Testing for Host Adaptive Evolution Using the *Maize streak virus* Model

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the

Department of Integrative BioMedical Sciences

at the

UNIVERSITY OF CAPE TOWN

November 2021

Author:

Kehinde Adewole Oyeniran

Supervisor:

A/Prof. Darren Martin

To the memory of my late twin brother Mr Taiwo Adebowale Oyeniran

Declaration

I declare that the work presented in this thesis is my original work. I also state that this work has not been presented in this, or any other university for any degree.

KEHINDE ADEWOLE OYENIRAN

Signed by candidate

NOVEMBER 2021

Abstract

Maize streak virus (MSV; Genus: *Mastrevirus*; Family: *Geminiviridae*) causes maize streak disease (MSD); a major biotic threat to maize farming especially in sub-Saharan Africa, and it neighbouring Indian and Atlantic Ocean Islands, where its insect vectors in the genus *Cicadulina* thrive. Of the eleven known MSV strains (called A through K), only MSV-A is economically significant as it is the only one that causes severe disease in maize. MSV is a single stranded DNA (ssDNA) virus which, like RNA viruses, has high mutation and recombination rates. Given that these processes can sometimes promote viral diversity and result in the rapid evolution of new, fitter MSV variants, continuous genomic surveillance of MSV is therefore important. Based on analyses of full genome sequences, MSV-A has been classified into five subtypes (-A_1, -A_2, -A_3, -A_4, and -A_6) and more than 20 recombinant lineages. Here, I showed using laboratory-based experiments that maize infecting mastreviruses such as *Maize streak Reunion virus* (MSRV) and MSV-C which have been found maize plants displaying severe streak symptoms do not in fact cause severe streak symptoms in maize when used to infect maize on their own. Although a mixed infection involving MSRV and MSV-B resulted in slight changes in symptom phenotypes it is unlikely that MSRV and MSV-C are responsible for emerging maize diseases. I carried out model-based phylogenetic and phylogeographic analyses of MSV-A movement dynamics in and out of Madagascar, Ethiopia and Rwanda using newly determined MSV-A genome sequences (Madagascar: n = 56; Ethiopia: n = 84) together with other sequences from GenBank. I showed that most movements of MSV-A into Madagascar have been from East Africa between the early 1990s and 2000s. My inferences show that MSV-A_1 variants currently found in Ethiopia likely arrived there from Uganda or Kenya between 1985 and 1988. Similarly the MSV-A_1 variants found in Rwanda likely also moved there from Ethiopia, Kenya or Uganda between 2007 and 2011. The time periods over which inferred movements of MSV-A_1 into Madagascar, Rwanda and Ethiopia occurred all correspond with the period during which trade between these and other East African nations was being liberalized. Although these temporally-scaled phylogeographic analyses

indicated that human activities are likely responsible for some of the long-range movements of MSV-A_1 variants (such as movements from East Africa to Madagascar), leafhopper-mediated dissemination of these variants also likely played a major role in long and short distance movements of these variants within both Madagascar and between East African countries. Over ~ 90 years of evolution that yielded MSV-A-ZW-MatA_1994 in the MSV-A_1 lineage, produced symptoms that have varied in a less concerted ways, or largely remained unchanged. Major harms (intensity of chlorosis, leaf deformation and stunting) have decreased while the amount of colonized cells (chlorotic areas) that determine onward transmission have increased. These data suggest MSV-A has evolved to optimize the number of cells it infects for effective onward transmission, while reducing excessive harm to its hosts. Altogether, these results suggest (1) synergism potentially plays a role in some instances of severe streak disease and (2) the movement of MSV-A_1 within the East African region and Madagascar emphasizes the importance of this MSV-A subtype as a major ongoing threat to maize production within these regions; and (3) over the last ~90 years, the MSV-A_1 subtype has evolved to produce greater chlorotic areas on the leaves of infected maize plants while at the same time either not increasing or reducing the degrees of chloroplast destruction, stunting and deformation caused by infections: characteristics that may have enhanced the transmissibility of this variant and therefore played an important role in the present rise to dominance of this subtype throughout East Africa and Madagascar.

Acknowledgements

My utmost gratitude goes to the Almighty God for His divine help and providence that saw me through this programme.

