877525%	-2.0567				
G23	GRV	0.003166%	-4.4995	B7	-	-	-
	PCSV	0.000553%	-5.2575		PCSV	0.007989%	-4.0975
	satRNA	0.009825%	-4.0074		satRNA	0.705992%	-2.1512
G24	GRV	0.020534%	-3.6875	B8	GRV	0.402601%	-2.3951
	CPPV1	0.000221%	-5.6554		CPPV1	0.006874%	-4.1628
	CPPV2	0.000677%	-5.1693		CPPV2	11.2829%	-1.9476
	PCSV	0.001387%	-4.8578				
	satRNA	0.062607%	-3.2034				
G25	GRV	0.033076%	-3.4804	B9	-		-
	CPPV1						

		0.001374%	-4.8619				
	CPPV2	0.003699%	-4.4319				
					PCSV	0.041859%	-3.3782
G27	PCSV	0.000741%	-7.1931	B10	CPPV1	0.499559%	-2.3014
					CPPV2	0.407578%	-2.3898
G28					CPPV1	0.010549%	-4.9768
	-		-	B11			
					CPPV2	0.033597%	-3.6810
G30	PCSV	0.000130%	-5.8846	B15	PCSV	0.002144%	-4.6687
G33	-		-	B16	PCSV	0.078664%	-3.1042
					CPPV2	0.058548%	-3.2325

Relative abundance = the log_{10} [clean reads mapped to the viral contig (s) divided by the total clean reads in each sample].

Table 4.3 Prevalence of viral species detected across Gem and Matayos Sub-counties, Western Kenya

Viral species detected on groundnut samples	**Viral Prevalence from Gem and Matayos Sub-counties (%)**
Pepper chlorotic spot virus (PCSV)	50
Cowpea polerovirus 1 (CPPV1)	25
Cowpea polerovirus 2 (CPPV2)	29
Groundnut rosette satellite RNA (SatRNA)	25
Groundnut rosette virus (GRV)	25
Peanut mottle virus (PeMoV)	04

Table 4.4: Viruses detected in groundnut leaf samples from Gem and Matayos sub-counties, Western Kenya

Sample ID	**Virus Species present**	**Numbers Contig (s)**	**Sampling site**
B1	-	-	Matayos Sub-county, Busia
B2	Pepper Chlorotic Spot virus	1	"
B3	-	-	"
B4	Groundnut RNA satellite virus	1	"
B5	Peanut Mottle virus	1	"
B7	-	-	Matayos Sub-county, Busia

			County
	Cowpea Polerovirus 2	4	
	Cowpea Polerovirus 1	3	
B8	Pepper Chlorotic Spot virus	1	“
	Groundnut rosette Virus	1	
	Groundnut RNA satellite virus	1	
B9	-		“
B10	Pepper Chlorotic Spot virus	1	
	Cowpea Polerovirus 1	3	“
	Cowpea Polerovirus 2	2	
B11	Cowpea Polero virus 1	2	
	Cowpea Polerovirus 2	8	“
B15	Pepper Chlorotic Spot virus	1	“
B16	Cowpea Polerovirus 2	7	
	Pepper Chlorotic Spot virus	2	“
G1	Groundnut RNA satellite virus	2	Gem Sub-county, Siaya County
	Pepper Chlorotic Spot virus	1	
G7	-	-	“
G8	Pepper Chlorotic Spot virus	1	“
G12	Groundnut RNA satellite virus	2	“
	Groundnut rosette virus	1	
	Groundnut RNA satellite virus	1	
	Cowpea Polerovirus 1	3	
G13	Cowpea Polerovirus 2	5	Gem Sub-county, Siaya County
	Groundnut rosette virus	1	

	Pepper Chlorotic Spot virus	1	
G23	Groundnut rosette virus	2	“
G24	Groundnut RNA satellite virus	2	
	Cowpea Polerovirus 1	1	
	Cowpea Polerovirus 2	3	“
	Groundnut rosette virus	3	
	Pepper Chlorotic Spot virus	1	
G25	Groundnut RNA satellite virus	3	
	Cowpea Polerovirus 1	2	
	Cowpea Polerovirus 2	5	“
	Groundnut rosette virus	5	
	Pepper Chlorotic Spot virus	1	
G27	Pepper Chlorotic Spot virus	1	“
G28	-	-	“
G30	Pepper Chlorotic Spot virus	1	“
G33	-	-	“

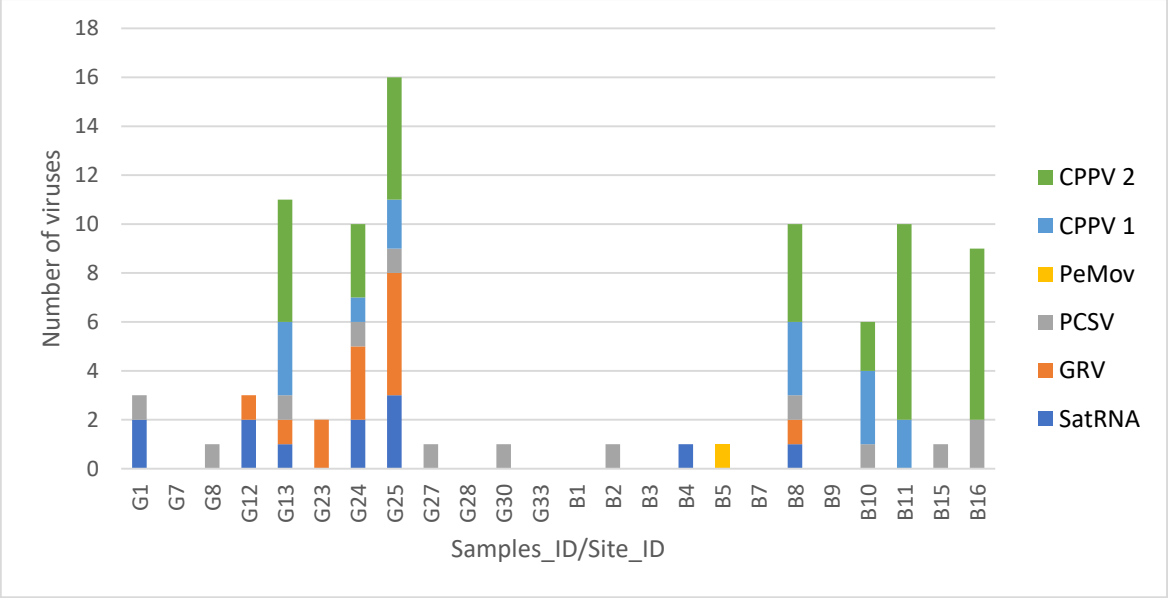
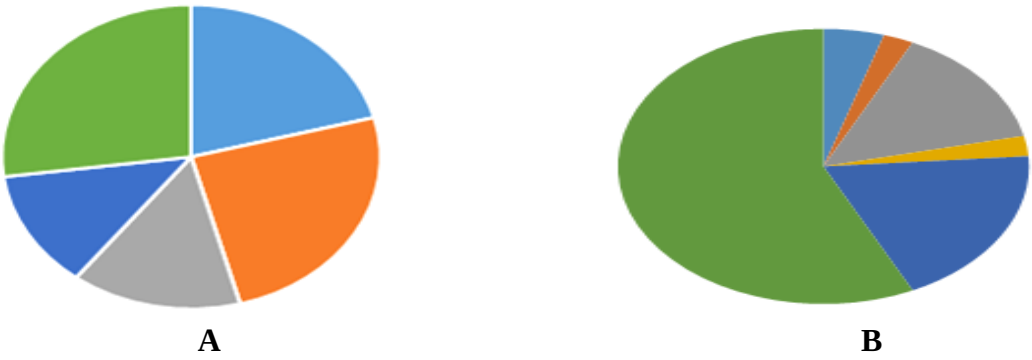


Figure 4.3: Composition of the six identified viruses for the individual samples across Gem and Matayos Sub-counties, Western Kenya.



KEY -viruses in the pie-charts: SatRNA GRV PCSV PeMov CPPV 1 CPPV 2

Figure 4.4: The distribution of the six identified viruses: Groundnut rosette satellite RNA virus (SatRNA), Groundnut rosette virus (GRV), Pepper chlorotic spot virus (PCSV), Peanut mottle virus (PeMoV), Cowpea polerovirus 1 (CPPV 1) and Cowpea polerovirus 2 (CPPV 2) (A) The pie chart for the distribution of the viruses in Gem Sub-county, Siaya County. (B) The pie chart for the distribution of the viruses in Matayos Sub-county, Busia County.

4.2 Simultaneous detection and identification of groundnut RNA viruses in Gem and Matayos Sub-counties, Western Kenya

4.2.1 Total RNA isolation, purification and quantification

The purified total RNA was obtained from groundnut leaf samples following extraction using Direct-zol MiniPrep kit (ZYMO Research), cooperating TRI Reagent Solution (Ambion, Life Technologies, CA, USA). Quantification of purified total RNA was done using agarose gel electrophoresis and the concentration measured with Qubit® RNA BR Assay Kit (Life Technologies, USA) (**Figures 4.5 and Figure 4.6)**.

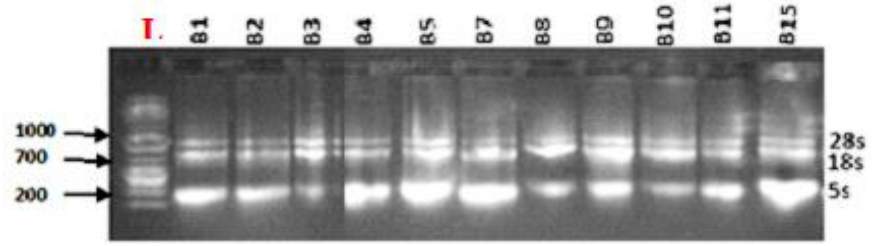


Figure 4.5: Total RNA gel electrophoresis image for groundnut leaf samples from Matayos Sub-county, Western Kenya coded with Bs; lane 1 is a 1kb DNA ladder (Thermo Fisher Scientific) followed by individual sample lanes as indicated.

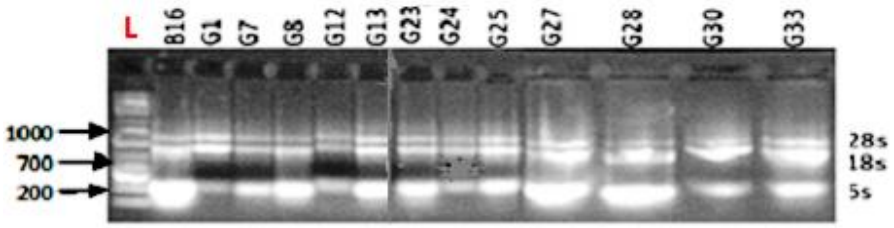


Figure 4.6: Total RNA gel electrophoresis image for groundnut leaf samples from Gem Sub-county, Western Kenya coded with Gs; lane 1 is a 1kb DNA ladder (Thermo Fisher Scientific) and lane 2 is for sample B16 while the rest of the lanes represent individual sample as indicated.

4.2.2 The pre-processing of the reads for the analyses of metagenomes

Before the commencement of the analysis process, the raw reads underwent several pre-processing stages prior to the actual analyses. These stages enhanced the efficiency of reads alignment and their assembly. To confirm the quality of reads for improvement following pre-processing, the assessment of the quality of the reads after pre-processing was done using FASTQC. A summary of the results obtained are provided in **Figure 4.7a and 4.7b** and **Table 4.5**.

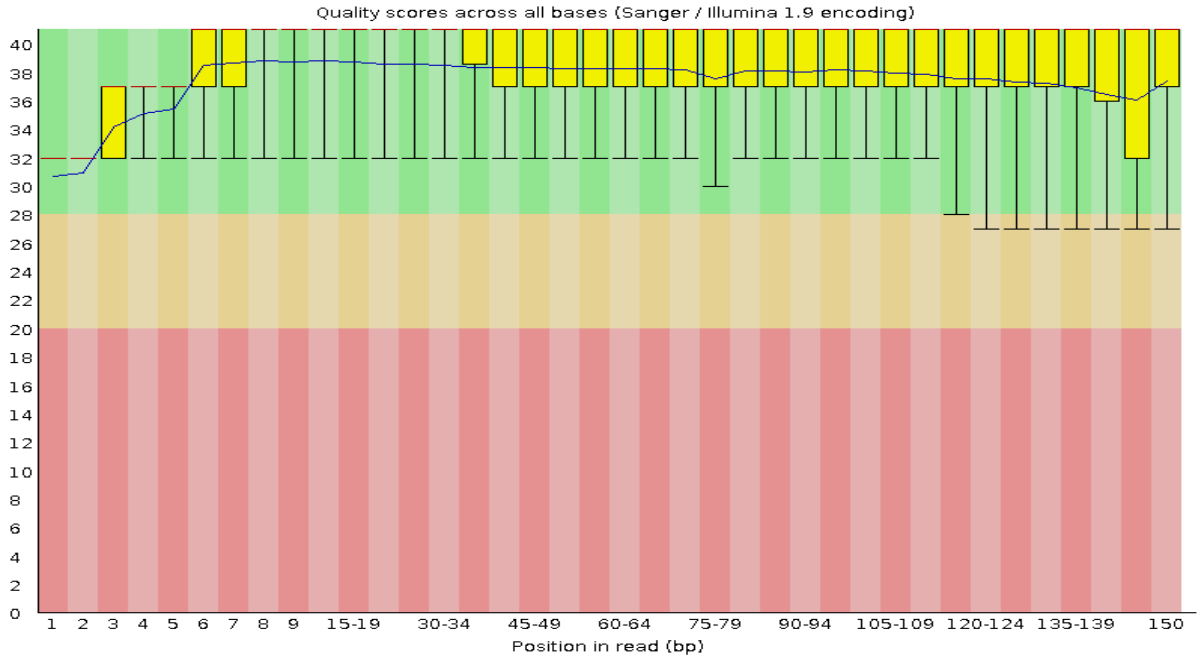


Figure 4.7a: FASTQC file for per base quality phred scores before pre-processing**.**

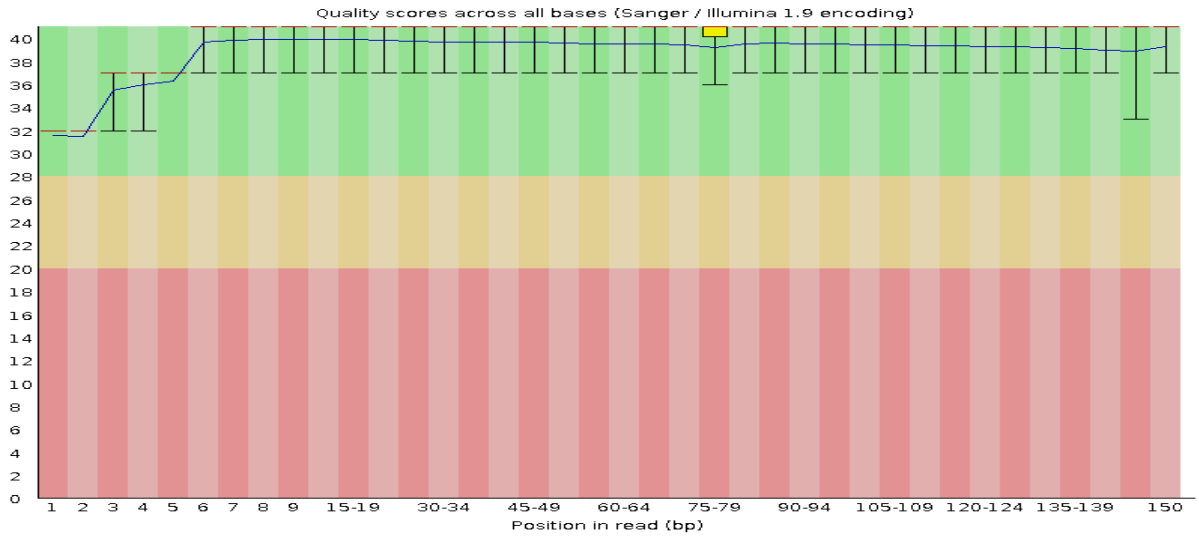


Figure 4.7b: FASTQC file showing the improved per base quality phred scores following pre-processing.

Table 4.5: Summary of the quality control of groundnut viral metagenomics dataset

Sample	Viral raw reads	Trimmed reads	Host-free reads	Error corrected reads
B1	8,709,082	7,659,238	3,104,972	3,037,524
B2	14,025,988	12,842,576	1,171,228	1,138,390
B3	10,901,674	9,719,754	542,266	513,654
B4	15,145,008	13,401,288	763,570	740,814
B5	9,195,594	8,336,288	1,167,580	1,133,384
B7	10,195,430	9,232,160	1,529,049	997,034
B8	16,674,828	15,051,646	2,185,994	2,152,998
B9	8,522,016	7,693,678	1,085,332	1,056,926

B10	5,707,668	4,637,402	371,410	363,120
B11	20,123,234	17,483,048	5,485,462	5,441,030
B15	13,838,288	11,956,266	2,416,092	2,353,552
B16	17,291,326	15,318,282	3,269,702	3,221,302
G1	5,585,458	4,830,708	1,438,472	1,391,908
G7	3,591,790	3,161,988	1,207,396	1,167,246
G8	8,284,306	7,272,324	828,172	768,500
G12	4,617,446	3,990,550	551,296	527,764
G13	29,074,936	25,127,114	4,423,918	4,328,766
G23	6,369,114	5,323,634	1,578,920	1,515,976
G24	18,399,678	15,424,836	14,812,090	14,473,454
G25	20,262,388	16,849,324	15,735,684	15,569,976
G27	11,028,814	9,635,140	3,048,854	2,970,226
G28	14,714,052	12,760,382	11,967,884	11,671,106
G30	12,393,254	10,608,252	6,285,670	6,132,304
G33	11,252,796	9,505,080	7,125,230	6,972,924
Total	295,904,168	257,820,898	92,096,243	89,639,878

4.2.3 Viral metagenomes identification using BlastN and Kraken Classifier

Total RNA from a total of 24 samples was used to construct paired-end reads which represented twenty four sampling sites in the two Sub-counties. A total of 295.9 million raw reads were obtained, which were reduced to 89.6 million reads after trimming, host subtraction and error correction (**Table 4.5**). The de-novo assembly of each individual sample gave a total of 42,591 contigs (**Table 4.6).** These contigs were used to determine the possible viruses present in the samples using BlastN against localized plant viral

database adopted from NCBI "nr" databases. Individual processed clean paired reads. To obtain better insight on groundnut viruses and their genetic variation in Gem Sub-county and Matayos Sub-county, Western Kenya, a metagenomics analysis was conducted based on next generation of total RNA sequencing, de-novo assembly, assessment of quality of assembly (**Table 4.7**) and groundnut viruses identification in Gem Sub-county and Matayos Sub-county, Western Kenya through bioinformatics (**Table 4.8**). The contigs were also subjected to taxonomic classification using Kraken classifier. Kraken classifier was able to discriminate RNA viruses from DNA viruses and classify the detected RNA viruses up to the species level and their visualization done using Krona (**Figure 4.8**). The results obtained indicated the presence of two of economically important groundnut viruses: Groundnut rosette virus (GRV) and Groundnut rosette satellite RNA virus (satRNA), (Groundnut rosette complex viruses) and Peanut mottle virus (PeMoV) (**Table 4.9**). Pepper Chlorotic spot virus (PCSV), Cowpea polerovirus 1 (CPPV 1) and Cowpea polerovirus 2 (CPPV 2) were detected for the first time in the two Sub-counties and on groundnut as a host (**Table 4.10**).

Table 4.6: A summary of the groundnut leaf samples reads assembled into contigs

Sample	Input Reads	Contigs	Largest contig	N50	Total assembly
B1	3,037,524	1,604	15,692	1,964	3,052,526
B2	1,138,390	11	5,048	2,067	23,082
B3	513,654	58	3,908	1,374	84,328
B4	740,814	19	2,912	2,088	33,156
B5	1,133,384	204	6,247	1,422	309,707
B7	997,034	21	18,934	8,192	78,025

B8	2,152,998	26	4,427	1,855	46,906
B9	1,056,926	1,543	11,031	1,694	2,601,832
B10	363,120	6	5,394	1,271	10,866
B11	5,441,030	1,009	5,184	1,306	1,343,442
B15	2,353,552	3,328	145,458	8,941	17,515,313
B16	3,221,302	1,666	9,459	1,456	2,500,548
G1	1,391,908	723	325,943	80,577	6,373,378
G7	1,167,246	692	14,670	1,470	1,127,792
G8	768,500	19	5,134	1,171	27,098
G12	527,764	13	3,250	1,690	21,555
G13	4,328,766	1,960	186,196	13,361	12,542,691
G23	1,515,976	2,245	14,400	1,520	3,548,247
G24	14,473,454	4,526	827,062	65,242	31,318,411
G25	1,569,976	3,567	754,216	63,900	27,334,871
G27	2,970,226	7,710	26,222	2,108	15,623,242
G28	11,671,106	3,774	147,949	5,360	10,949,433
G30	6,132,304	5,522	111,174	3,085	6,118,173
G33	6,972,924	2,345	402,663	91,460	2,637,150

Table 4.7: Misassemblies report from MetaQUAST comparing the results of the two reads assembly software