My sincere and thoughtful appreciations go to my supervisor; Associate Professor Darren Martin for supervising this project and ensuring that I am on the right track to becoming the scientist of my dream. Working with Darren, I learnt a lot from his wealth of experience. Striving through the stormy weather and wading through the murky water, day and night, I was able to hone my wet laboratory, programming and writing skills and got more confident. I am also indebted to Associate Professor Arvind Varsani for sending me viral clones for this project, his valuable advice, as well as timeous intervention, support, and dedication towards helping me sequence my samples. I am grateful to Dr Lara Donaldson for the crucial support I received from her. I sincerely appreciate Dr Gordon Harkins, Dr Aderito Monjane and Dr Simon Dellicor for their invaluable assistance and support towards the successful completion of this project.

My appreciation also goes to Associate Professor Joan Passmore and everyone in the Mucosal Immunology Laboratory. I especially thank Hoyam, Shameem, Nadia, Monalisa, Bridget and Brian for being of immense help and good friends. I sincerely thank the members of the Cbio group for being great friends and colleagues. My appreciation also goes to Professor Nicky Mulder for her kind words of encouragement and motivations. I sincerely appreciate my workmate Penelope Hartnady for her crucial help and support; thanks for being ruthful!

I am grateful to the National Research Foundation (NRF) South Africa, The World Academy of Sciences (TWAS) Italy, and the University of Cape Town for funding this programme and providing me the world class PhD training of my dream.

My heartfelt appreciations go to my mother: Alice Olorunfemi Oyeniran, and siblings: Mrs Adenike Lashilola, Mrs Olufunmilola Ogundana, and Architect Olumide Oyeniran and my late twin brother; Taiwo Oyeniran for their unparalleled supports in all fronts. I specially thank my darling wife; Dr (Mrs)

Olubukola Helen Oyeniran for her understanding, patience and perseverance juggling her demanding PhD commitments with work, and looking after our son Adebare while I was inevitably absent. I can't thank you enough! I sincerely appreciate my in-laws: Mrs Felicia Aladeselu, Mrs Temidayo Onifade, Oluwamodupe and Omolayo for the supports I received from them.

Moreover, I thank my colleagues and friends: Dr Isaac Sanusi, Dr Abiola Taiwo, Samuel Obimakinde, Solomon Aremu, Daniel Muthitu, Javan Okendo, Dr Sinkala Musalula, Dr Fredrick Nindo, and Dr Anna Ojo. I am grateful to the families of Professor & Professor (Mrs) Okunlola, Professor Oladunmoye, Dr and Mrs Ekunola, Professor Oyetayo, and Professor (Mrs) Adebolu for their timely advice, supports and encouragements. As for me, the best part of the journey was everything I learnt from everyone, and to every individual who has contributed to my success and progress, I want to say a big thank you! Once again, I acknowledge and thank you all.

Acronyms

AbMV	:	*Abutilon mosaic virus*
AFTZ	:	African Free Trade Zones
ALCV	:	*Alfafa leaf curl virus*
BCTV	:	*Beet curly top virus*
BCTIV	:	*Beet curly top Iran virus*
BEAST	:	Bayesian Evolutionary Analysis Sampling Tree
BeYDV	:	*Bean yellow dwarf virus*
BF	:	Bayes Factor
BGMV	:	*Bean golden mosaic virus*
BSSVS	:	Bayesian Stochastic Search Variable Selection
cccdsDNA	:	Covalently Closed Supercoiled Circular Double Stranded Deoxyribonucleic Acid
CMV	:	*Cucumber mosaic virus*
CP	:	Capsid Protein
CpCDV	:	*Chickpea chlorotic dwarf virus*
CRA	:	Common Region A
CRB	:	Common Region B
CSMV	:	*Chloris striate mosaic virus*
cssDNA	:	Circular Single Stranded Deoxyribonucleic Acid
dsDNA	:	Double Stranded Deoxyribonucleic Acid
dsRNA	:	Double Stranded Ribonucleic Acid
DSV	:	*Digitaria streak virus*
EAC	:	East African Community
EACM-UG	:	East African Begomoviral Cassava Mosaic Recombinant Variant Uganda