	MegaHIT	**MetaSPAdes**
# misassemblies	0	0
#contig misassemblies	0	0

# c. relocations	0	0
# c. translocations	0	0
# c. inversions	0	0
#c. interspecies translocations	0	0
#scaffold misassemblies	0	0
# s. relocations	0	0
# s. translocations	0	0
# s. inversions	0	0
# s. interspecies translocations	0	0
# misassembled contigs	0	0
Misassembled contigs length	0	0
# possibly misassembled contigs	1	1
# possible misassemblies	1	1
# local misassemblies	0	0
# Scaffold gap ext. mis.	0	0
# Scaffold gap loc. mis.	0	0
# Unaligned mis. Contigs	1	1
# mismatches	60	86
# indels	0	10
# indels (<= 5 bp)	0	9
# indels (> 5 bp)	0	1
Indels length	0	22

All statistics are based on contigs of size >= 1000 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

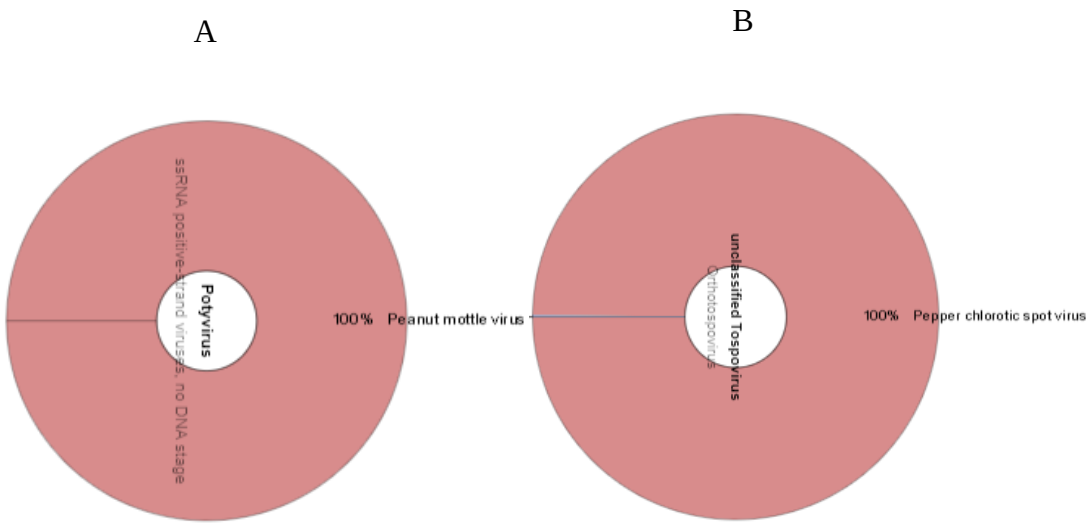
Table 4.8: Contig (s) details for the viruses identified

Sample ID	Viral contig (s)	Viral species	Av. length (nt)	Best Matched (NCBI - Accession)	% Identity	E-value
B1	-	-	-	-	-	-
B2	1	PCSV	386	NC_033774	96	2.45e-24
B3	-	-	-	-	-	-
B4	1	satRNA	308	NC_002738	90	2.83e-117
B5	1	PeMoV	339	NC_002600	98	1.36e-165
B7	-	-	-	-	-	-
B8	4	CPPV2	646	NC_034247	88	0.0
	3	CPPV1	686	NC_034246	88	0.0
	1	PCSV	512	NC_033774	98	1.18e-23
	1	GRV	4139	MG646923	86	0.0
	1	satRNA	980	NC_002738	90	0.0
B9	-	-	-	-	-	-
B10	1	PCSV	379	NC_033774	100	5-21e-21
	3	CPPV1	681	NC_034246	88	1.18e-24
	2	CPPV2	2777	NC_034247	84	0.0
B11	2	CPPV1	1231	NC_034246	88	2.57e-11
	8	CPPV2	948	NC_034247	87	2.3e-17
B15	1	PCSV	520	NC_033774	98	1.20e-23
B16	7	CPPV2	461	NC_034247	92	4.01e-94
	2	PCSV	674	NC_033774	96	3.38e-25

G1	2	satRNA	761	NC_002738	90	0.0
	1	PCSV	610	NC_033774	98	1.42e-23
G7	-	-	-	-	-	-
G8	1	PCSV	365	NC_033774	98	8.32e-24
G12	2	satRNA	804	NC_002738	89	0.0
	1	GRV	1608	MG646923	86	0.0
G13	1	satRNA	1124	NC_002738	90	0.0
	3	CPPV1	977	NC_034246	90	5.86e-46
	5	CPPV2	2102	NC_034247	88	0.0
	1	GRV	3976	MG646923	87	0.0
	1	PCSV	371	NC_033774	100	5.10e-21
G23	2	GRV	347	MG646923	87	8.095e-99
	2	satRNA	719	NC_002738	89	1.31e-14
	1	CPPV1	407	NC_034246	96	2.32e-104
G24	3	CPPV2	525	NC_034247	90	1.76e-10
	3	GRV	1455	MG646923	87	0.0
	1	PCSV	789	NC_033774	96	3.97e-25
	3	satRNA	701	NC_002738	93	5.36e-70
	2	CPPV1	988	NC_034246	86	0.0
G25	5	CPPV2	858	NC_034247	93	1.00e-32
	5	GRV	1208	MG646923	86	1.53e-31
	1	PCSV	373	NC_033774	96	1.83e-25
G27	1	PCSV	764	NC_033774	96	3.84e-25
G28	-	-	-	-	-	-

G30	1	PCSV	306	NC_033774	98	3.2e-22
G33	-	-	-	-	-	-

GRV=Groundnut rosette virus; satRNA=Groundnut rosette satellite RNA virus; PeMoV=Peanut mottle virus; PCSV= Pepper chlorotic spot virus; CPPV 1=Cowpea polerovirus1; CPPV 2=Cowpea polerovirus 2



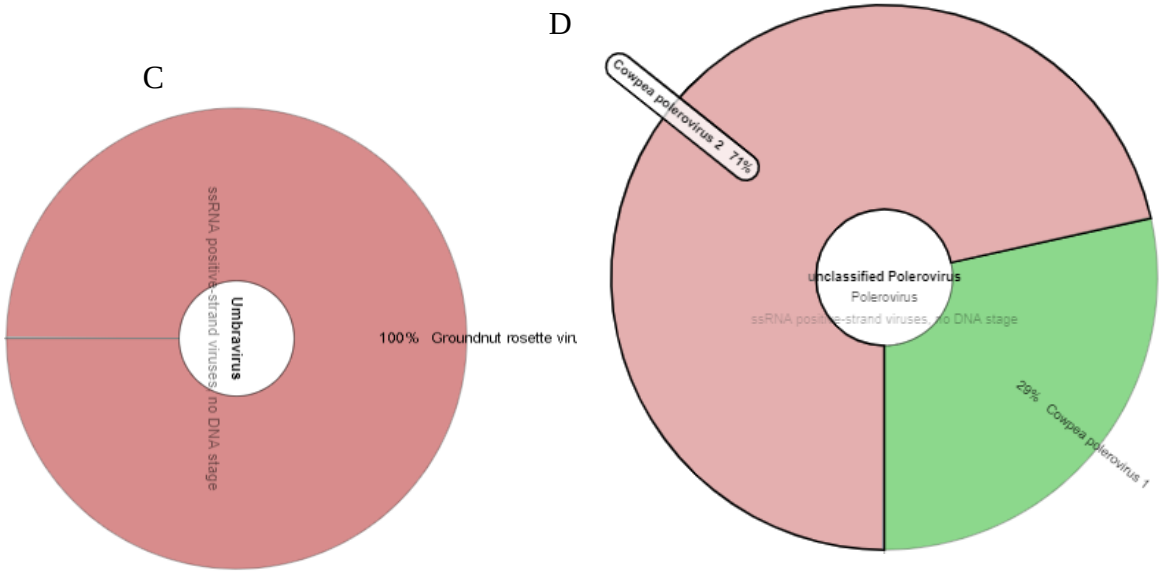


Figure 0.8: Taxonomic classification of the identified RNA viruses namely, (A) Peanut mottle virus (PeMoV), (B) Pepper chlorotic spot virus (PCSV), (C) Groundnut rosette virus (GRV), (D) Cowpea polerovirus 1 (CPPV1) and Cowpea polerovirus 2 (CPPV2) using Kraken and their visualization by Krona.

Table 4.9: Detection of major economically important groundnut RNA viruses

Sample	**Virus**								
				Satellite					
	CMV	GBNV	GRV	RNA	PCV	PeMoV	PStV	TSWV	IPCV
B1	-	-	-	-	-	-	-	-	-
B2	-	-	-	-	-	-	-	-	-
B3	-	-	-	-	-	-	-	-	-

B4	-	-	-	+	-	-	-	-	-
B5	-	-	-	-	-	+	-	-	-
B7	-	-	-	-	-	-	-	-	-
B8	-	-	+	+	-	-	-	-	-
B9	-	-	-	-	-	-	-	-	-
B10	-	-	-	-	-	-	-	-	-
B11	-	-	-	-	-	-	-	-	-
B15	-	-	-	-	-	-	-	-	-
B16	-	-	-	-	-	-	-	-	-
G1	-	-	-	+	-	-	-	-	-
G7	-	-	-	-	-	-	-	-	-
G8	-	-	-	-	-	-	-	-	-
G12	-	-	+	+	-	-	-	-	-
G13	-	-	+	+	-	-	-	-	-
G23	-	-	+	-	-	-	-	-	-
G24	-	-	+	+	-	-	-	-	-
G25	-	-	+	+	-	-	-	-	-
G27	-	-	-	-	-	-	-	-	-
G28	-	-	-	-	-	-	-	-	-
G30	-	-	-	-	-	-	-	-	-
G33	-	-	-	-	-	-	-	-	-

Reference: (+) Presence of viral species and (-) absence of viral species.

CMV=Cucumber mosaic virus; GBNV=Groundnut bud necrosis virus; GRV=Groundnut rosette virus; satRNA=Groundnut rosette satellite RNA virus; PCV=Peanut clump virus;

PeMoV=Peanut mottle virus; PStV=Peanut stripe virus; TSWV=Tomato spotted wilt virus; IPCV=Indian peanut clump virus.

Table 4.10: First time reported RNA viruses on groundnut in Gem and Matayos Sub-counties

Sample	**Virus**								
	PCSV		CPPV1		CPPV2				
B1	-	-	-	-	-	-	-	-	-
B2	-	-	-	-	+	-	-	-	-
B3	-	-	-	-	-	-	-	-	-
B4	+	-	-		-	-	-	-	-
B5	-	-	+	-	-	-	-	-	-
B7	-	-	-	-	-	-	-	-	-
B8	+	-	+		+	-	-	-	-
B9	-	-	-	-	-	-	-	-	-
B10	-	-	+	-	+	-	-	-	-
B11	-	-	-	-	+	-	-	-	-
B15	+	-	-	-	-	-	-	-	-
B16	+	-	-	-	+	-	-	-	-
G1	+	-	-		-	-	-	-	-
G7	-	-	-	-	-	-	-	-	-
G8	+	-	-	-	-	-	-	-	-

G12	-	-	+		-	-	-	-	-
G13	+	-	+		+	-	-	-	-
G23	-	-	-	-	-	-	-	-	-
G24	+	-	+	-	+	-	-	-	-
G25	-	-	+	-	+	-	-	-	-
G27	+	-	-	-	-	-	-	-	-
G28	-	-	-	-	-	-	-	-	-
G30	+	-	-	-	-	-	-	-	-
G33	-	-	-	-	-	-	-	-	-

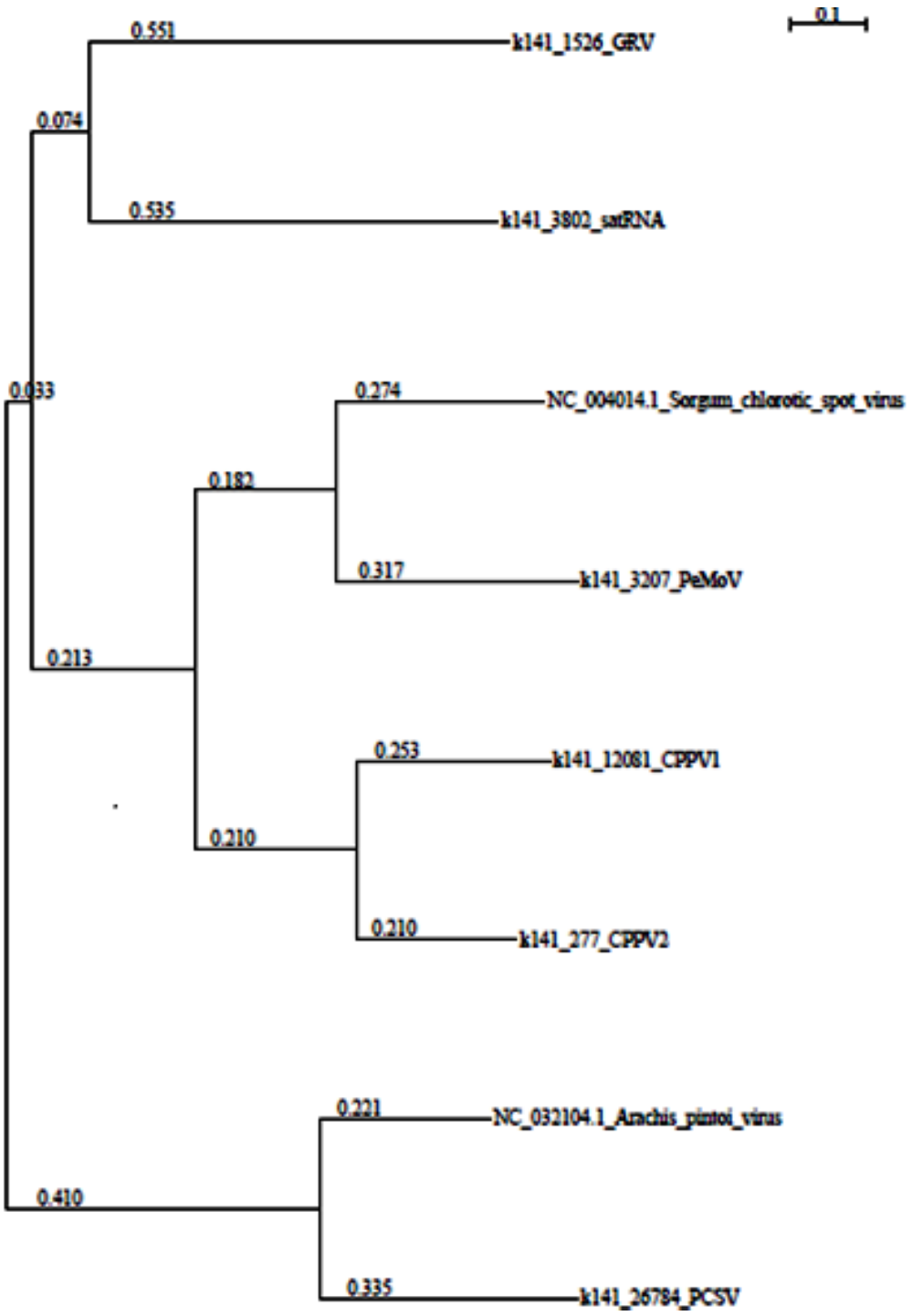
Reference: (+) Presence of viral species and (-) absence of viral species.

PCSV=Pepper chlorotic spot virus; CPPV 1=Cowpea polerovirus 1; CPPV 2=Cowpea polerovirus 2.

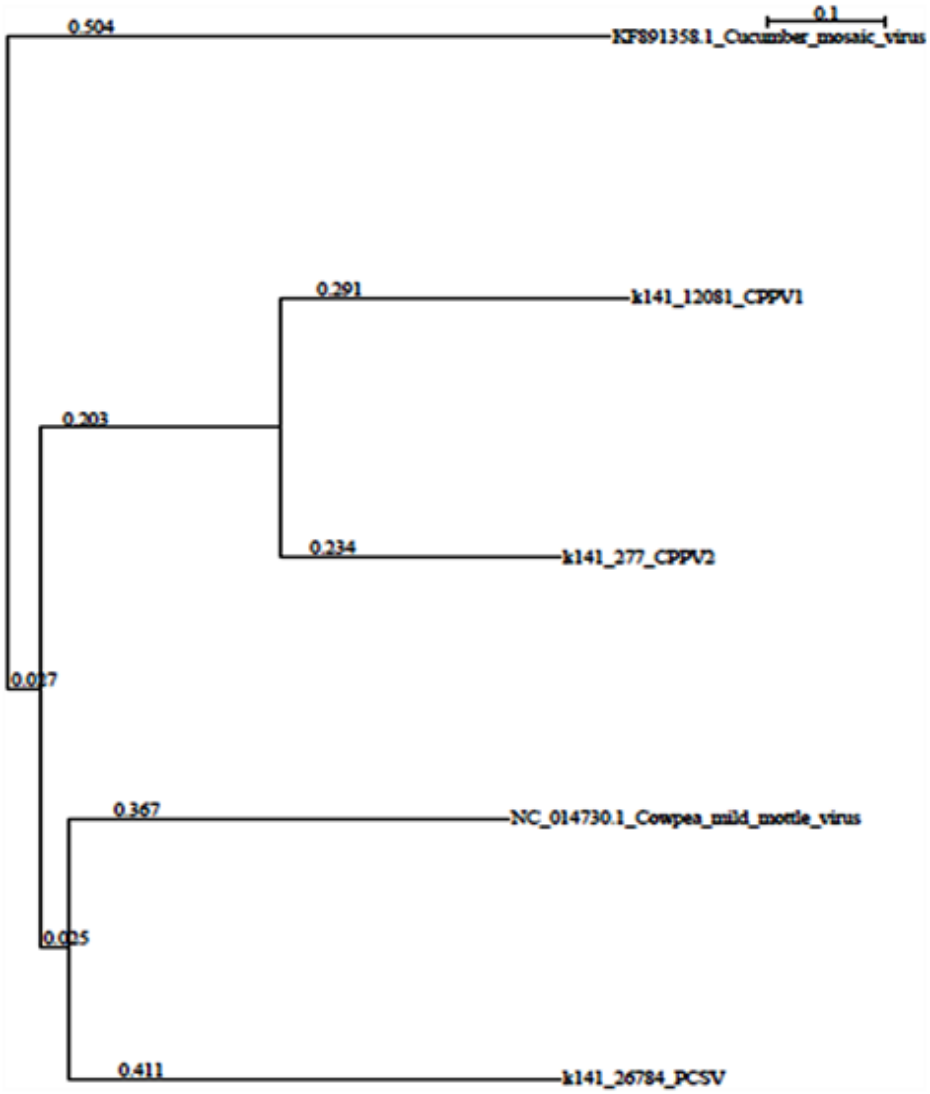
4.2.4 Phylogenetic analysis of the identified viruses on groundnut leaf samples across Gem and Matayos Sub-couties, Western Kenya

To determine the evolutionary relationships among the six viruses identifies on groundnut leaf samples from across Gem Sub-county and Matayos Sub-county, Western Kenyan the best recovered contigs for each of the viruses k141_3802_satRNA (Groundnut rosette satellite RNA), k141_1526_GRV(Groundnut rosette virus), k141_3207_PeMoV (Peanut mottle virus), k141_26784_PCSV (Pepper chlorotic spot virus), k141_277_CPPV2 (Cowpea polerovirus 2), k141_12081_CPPV1 (Cowpea polerovirus 1) and two out-groups; Arachis pintoi virus and Sorghum chlorotic spot virus (NCBI Refseq Accession Numbers NC_032104.1 and NC_004014.1 respectively) (**Figure 4.9A**). Additionally, to establish the relationship among Cowpea poleroviruses detected and Pepper chlorotic spot virus and with known cowpea viruses; Cucumber mosaic virus and Cowpea mild mottle virus (NCBI

Refseq Accession Numbers KF_891358.1 and NC_014730.1 respectively) were used to construct a phylogenetic tree by pooling representative contigs across the two Sub-counties (**Figure 4.9B**). The analysis divided all the six identified viral contigs and the out-groups into two distantly cluster (**Figure 4.9A**). The newly reported Cowpea poleroviruses and Pepper chlorotic spot virus fell under three distinct clusters; Cucumber mosaic virus clustered alone, the two Cowpea poleroviruses clustered on their own too while Pepper chlorotic spot virus and Cowpea mild mottle virus clustered together (**Figure 4.9B**).


0.1
0.551
k141_1526_GRV
0.074
0.535
k141_3802_satRNA
0.033
0.274
NC_004014.1_Sorgum_chlorotic_spot_virus
0.182
0.317
k141_3207_PeMoV
0.213
0.253
k141_12081_CPPV1
0.210
0.210
k141_277_CPPV2
0.221
NC_032104.1_Arachis_pintoi_virus
0.410
0.335
k141_26784_PCSV

A



B

Figure 4.9: Phylogenetic analysis of six RNA viruses identified in Gem and Matayos Sub-counties, Western Kenya. The phylogenetic tree diagrams were based on the best recovery of each of the viral nucleotide sequences. The evolutionary history was inferred by using the Maximum-likelihood method based on K2 substitution model with 1,000 bootstraps in SeaView version 4 using PhyML in-built module. (A) Phylogenetic tree of the six identified viruses and the two out-groups: **Arachis pintoi virus** and **Sorghum chlorotic spot virus** and (B) a phylogenetic tree relating Cowpea polerovirus 1 (**CPPV1**), Cowpea polerovirus 2 (**CPPV2**) and Pepper chlorotic spot virus with **Cucumber mosaic virus** and **Cowpea mild mottle virus**; known cowpea infecting viruses.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1 Discussion

Viruses are known as a major biotic constraint in groundnut production and cause loss of quality and quantity of produce. But only a few of the host-pathogen causing the losses have been investigated in Western Kenya. Moreover, not much has been reported on none pathogenic interactions because they have few references in this region and hence unknown at the moment.