ECSV	:	*Eragrostis curvula streak virus*
EcmLV	:	*Euphoria caput-medusae latent virus*
ESS	:	Effective Sample Size
GTR	:	General Time Reversible
GWAS	:	Genome Wide Association Studies
HIV	:	*Human immunodeficiency virus*
HrCTV	:	*Hoseradish curly top virus*
HPD	:	Highest Probability Density
IOC	:	Indian Ocean Commission
IR	:	Intergenic Region
LB	:	Luria-Bertani
LIR	:	Long Intergenic Region
MCC	:	Maximum Clade Credibility
MCMC	:	Markov chain Monte Carlo
MMV	:	*Maize mosaic virus*
MRCA	:	Most Recent Common Ancestor
MP	:	Movement Protein
MSD	:	Maize Streak Disease
MSRV	:	*Maize streak Réunion virus*
MSRV-YN	:	*Maize streak Réunion virus*- Yunnan isolate
MSV	:	*Maize streak virus*
MT	:	Metric Tons
NSP	:	Nuclear Shuttle Protein
OD	:	Optical Density
ORF	:	Open Reading Frame

PDV	:	*Potato dwarf virus*
PPS	:	Posterior Probability Support
PrLV	:	*Prunus latent virus*
RBR	:	Retinoblastoma Related Protein
RCR	:	Rolling Circle Replication
RDP	:	Recombination Detection Program
RDR	:	Recombination-Dependent Replication
REN	:	Replication Enhancer
REP/REPA	:	Replication Associated Protein
RFI	:	Replicative Form Intermediate
RNA	:	Ribonucleic Acid
RPM	:	Rotation Per Minute
seCTV	:	*Sesame curly top virus*
SD	:	Symptom Determinant
SIR	:	Short Intergenic Region
SPREAD3	:	Spatial Phylogeographic Reconstruction of Evolutionary Dynamics
SS	:	Silencing Suppressor
ssDNA	:	Single Stranded Deoxy-Ribonucleic Acid
ssRNA	:	Single Stranded Ribonucleic Acid
TCT	:	*Turnip curly top virus*
TPCTP	:	*Tomato yellow curl virus*
TRAP	:	Transcription Activator
TYDV	:	*Tobacco yellow dwarf virus*
TYLCV	:	*Tomato yellow curl virus*
WDV	:	*Wheat dwarf virus*

Table of Contents

List of tables

List of figures

Chapter 1: Introduction

1.1 Evolutionary origin of viruses

Evolution is a natural process that involves the persistent adaptations of organisms to increase resource utilization efficiency within their ecological environments. Evolution frequently involves the modification and/or stabilization of these environments to expedite the survival of the best adapted organisms to those environments (Johansson, 2008; Lefeuvre *et al*., 2019; Mergeay & Santamaria, 2012). The pervasiveness of human influences on, or human disruptions of, environments can have a major impact on the ongoing evolution of organisms on Earth by determining how these organisms respond (Acosta-Leal *et al*., 2011).

The most familiar manifestation of evolution is sexual reproduction where by progenitors do not simply clone themselves in their descendants (Penn & Smith, 2007). The genetic variation created by processes such as sexual reproduction is the raw material upon which evolution operates. Another important process for generating genetic diversity is mutation, the most familiar evolutionary impacts of which are seen in cancers where nucleotide changes, nucleotide insertions and nucleotide deletions create genetically heterogeneous host cell populations that begin competing with one another and healthy host cells for resources: a process that culminates in tumour formation (Meacham & Morrison, 2013; Shackleton *et al*., 2009).

Evolution will occur as long as reproduction of genetically variable individuals with differing degrees of fitness occur (Penn & Smith, 2007; Scoville, 2018). The fossil record of the macroscopic and microscopic remnants of past life on Earth have provided only the slightest glimpses of the evolutionary pathways that have yielded the higher organisms that exist today

(Betts *et al*., 2018; Javaux & Lepot, 2018). However, viruses of the past have not left conventional fossils; although amber-preserved insects and seeds may have displayed evidence of viral infections (Hull & Hull, 2014c).