The study of groundnut's virome could be essential in the attempt to understand groundnut diseases dynamics and the diversity of these viruses. There is possibility of drastic changes in virome composition, which is not only reinforced by the infidelity of the RNA polymerase during replication but also by the changing climatic conditions in plant habitat. This could be a source of selective pressure, a fodder for viral emergence. In addition, the ubiquitous presence of viruses have enabled them to establish several complex relationships with different organisms and host ranges either pathogenic or otherwise (Roossinck, 2015), and not limited to groundnut. There is little knowledge of the visible impact of some of these relationships with the host.

Presently, only a few studies have been done pertaining to groundnut viruses but most of these lie outside the wider scope of this study. The majority of these studies have reported on a single investigation of individual groundnut viruses without any focusing on a platform to simultaneously tackle the entire groundnut virome. Further emphasis is that most of these studies have been biased towards groundnut rosette complex viruses (Naidu et al., 1999; Wangai et al., 2001; Waliyar et al., 2007; Kayondo et al., 2014). Although,

several studies in the recent years have reported success with NGS for discoveries of viruses on other crops (Al Rwahnih et al., 2009; Coetzee et al., 2010; Ng, 2010; Al Rwahnih et al., 2011; Li et al., 2012; Maina et al., 2016; Mollov and Malapi-Wight, 2016), but only one for groundnut (Gutiérrez et al., 2016).

A comprehensive analysis of virome diversity is a challenge because the mundane methods of virome detection such as culturing often fail to spot a wide range of virus types present in the sample. This is analog to bacterial cultivation where only a small percentage of the population can be cultivated. This limits the use of culturing as a standard protocol for isolation and identification of diverse organisms.

Therefore, most components of the existing and ever-changing virome are assumed to be untenable through cultivation standard protocols. Consequently, the available genomic data on viruses is suggestively biased towards predominantly culturable viruses, and little genetic diversity. Wider virome analyses are also limited by technology bottlenecks and were only able to afford new lease of life with the plunging cost of NGS and high throughput data generation platforms (Rose et al., 2016). Coupled with the aforementioned developments, metagenomics is enabling the possibility of utilizing a universal platform for a real-time, simultaneous and rapid identification of entire virome community diversity without the prior knowledge of the sequence information for priming and even circumvent the cultivation step altogether.

However, due to the lack of reported investigations on groundnut metagenomics analysis, it was worthwhile developing procedures for simultaneous detection and characterization of the groundnut virome composition at any instance. The findings of this thesis result

from unbiased amplification and identification of viral genomes of environmental samples. A similar protocol was used by Sachsenröder et al., (2012) in their investigation of pig virome. Hence the findings of this study can be optimized following the recommendation of studies within this scope for achieving routine use in groundnut viruses' discovery.

The detection groundnut RNA viruses results (**Table 4.4**), established a correlation with the severity and incidence data that was established the time of at the time of sample collection, with the findings obtained by analysis of reads corresponding to different viral genomes. The study found out that there was a higher incidence and severity of viral symptoms in Gem Sub-county in Siaya County as compared to Matayos Sub-county in Busia County. This correlates well with the analysis of reads, where nine (9) samples out of twelve from Gem sub-county, Siaya, County were positive for at least one of the six (6) identified viruses in this study as compared to eight (8) positive samples out of twelve in Matayos Sub-county, Busia County (**Table 4.4** and **Figure 4.3**). The findings of this study demonstrated the reliability of this technique in representing the most probable viral disease prevalence by identifying the specific viral species present in a given sample (**Table 4.3**). There is a marked difference in disease prevalence at the start of the study which provided generalization of viral disease presence but at the time of reporting the findings of this study, specific samples with viral genomes were established. One of the samples under asymptomatic category, sample G30 was positive for viral genome presence while the remaining two had no viral genome detected or were of lower titer hence could not be detected. This study investigated both symptomatic and asymptomatic samples. All the asymptomatic and one symptomatic sample were negative for RNA viruses. The symptomatology witnessed earlier on virus negative samples can be suggestive of other reasons other than RNA viruses presence but does not fully exclude all viruses in totality,

given that viruses have very small genomes, which can be overwhelmed or domination by host redundant rRNA and other contaminants such as bacteria in very low concentrations hence failed detection.

Moreover, the general prevalence of viral disease calculated at the start of the study was 91.67% in Matayos and 83.33% in Gem and overall of 87.5% in the two locations. This is in contrast with the reported findings of 66.67% in Matayos and 75% in Gem and an overall prevalence of 70% in the end –based on actual viral sequences detected in each sample/site. The difference in the prevalence of groundnut viruses' in these areas can be due to variations in weather conditions that occurred in these areas. Gem Sub-county was drier during the groundnut growing period with an observed common infestation of Aphids major vectors to the groundnut viruses were highly noticeable. The influence of weather conditions on the distribution of aphids was reported earlier by Naidu et al., (1999).

The detection and identification of Peanut mottle virus (PeMoV) in one of the samples confirmed its worldwide distribution and the presence of at least one of the Groundnut rosette complex viruses in eight (8)samples also affirms their being endemic to sub-Saharan Africa all asserting the findings by Naidu et al., (1999).

The presence of five (5) of the six viruses detected across the two locations, Gem sub-county and Matoyos sub-county, may indicate viral vagility, which may be in relation to vector (s), aphids population at any given time. Pepper chlorotic spot virus (PCSV) was the most prevalent of the viruses detected across the two locations. PCSV was first reported in Taiwan in 2009 in a study involving sweet pepper as a host, followed by a second reporting in chili pepper as the next host which was detected in 2017 in mainland China (Chen et al., 2017). The latest reporting was its discovery on *Biden pilosa*, a weed in

Yunnan, China in 2019 (Huang et al., 2019). Having been reported in more than three occasions so far, PCSV also closely clusters with Cowpea mild mottle virus (CPMMV) (**Figure 4.9B**), and was reported as an emerged problem in soybean and a re-emerged problem in common beans (Zanardo and Carvalho, 2017). Peanut mottle virus (PeMoV) in the phylogenetic tree diagram (**Figure 4.9A**) is another member of the same genus *Potyvirus* which Biden mottle virus belongs and also shared Biden pilosa as host with PCSV though they do not group together. This therefore, confirms PCSV as the newly emerging virus with its new discovery in groundnut.

The two poleroviruses and Groundnut rosette complex viruses had almost similar prevalence (**Table 4.3**). The presence of CPPV1 and CPPV2 together in 6 out of 7 samples that were either positive for one or both of them and the clustering two viruses together on the phylogenetic tree (**Figure 4.9A and B**) branches may allude to a mixed infection of the two viruses. This corroborates the findings of Palanga et al., (2017). These two viruses were first reported collectively after their discovery using metagenomics-based screening (Palanga et al., 2016). To the best of my knowledge, this is the second finding of these viruses outside Burkina Faso and in another host other than cowpea. Similarly, the two Groundnut rosette complex viruses detected in this study occurred together on many occasions. This is in agreement with the findings by Naidu and Kimmins, (2007) that the two viruses frequently occur as a pair being part of the three complex viruses causing Groundnut rosette disease (GRD). The most probable justification for a possible co-infection or mixed infections of viruses in various samples in this study can be attributed to the fact that most of the viruses belonging to the families involved in this study are transmitted by aphids (NG and Perry, 2004).

The appearance of GRV and the two poleroviruses on the same major branch show a close relationship among them as compared to other major viruses infecting groundnut. This can be confirmed by the fact that they all belong to the family *Luteoviridae* (Palanga et al., 2016). Members of family *Luteoviridae* have been reported as having a worldwide distribution and are known to infect the two main classes of angiosperm (Palanga et al., 2017).

This PeMoV was identified and classified using Kraken taxonomic classifier. Only one sample was positive for PeMoV virus and with a single contig recovered. The contig of 339 nt with a 98% identity to the NC_ 002600 accession in the NCBI represented a partial PeMoV genome recovery (**Tables 4.4, 4.8 and 4.9**). Sanchez' et al., (2016) in their investigation of forage peanut obtained a larger genome of up to 9707 nt. It is noteworthy to note that the most abundant virus in their samples was Peanut mottle virus though only one sample was positive in this case.

PCSV was the most prevalent virus in the present study. It was detected either singly or in the presence of other viruses. The best recovery of the genome was about 789 nt with up to 100% identity to NC_033774 accession. The Kraken classification and confirmation by NCBI BLAST classified PCSV as a member of family *Peribunyaviridae*, genus *Tospovirus,* and still unclassified Tospovirus (**Tables 4.4, 4.8 and 4.9**).

CPPV 1 and CPPV 2 were classified in the family *Luteoviridae* and the genus *Polerovirus,* though still appearing as unclassified poleroviruses. Like other members of *Polerovirus* genus, they are characterized by isometric virions of about 25 to 30 nm in diameter of a positive single-stranded RNA genome of up to 6.0 kb. These detailed findings were reported by Gutierrez et al., (2016) who had similar findings on *Poleroviruses* though not

exact viruses as in this present study. The largest viral contig recovered was of CPPV2, which was about 5394 nt and the average identity was about 89%. CPPV1 1164 nt was the largest fragment obtained from assembling the groundnut reads with an average identity of 89% to NC_034246 accession. **Table 4.4** shows the average of all the viruses in these two species.

GRV and its satellite RNA work together to signify Groundnut rosette disease GRD symptoms manifestations in most cases but do not work as pair exclusively. In some instances a third virus, Groundnut rosette assistor virus, (GRAV) may be involved and in most cases leading to increase in the intensity of the symptomatology of GRD (Naidu and Kimmins, 2007). The largest GRV fragment obtains through assembly in the present study was 4139 nt and its average identity was 87% to NCBI accession number MG646923 (**Table 4.4).** Like the poleroviruses, GRV also belongs to the family *Luteoviridae* and but genus *Umbravirus.* GRV is a positive single-stranded virus with no DNA stage. Sat-RNA is a small linear single-stranded RNA satellite fragment and with a broader grouping as Satellite Nucleic Acids (**Figure 4.8**). The best recovered assembled contig for satRNA was about 1124 nt with an average identity of 90% giving a reliable similarity to the NCBI accession NC_002738. **Table 4.4** shows average size of all the satRNA each sample and other relevant details.

The overall number of paired-end sequenced reads in this study ranged from 3 million ~ 29 million (**Table 4.5**). Depending on the objective (s) of the study and the design, the number of reads that might be recommended varies as long as the diagnosis is achieved. This study provides an experience of using as low as 3 million reads to sufficiently and reliably detect viruses. Though, there was a slump in the number of reads after the entire process of reads

quality control (**Table 4.5**). This can be traced to poor quality reads that were determined by the quality phred scores indicated by FASTQC and hence removal by the quality and adapter removal software (**Figure 4.7a and b**). In addition, there was an overwhelming number of host reads which was clearly indicated after host-subtraction step (**Table 4.5**). Similar procedures were followed and results akin to these findings obtained were reported in a review by Jones et al., (2017).

The assembly stage was conducted by both megaHIT and metaSPAdes but the results obtained by megaHIT took precedence because they showed longer contigs with a larger N50 (**Table 4.7**) and were relevant in downstream function annotation analysis (Quince et al., n.d.). Finally, the finding of the study demonstrated capacity to simultaneously detect and characterize viruses present in samples collected by discriminating between RNA viral strains that infected groundnut samples. In addition, the advantage of this technique and procedures is that they offer an additional element of information above the presence and absence of pathogen synonymous with traditional routine methodologies for virus detection and identification. The availability of sequence and sequence information complement rather than replace the routine technique used currently.

5.2 Conclusion

The following can be inferred from the findings of this study.

a) Metagenomics enabled efficient analysis of the viral relative abundances, composition, distribution and prevalence of groundnut RNA viruses.
b) The detection of groundnut RNA viruses from different groundnut leaf samples with a few intermediary upstream steps involved offers a basis for developing an optimized protocol for groundnut RNA viral diagnostics.

c) Through the use of viral metagenomics, all samples were tested and confirmed for presence or absence for virus by obtaining the actual viral sequence, i.e. both the asymptomatic and the symptomatic samples.

d) Two major economically important groundnut RNA viruses, i.e. Groundnut rosette virus (GRV) and Peanut mottle virus (PeMoV) were found in nine of the samples tested. With eight (8) samples testing positive for GRV and only one sample being positive for PeMov.

e) This study established a possible co-infection or a mixed infection involving the CPPV1 and CPPV2 virus. This was also the first reporting of these viruses outside Burkina Faso and on groundnut as a host.

f) PCSV was the most prevalent virus in Gem and Matayos Sub-counties in Western Kenya.

g) The findings of this thesis strongly suggest that the use of viral metagenomics was still reliable 3 million reads or less for the identification of viruses from both asymptomatic and symptomatic samples.

5.3 Significance of the findings

The findings of the present study envisages a profound premise and promise for routine generic diagnostic procedures for groundnut RNA viruses' detection, groundnut phytosanitation during quarantine, certification and groundnut seed breeding system.

5.4 Recommendations

1) Further study elucidating the specific types of relationships and correlations among all the various viruses infecting groundnut and the impact of these relationships on the manifestation of the symptoms on the host plant should be conducted.

2) Establishing a comprehensive outlook of the groundnut viruses' diversity by undertaking an extensive metagenomics study on the larger Western region can follow this study.
3) In depth study to understanding the host-virus interactions and their influence on the viruses' population, types of strains, identification virus-resistant genes and change trajectory due to these interactions in a multifaceted infection spectrum will be provide a better understanding than what is possible currently.
4) Studying the effects of the relative abundances, composition and distribution of other groundnut viral groups, co-evolution and the effect of their overall interactions impacting on groundnut production, food security and the agro-ecosystem can also be appropriate as a follow-up study.

REFERENCES

Abate, T., Alene, A. D., Bergvinson, D., Shiferaw, B., Silim, S., Orr, A., & Asfaw, S. (2012). Tropical Grain Legumes in Africa and South Asia: Knowledge and Opportunities. International Crops Research Institute for the Semi- Arid Tropics.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology*, *10*(4), 537–545.

Al Rwahnih, M., Daubert, S., Golino, D., & Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology*, *387*(2), 395–401.

Al Rwahnih, M., Dolja, V. V, Daubert, S., Koonin, E. V, & Rowhani, A. (2011). Genomic and biological analysis of Grapevine leafroll-associated virus 7 reveals a possible new genus within the family Closteroviridae. *163*, 302–309.

Alemu, K. (2015). Detection of diseases, identification and diversity of viruses : A review. *5*(1), 204–214.

Altschul, S. F., Gish, W., Pennsylvania, T., & Park, U. (1990). Basic Local Alignment Search Tool. *Mol. Biol*, *215*, 403–410.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, *19*(10), 535–544.

Argonne, G., Meyer, F., National, A., & Meyer, F. (2016). Analysis of metagenomics dat*a*.

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., & Palù, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology, pp 7861–7884.

Ben Langmead, & Salzberg, S. L. (2013). Bowtie2. *Nature Methods*, *9*(4), 357–359.

Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F. M., M., S. C., Leal-Bertioli, & Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. *Plant Genetic Resources:Characterization and Utilization*, *9*(1), 1341–149.

Bock, K. R. (1973). *Peanut mottle virus in East Africa*. 171–179.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Booker, R. H. (1963). The effect of sowing date and spacing on rosette disease of groundnut in Northern Nigeria , with observations on the. *Annals of Applied Biology*, 125–131.

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J., & Mumford, R. (2014). Methods in virus diagnostics: From ELISA to next generation sequencing. *Virus Research*, *186*, 20–31.

Breuil, S. De, Nievas, M. S., Giolitti, F. J., Fitopatología, I. De, Giorda, L. M., Experimental, E., Lenardon, S. L. (2008). Special report occurrence, prevalence, and distribution of viruses infecting peanut in Argentina. 1237–1240.

Canavar, Ö., & Kaynak, M. A. (2016). Effect of Different Planting Dates on Yield and Yield Components of Peanut (Arachis hypogaea L .) Effect of Different Planting Dates on Yield and Yield Components of. *Turkish Journal of Agric*, *32*, 521–528.

Castrignano, S. B., & Nagasse-sugahara, T. K. (2015). The metagenomic approach and causality in virology Abordagem metagenômica e causalidade em virologia. 10–14.

CEFA. (2011). Good agronomic practices for Groundnut in Western Kenya Good agronomic practices for Groundnut in Western Kenya Training manual for Trainers.

Chen, K. Z., Chen, T., Yeh, S.-D., Rahman, M. S., Su, X., Wu, K., Dpng, J. (2017). Characterization of a new isolate of pepper chlorotic spot virus from Yunnan province

, China. *Archives of Virology*.

Chirwa, M., Mrema, J. P., Mtakwa, P. W., Kaaya, A., & Lungu, O. I. (2017). Yield response of Groundnut (*Arachis hypogaea* L .) to Boron, Calcium, Nitrogen, Phosphorus and Potassium Fertilizer Application. *International Journal of Soil Science*, *12*, 18–24.

Coetzee, B., Freeborough, M. J., Maree, H. J., Celton, J. M., Rees, D. J. G., & Burger, J. T. (2010). Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. *Virology*, *400*(2), 157–163.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-cabrero, D., Cervera, A., Mcpherson, A., … Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. 1–19.

Costa, V., Angelini, C., Feis, I. De, & Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. 2010.

Coulibaly, M. A., Ntare, B. R., Gracen, V. E., & Danquah, E. (2017). Groundnut production constraints and farmers' preferred varieties in Niger. *4*(1).

Culbreath, A. K., Todd, J. ., & Brown, S. L. (2003). Epidemiology and management of tomato spotted wilt in p eanut. *Annual Review of Phytopathology*, *41*(1), 53–75.

D'Arcy, C. J., Torrance, L., & Martin, R. . (1989). Discrimination among *Luteoviruses* and their strains by monoclonal antibodies and identification of common epitopes . *The American Phytopathological Society*, pp 5.

de Breuil, S., Giolitti, F. J., Bejerman, N., & Lenardon, S. L. (2012). Effects of Cucumber mosaic virus on the yield and yield components of peanut. *Journal of Plant Pathology*, *94*(3), 669–673.

Delcher, A. L. (2006). Glimmer Release Notes Version 3 . 02. 1–18.

Demski, J. W. (1975). Source and spread of Peanut mottle virus in soybean and peanut.

Desmae, H., & Sones, K. (2017). Groundnut cropping guide.

Devi, Shoba P., Maraite, H., & Reddy, D. V. R. (2002). Dynamics of *Polymyxa graminis* and Indian peanut clump virus (IPCV) infection on various monocotyledonous crops and groundnut during the rainy season. 546–560.

Dietzgen, R. G., Callaghan, B., Higgins, C. M., Birch, R. G., Chen, K. R., & Xu, Z. Y. (2001). Differentiation of peanut seedborne potyviruses and cucumoviruses by RT-PCR. *Plant Disease*, *85*(9), 989–992.

Directorate Plant Production, Department of Agriculture, F. and F. S. A. (2010). Groundnuts production guideline.

Domola, M. J., Thompson, G. J., Aveling, T. A. S., & Laurie, S. M. (2008). Sweet potato viruses in South Africa and the effect of viral infection on storage root yield.

Dorak, M. T. (2006). Real-time *PCR*. Taylor & Francis Group.

Duhaime, M. B., & Sullivan, M. B. (2012). Ocean viruses : Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology*, *434*(2), 181–186.

Edgar, R. C., Drive, R. M., & Valley, M. (2004). MUSCLE : multiple sequence alignment with high accuracy and high throughput. *32*(5), 1792–1797.

Food and Agriculture Organization (2003). FAOSTAT data.

Fang, Y., & Ramasamy, R. P. (2015). Current and prospective methods for plant disease detection. *Biosensors*, *5*(3), 537–561.

Farrell, J. A. K. (1976). Effects of groundnut sowing date and plant spacing on rosette virus disease in Malaw*i*. (1964), 159–171.

Ferguson, M., Bramel, P., & Chandra, S. (2004). Gene diversity among botanical varieties in peanut (*Arachis hypogaea* L.). *Crop Sciences*, *44*(5), 1847–1854.

Gascuel, O., & Gouy, M. (2017). SeaView Version 4 : A multiplatform graphical user interface for sequence alignment and phylogenetic tree building.

Getachew, G., Alemu, T., & Tesfaye, K. (2014). Evaluation of disease incidence and

severity and yield loss of finger millet varieties and mycelial growth inhibition of Pyricularia grisea isolates using biological antagonists and fungicides in vitro condition , pp5883–5901.

Ghanekar A.M., D. V. R. Reddy, N. Iizukat, P. W. A. R. W. G. (1979). Tomato spotted wilt virus on tomato. (49), 583–584.

Ghanekar, A. M. (1980). Groundnut Virus Research at ICRISAT, pp 211–216.

Green, S. K. (1971). Guidelines for Diagnostic Work in Plant Virology, (15).

Guindon, S., Dufayard, J., & Lefort, V. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3 . 0. *Systematic Biology*, *3*(59), 307–321.