Fortunately, analyses of the nucleic acid and protein sequences of any group of evolutionarily related organisms can provide much information on the evolutionary pathways that yielded these organisms since the time of their most recent common ancestors (Jordan *et al.* 2002). Although the large numbers of viral genome sequences that have so far been determined are very useful for viral evolutionary studies, appropriate care must be taken when interpreting the evolutionary significance of the patterns of nucleotide and amino acid variations that are detectable within these sequences (Lodish, 2000; Martin *et al.* 2011). Paradoxically, it is the genome regions that accumulate the lowest numbers of amino acid and/or nucleotide changes during the evolution of organisms that tend to have the greatest evolutionary significance: these unchanging or "conserved" regions tend to be the ones that, when changes occur within them, have the greatest negative impact on the survival of organisms (Goh *et al*., 2000).

There are three main theories of how viruses originated. (i) They descended from primitive pre-cellular life forms closely related to the earliest RNA prebiotic polymers with enzymatic properties (like ribozymes) that became parasites of the earliest cells. (ii) They developed from normal cell constituents such as jumping genes and mRNAs that went rogue and successfully evaded normal cellular control mechanisms, ultimately becoming self-replicating entities; or (iii) they evolved from degenerated cells that parasitized normal cells. (Forterre, 2006; Hull & Hull, 2014c).

There is growing evidence for a pre-cellular origin of viruses (Durzyńska & Goździcka-Józefiak, 2015; Jordan *et al*., 2002). Specifically, it is hypothesised that the virus world developed within intertwined networks of inorganic compartments within which virus-like genetic elements evolved into both cellular genes and viral genomes (Figure 1.1). In this pre-cellular life (virus world), RNA viruses may have been the first to evolve (Cuevas *et al*., 2003; Hull & Hull, 2014a; Ivanowski, 1892), followed by retro-elements and then DNA viruses (Hull & Hull, 2014c) (Figure 1.1).

The main support for this hypothesis is that most viral proteins that are key to replication and morphogenesis, have no detectable homologues in cellular lifeforms and it is therefore unlikely that these genes evolved from those found in cellular organisms (Koonin *et al*., 2015). An example of one of these key viral proteins is the double beta-barrel capsid protein of isometric virus particles and superfamily three helicases, that are encoded within the genomes of RNA and DNA viruses of eukaryotes and prokaryotes (Koonin *et al*., 2006; Krupovič & Bamford, 2010).