Gutiérrez Sánchez, P. A., Jaramillo Mesa, H., & Marin Montoya, M. (2016). Next generation sequence analysis of the forage peanut (Arachis pintoi) virome. *Revista facultad Nacional de Agronomía*, *69*(2), 7881–7891.

Hall, R. J., Wang, J., Todd, A. K., Bissielo, A. B., Yen, S., Strydom, H., Peacey, M. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*, *195*, 194–204.

Hasegawa, M. (1985). Dating of the human-ape splitting by molecular clock of mitochodrial DNA. *Journal of Molecular Evolution*, 160–174.

Hebert W.Virgin, H. (2015). The virome in mammalian physiology and Disease. *Cell*,*157*(1), 142–150.

Huang, C. J., Zeng, J. M., Yu, H. Q., & Liu, Y. (2019). First Report of Pepper Chlorotic Spot Virus in Bidens pilosa in Yunnan , China. *Plant Disease*, 1–4.

IITA (1971). *Methods for the Diagnosis of Plant Virus Diseases Laboratory Manual* (P. Lava Kumar, Ed.).

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION :

delivery of nanopore sequencing to the genomics community. *Genome Biology*, 1–11.

Jo Vandesompele (2009). *qPCR guide*.

Jones, S., Baizan-edge, A., Macfarlane, S., & Torrance, L. (2017). Viral diagnostics in plants using next generation sequencing : computational analysis in practice.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. *30*(14), 3059–3066.

Kayondo, S. I., Rubaihayo, P. R., Ntare, B. R., Gibson, P. T., Edema, R., Ozimati, A., & Okello, D. K. (2014). Genetics of resistance to groundnut rosette virus disease. *African Crop Science Journal*, *22*(1), 21–30.

Kenchanagoudar, P. V, Naragund, V. B., & Naik, M. K. (2005). Management of peanut bud necrosis disease through intercropping. *58*(2), 207–211.

Koech et al., 2007. (2007). Full length research article production efficiency and economic potential of different soil fertility management strategies among groundnut farmers of kenya. *World*, *2*(1), 15–21.

Kokalis-Burelle, Porter D.M., Rodriguez-Kabana, R., Smith D.H., & Subrahnyam P., (1997). The Disease compendium series of the American Phytopathological Society. *The American Phytopathological Society.*

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

Krapovickasi, A., & Gregory, W. C. (1994). Taxonomia del Genero Arachis (Leguminosae). *Bonplandia*, *8*, 1–186.

Krausslich, H.G. (2009). Viral Strategies, pp 1–22.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., & Simon, R.

(2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*, *388*(1), 1–7.

Kuhn, C. W., Demski, J. W., Reddy, D. V. R., Bennep, C. P., & Bijaisoradaf, M. (1973). identification and incidence of peanut viruses in Georgia, pp 67–69.

Laing, M. D., Daudi, H., Shimelis, H., Laing, M., Okori, P., Mponda, O., Mponda, O. (2018). Groundnut production constraints , farming systems , and farmer-preferred traits in Tanzania.

Le, S. Q., & Gascuel, O. (2001). An improved general amino acid replacement matrix.

Li, D., Liu, C.M., Luo, R., Sadakane, K., & Lam, T.W. (2016). MEGAHIT: An ultra-fast single-node solution for large and com- plex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *32*(22), 3498–3500.

Li, R., Gao, S., Hernandez, A. G., Wechter, W. P., Fei, Z., & Ling, K. S. (2012). Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS ONE*, *7*(5), 1–10.

Lim, E. S., & Gumpil, J. S. (1984). The flowering, pollination and hybridization of groundnuts *(Arachis hypogaea L .)*. *7*(2), 61–66.

Liu, S., Vijayendran, D., & Bonning, B. C. (2011). Next generation sequencing technologies for insect virus discovery. *Viruses*, *3*(10), 1849–1869.

Maina, S., Edwards, O. R., J, B. M., de Almeida, L., Ximenes, A., & Roger A C Jones. (2016). Deep Sequencing reveals the complete genome sequence of sweet potato virus G from East Timor. *4*(5), 957–16.

Mari Kamitani, Atsushi J Nagano, Mie N.Honjo, H. K. (2016). *RNA-S*eq reveals virus-virus and virus-plant interactions in nature. 1–27.

Marston, D. A., Mcelhinney, L. M., Ellis, R. J., Horton, D. L., Wise, E. L., Leech, S. L.,

Fooks, A. R. (2013). Next generation sequencing of viral RNA genomes. *BMC Genomics*, *14*(1), 1.

Masira, I. (2017). Agriculture and food authority Busia county production and market research.

Mathews, D. M., & Ph, D. (2010). Optimizing detection and management of virus diseases of plants. 1–7.

Matranga, C. B., Gladden-young, A., Qu, J., Winnicki, S., Nosamiefan, D., Levin, J. Z., & Sabeti, P. C. (2016). Unbiased deep sequencing of RNA viruses from clinical samples. *Journal of Visualized Experiments*, 9.

Mcclean, P. (2004). Viral Genome Organization : A General Overview..

Mikheenko, A., Saveliev, V., & Gurevich, A. (2016). MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics*, *32*(7), 1088–1090.

Misari, S., Abraham, J., Demski, J., O.A, A., C.W, K., R., C., & E., B. (1988). Aphid transmission of the viruses causing chlorotic rosette and green rosette disease of peanut in Nigeri*a*. *The American Phytopathological Society*, pp 4.

MOA (2004a). Annual report 2004, Nyanza Province. Kisumu, Kenya.

MOA (2004b). Annual report 2004, Western Province. Kakamega, Kenya.

MoALF. (2016). Climate risk profile for Siaya. Kenya County climate risk profile series. The Kenya Ministry of Agriculture, Livestock and Fisheries (MoALF), Nairobi, Kenya. 1–20.

Mollov, D., & Malapi-Wight, M. (2016). Next Generation Sequencing: a useful tool for detection of sugarcane viruses in quarantine programs. *Proceedings of the International Society of Sugar Cane Technologists 29*, *29*, 1631–1635.

Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C. M., Valls, J. F. M., & Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (Arachis

hypogaea) and its most closely related wild species using intron sequences and microsatellite markers. *Annals of Botany*, *111*(1), 113–126.

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *24*(16), 1757–1764.

Mukoye B., Mangeni.B.C., Leitich.R.K., Wosula.D.W., Omayio.O.D., Nyamwamu.P.A., Arinaitwe.W., Winter.S., A. M. . and W. K. (2015). First report and biological characterization of Cowpea mild mottle virus (Cpmmv) infecting groundnuts in Western Kenya, pp 0–5.

Naidu, R. A., Bottenberg, H., Subrahmanyam, P., Kimmis, F., Robinson, D., & JM Thresh. (1998). Epidemiology of groundnut rosette virus disease : Current status and future research needs. *Journal of Applied Biology*, 23.

Naidu, R. A., & Kimmins, F. M. (2007). The effect of Groundnut rosette assistor virus on the agronomic performance of four groundnut *(Arachis hypogaea L .)* genotypes. *356*, 350–356.

Naidu R.A, Kimmins, F.M, Deom, C.M, Subrahmanyam, P, Chiyembekeza, A.J, van der Merwe, P. J. . (1999). Groundnut Rosette: A virus Disease Affecting Groundnut Production in Sub-Saharan Africa. *Plant Disease*, *83*(8), 700–709.

Naim, A. M. El, Eldouma, M. A., Ibrahim, E. A., & Zaied, M. B. (2011). Influence of plant spacing and weeds on growth and yield of peanut (*Arachis hypogaea*. L) in rain-fed of Sudan. *1*(2), 45–48.

Ndunguru, J., Kapinga, R., Sseruwagi, P., Sayi, B., Mwanga, R., Rugutu, C., Salaam, D. (2009). Assessing the sweetpotato virus disease and its associated vectors in northwestern Tanzania and central Uganda. *4*, 334–343.

Nduru, G., & Bein, F. L. (2011). Survey of traditional farming practices in Western Kenya, pp 2–4.

NG, J. C., & Perry, K. L. (2004). Transmission of plant viruses by aphid vectors. *Molecular Plant Pathology*, *5*, 505–511.

Ng, T. F. (2010). Discovery of Novel Viruses From Animals , Plants , and Insect Vectors Using Viral Metagenomics by. *Discovery*.

Nigam, S., Giri, D., & Reddy, A. G. (2013). Groundnut seed production manual. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Nontajak, S., Vulyasevi, S., Jonglaekha, N., Program, B., Royal, T., Foundation, P., & Mai, C. (2014). Effect of mixed viruses infection on symptom expression in Zucchini (Cucurbita pepo. *10*(5), 1329–1339.

Ntare, B., & Olorunju, P. (2001). Variation in yield and resistance to groundnut rosette disease in early and medium-maturing groundnut genotypes in Nigeria. *African Crop Science Journal*, *9*(2), 451–461.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res.*, *1*(27), 30–47.

Okello, D.K, Biruma, M, and Deom, C. (2010). Overview of groundnuts research in Uganda: Past, present and future. *African Journal of Biotechnology*, *9*(20), 2843–2850. 3

Okello, D. ., Akello, L. B., Tukamuhabwa, P., Odong, T. L., Ochwo-Ssemakula, M., Adriko, J., & Deom, C. M. (2014). Groundnut rosette disease symptoms types distribution and management of the disease in Uganda. *African Journal of Plant Science*, *8*(3), 153–163.

Okello, D. K., Monyo, E., C.M., D., Ininda, J., & Oloka, H. K. (2013). Groundnut production guide for Uganda : Recommended practices for farmers.

Olorunju, P. ., Kuhn, C. W., Demski, J. W., Misari, S. ., & O.A, A. (1991). Disease reactions and yield performance of peanut genotypes grown under Groundnut rosette

and rosette-free field environments. *The American Phytopathological Society*, pp 5.

Onyuka, E. O. (2016). An assessment of profitability of groundnut production using gross margin, the case of Ndhiwa Sub-county, Kenya. *3*(3), 116–123.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., Mcinerney, P., Milos, P. M. (2009). Direct RNA sequencing. *Nature*, *461*(7265), 814–818.

Palacios, G., Ph, D., Druce, J., Ph, D., Du, L., Ph, D., Paddock, C. D. (2008). *New England journal*. 991–998.

Palanga, E., Filloux, D., Martin, D. P., Fernandez, E., Bouda, Z., Gargani, D., Roumagnac, P. (2016). Metagenomic-based screening and molecular characterization of cowpea-infecting viruses in Burkina Faso. 1–21.

Palanga, E., Martin, D. P., Galz, S., Zabre, J., Bouda, Z., Neya, J. B., Fillloux, D. (2017). Complete genome sequences of cowpea polerovirus 1 and cowpea polerovirus 2 infecting cowpea plants in Burkina Faso. 1–4.

Pearson, M. N., & Wei, T. (2007). Microarrays for the detection and identification of plant viruses : a feasibility study using potyviruses. 149–154.

Plain, C., Station, E., Carolina, N., So, V., & States, U. (1988). Tomato spotted wilt virus on peanuts.

Popgeorgiev, N., Boyer, M., Fancello, L., Monteil, S., Robert, C., Rivet, R., Desnues, C. (2013). Marseillevirus-like virus recovered from blood donated by asymptomatic humans, *208*.

Quail, M. A., Quail, M., Smith, M., Coupl, P., Otto, T. D., Harris, S., Coupland, P. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 341.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (n.d.). Shotgun metagenomics , from sampling to sequencing and analysis.

Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., & Candela, M. (2016). ViromeScan : a new tool for metagenomic viral community profiling. *BMC Genomics*, 1–9.

Renaut, J., Planchon, S., Przybylska, A., Barylski, J., & Palukaitis, P. (2015). Effect of temperature on the pathogenesis , accumulation of viral and satellite RNAs and on plant proteome in peanut stunt virus and satellite RNA-infected plants, pp 1–14.

Robledo, G., & Seijo, G. (2008). Characterization of the Arachis (Leguminosae) D genome using fluorescence in situ hybridization (FISH) chromosome markers and total genome DNA hybridization, *724*, 717–724.

Robledo, G., & Seijo, G. (2010). Species relationships among the wild B genome of Arachis species (section *Arachis*) based on FISH mapping of rDNA loci and heterochromatin detection : a new proposal for genome arrangement, pp1033–1046.

Roossinck, M. J. (2012). Plant Virus Metagenomics: Biodiversity and Ecology. *Annu. Rev. Genet*, *46*, 359–369.

Roossinck, M. J. (2015a). Plants, viruses and the environment: Ecology and mutualism. *Virology*, *479–480*, 271–277.

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015b). Plant virus metagenomics: Advances in virus discovery. *Phytopathology*.

Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L., & Prosperi, M. (2016). Challenges in the analysis of viral metagenomes. *2*(2), 1–11.

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter : mining viral signal from microbial genomic data. 1–20.

Sachsenröder, J., Twardziok, S., Hammerl, J. A., Janczyk, P., Wrede, P., Hertwig, S., & Johne, R. (2012). Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS ONE*, *7*(4).

Sastry, K. S., & Zitter, T. A. (2014). Plant virus and viroid diseases in the tropics: Volume 2: Epidemiology and management. In plant virus and viroid diseases in the tropics: Volume 2: Epidemiology and management.

Schlaberg, R. (2016). Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics : a systematic comparison to a commercial PCR panel. *54*(4), 1000–1007.

Seijo, G., Lavia, G. I., Fernandez, A., Krapovickas, A., Ducasse, D. A., Bertioli, D. J., & Mosconet, E. A. (2007). Genomic relationships between the cultivated peanut (*Arachis hypogaea*, *Leguminosae*) and its close relatives revealed by double GISH. *American Journal of Botany*, *94*(12), 1963–1971.

Seijo, G., Lavia, G. I., Fernandez, A., Krapovickas, A., Ducasse, D., & Moscone, E. A. (2004). Physical mapping of the 5S and 18S–25S RRNA genes by FISH as evidence that *Arachis duranesis* and *A. Ipaensis* are the wild diploid progenitors of *A. hypogaea (Leguminosae)*. *91*(9), 1294–1303.

Settaluri, V. S., Kandala, C. V. K., Puppala, N., & Sundaram, J. (2012). Peanuts and their nutritional aspects-a review. *Food and Nutrition Sciences*, 1644–1650.

Singh, A. K., & Nigam, S. N. (2016). Gene pool diversity and crop improvement .

Sogut, T., Ozturk, F., & Kizil, S. (2016). Effect of sowing time on peanut (*Arachis hypogaea* L .) cultivars : I . yield, yield components, oil and protein. *Scientific Papers. Series A. Agronomy*, *LIX*, 415–420.

Soueidan, H., Schmitt, L. A., Candresse, T., & Nikolski, M. (2015). Finding and identifying the viral needle in the metagenomic haystack: Trends and challenges. *Frontiers in Microbiology*, pp 1–7.

Sreenivasulu, P, Subba Reddy, CH.V, Ramesh, B, Lava Kumar, P. (2008). *Virus Diseases of Groundnut*.

Sreenivasulu, P and Demski, J. . (1988). Transmission of Peanut Mottle and Peanut Stripe Viruses by Apis craccivora and Myzus persicae. *Plant Disease*, *72*(8), 722–723.

Sreenivasulu, P., Kuhn, C. W., Naidu, H. A., Demski, J. W., Reddy, V. R., & Nayudu, M. V. (1991). Viruses Infecting Peanuts *(Arachishypogaea).* taxonomy, identification, and disease management.

Storey, H. H., & Ryland, A. K. (1957). Viruses causing rosette and other diseases in groundnuts.

Subrahmanyam, P., Wyk, P. S. Van, Kisyombe, C. T., Cole, D. L., Hildebrand, G. L., Chiyembekeza, A. J., & Merwe, P. J. A. Van Der. (1997). International Journal of Pest Management Diseases of groundnut in the Southern African Development Community (SADC) region and their management, pp 37–41.

Suttle, C.A. (2017). Nutrients and Other Environmental Factors Influence Marine Environments. 1–15.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6 : Molecular evolutionary genetics analysis version 6 . *0*. *30*(12), 2725–2729.

Thekisoe, O. M. M., Rambritch, N. E., Nakao, R., Bazie, R. S., Mbati, P., Namangala, B., … Inoue, N. (2009). Loop-mediated isothermal amplification (LAMP) assays for detection of Theileria parva infections targeting the PIM and p150 genes. *International journal for parasitology*.

Thottappilly, G., & Rossel, H. W. (1997). Identification and characterization of viruses infecting bambara groundnut (Vigna subterranea) in Nigeria. *International Journal of Pest Management*, *43*(3), 177–185.

Uritskiy, G. V, DiRuggiero, J., & Taylor, J. (2018). MetaWRAP - a flexible pipeline for genome-resolved metagenomic data analysis. *BioRxiv*.

Valls, J. F. M., & Simpson, C. E. (2005). New Species of Arachis (Leguminosae) from

Brazil, Paraguay and Bolivia. *Bonplandia*, *14*, 35–63.

Waliyar., F., Kumar, P., B.R., N., E., M., Nigam, S., A.S, R., Diallo. (2007). A Century of Research on Groundnut Rosette Disease and its Management. *Information Bulletin*, (75).

Waliyar, F., Kumar, P. L., Ntare, B., Monyo, E., Nigam, S., Reddy, A., Diallo, A. (2007). A Century of Research on Groundnut Rosette Disease and its Management. *International Crops Research Institute for Semi-Arid Tropics, Informatio*, 44.

Wangai, A. W., Pappu, S. S., Pappu, H. R., Deom, C. M., and Naidu, R. A. (2001). Distribution and characteristics of groundnut rosette disease in Kenya. *Plant Dis.*, *85*(5), 470–474.

Webster, C. G., Frantz, G., Reitz, S. R., Funderburk, J. E., Mellinger, H. C., Mcavoy, E., Adkins, S. (2015). Emergence of Groundnut ringspot virus and Tomato chlorotic spot virus in vegetables in Florida and the Southeastern United States. 105, 388–398.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3).

Wright, D. L., Tillman, B., Small, I. M., Ferrell, J. A., & Dufault, N. (2016). Management and cultural practices for peanuts 1. 1–10.

Wu, Q., Ding, S.W., Zhang, Y., & Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Annual Review of Phytopathology*, *53*(1).

Zanardo, L. G., & Carvalho, C. M. (2017). Cowpea mild mottle virus *(Carlavirus, Betaflexiviridae)*: a review.

Zeng, R., Liao, Q., Feng, J., Li, D., & Chen, J. (2007). Synergy between Cucumber Mosaic virus and Zucchini Yellow Mosaic Virus on Cucurbitaceae hosts tested by real-time reverse transcription-Polymerase Chain Reaction. 39–42.

Zerbino, D. R., & Birney, E. (2008). Velvet *:* Algorithms for de novo short read assembly using de Bruijn graphs. 821–829.

https://www.biostars.org/p/89123, [accessed 27 August 2018].

https://en.climate-data.org/location/11165-8/, [accessed 18 June 2017].

APPENDICES

Appendix 1: A summarized details of sample collected

Sample ID	Sampling Location	Groundnut variety type	GPS coordinate Latitude	GPS coordinate Longitude
B1	Matayos Sub-county, Busia	Local	0.443611	34.1213889
B2	“	Hybrid	0.4441667	34.1194444
B3	“	Hybrid	0.4433333	34.1194444
B4	“	Local	0.4441667	34.12
B5	“	Hybrid	0.4605556	34.1141667
B7	“	Hybrid	0.4133333	34.1077778
B8	Matayos Sub-county, Busia	Hybrid	0.4133333	34.1075
B9	“	Hybrid	0.4133333	34.1080556
B10	“	Hybrid	0.4566667	34.1086111
B11	“	Hybrid	0.43	34.1083333
B15	“	Hybrid		34.13

			0.4088889	
B16	“	Hybrid	0.415	34.1302778
G1	Gem Sub-county, Siaya	Hybrid	0.0508333	34.4072222
G7	“	Hybrid	0.113889	34.4352778
G8	“	Hybrid	0.0366667	34.4661111
G12	“	Hybrid	0.0194444	34.5113889
G13	“	Hybrid	0.0011111	34.5
G23	“	Hybrid	0.1694444	34.3405556
G24	Gem Sub-county, Siaya	Hybrid	0.0341667	34.3238889
G25	“	Hybrid	0.016944	34.4094444
G27	“	Hybrid	0.066667	34.3344444
G28	“	Hybrid	0.0505556	34.4113889
G30	“	Local	0.0230556	34.4808333
G33	Gem Sub-county, Siaya	Hybrid	0.0191667	34.4808333

Appendix 2: Useful and frequently used commands in the analysis

Trimmomatic command for trimming raw reads

java -jar <path to trimmomatic.jar> PE [-threads <threads] [-phred33 | -phred64] <input 1><input 2><paired output 1><unpaired output 1><paired output 2><unpaired output 2>ILLUMINACLIP: <fastaWithAdaptersEtc>:<seed mismatches>: <palindrome clip threshold>:<simple clip threshold> LEADING:<quality> TRAILING:<quality> SLIDINGWINDOW: <window Size>: <requiredQuality> HEADCROP: <length> MINLEN: <length>

Similarity search command for BLAST software

/path/to/makeblastdb -in nr -dbtype nucl -out

~/path/to/blastp -db ~/path/to/nr -query /path/to/query.fa -out /path/to/results -evalue 0.00001 -outfmt 6 max_target_seqs 1

Reads mapping using bowtie2-commands

~/bowtie2-build <fasta.file><idxprefix>

~/bowtie2 [options]* -x <bt2-idx> {-1 <m1> -2 <m2>} –S [<sam>]