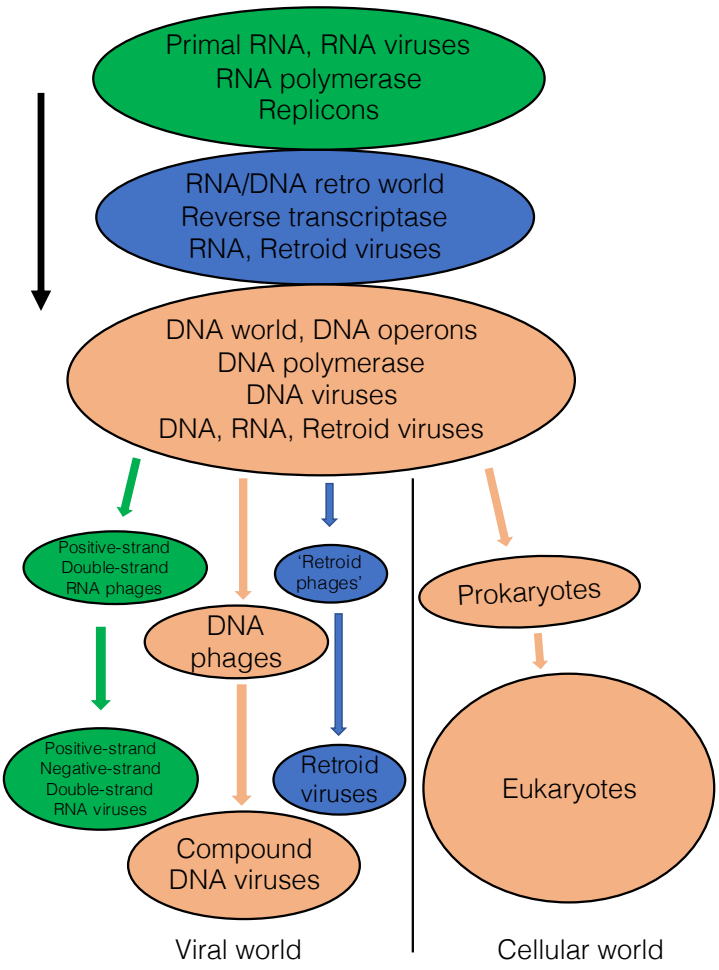


Figure 1.1: Evolution of the cellular and viral worlds from Koonin & Dolja, (2006). The time direction is from top to bottom according to the black arrow. The far left levels show three key stages of precellular evolution involving different genetic entities comprising interactions between proto-viral (selfish) and proto-cellular (altruistic) replicating entities (Koonin, Dolja, & Krupovic, 2015; Koonin, Senkevich, & Dolja, 2006). Differentiation of ancient genetic entities into free-living prokaryotic cells and the viral parasites of these cells is shown in the middle with the DNA phages. The emergence of eukaryotes likely promoted the structural complexity of viruses as is illustrated in the bottom, far right side of the diagram (Iyer *et al*., 2006). The continuous line demarcating the two worlds represents continuous gene exchanges between cells and viruses. Major directions of gene flow in the viral world are represented by coloured arrows.

1.2 Viruses and their hosts

Although it is difficult to precisely state when the very first virus was discovered, *Tobacco mosaic virus* was the first virus to be definitively documented as a non-cellular disease agent (Ivanowski, 1892). Subsequent research revealed that viruses are obligate parasites that require viable hosts for survival, occupying the borderline between the living and non-living (Coffin, 2013; Hohn & Sharma, 2014; Hull & Hull, 2014a).

Many viruses have evolved to become masters of stealth that commonly hide their replication, gene expression activities, and the presence of infectious viral particles within the cells that they infect (Worobey, Bjork, & Wertheim, 2007). In addition, viruses commonly have high mutation and recombination rates, which produce large amounts of genetic variation and therefore plentiful opportunities for natural selection to promote the survival of variants with increased fitness (Lodish, 2000; Worobey, Bjork, & Wertheim, 2007). Viruses are essentially highly flexible collections of protein coated genetic material that can easily hijack sophisticated host biochemical mechanisms for their reproduction. The genetic composition of a given virus population is also related to its host environment, thus vital interactions between viruses and their hosts are common (Acosta-Leal *et al*., 2011). A plant virus may infect and cause mild to severe diseases in one or more hosts depending on its genes and their interactions with the host environment (Borrow, 1997; Coffin, 2013; Lefeuvre *et al*., 2019).

Basically, interaction of viral particles with host cells can either result in complete cell lysis (lytic) or lysogeny (Cohen, 2016; Fermin, 2018). Whether a virus interaction with its host becomes lytic or lysogenic is subject to its genetics (Coffin, 2013; Holland *et al*., 1980) as most lysogenic infections are commonly associated with mild and less virulent variants (Oldstone,

2009). Impacts of co-evolved genomic virus-host interactions are likely quite crucial to producing fitter offspring in viruses that undergo frequent host switching (Lawrence *et al*. 2012; Martin *et al*. 2011).

Virus genetics definitely determines the natures and extents of how viruses interact with their hosts. From cell attachment to replication cycle completion there is a strong selective pressure for viruses to not kill (or excessively damage) their hosts prior to transmission. The severity of viral diseases in a host are therefore directly constrained by how disease symptoms impact the onward transmission of viruses (Monjane *et al*., 2020; Pazos & Valencia, 2008).

Hosts and cell ranges of a virus are often a useful basis for classifying them (Parrish & Kawaoka, 2005; Varsani *et al*. 2014). For instance, bacteriophages are viruses that only infect bacteria (Labrie, Samson, & Moineau, 2010) whereas, in general, evolutionarily distinct viruses infect animals and plants (Gutiérrez *et al*., 2013; Roossinck, 2011).

While most known plant viruses are transmitted by insects, only a small proportion of plant viruses are actually able to infect both plants and insects (Ammar *et al*., 2009; Maramorosch, 1955). There are two main mechanisms by which plant viruses are transmitted by insects according to Dáder *et al*. (2017): (1) persistent circulative transmission whereby ingested viral particles traverse the gut epithelial barrier, access the haemocoel, and end up in the salivary duct (Ammar et al., 2009); and (2) non-persistent, and non circulative transmission whereby virus particles bind with molecular components of mouthparts, and are released in excreted saliva when insects probe host cells for feeding sites (Uzest *et al*., 2007). *Maize streak virus* (MSV) is an example of a virus that is transmitted in a persistent circulative manner by *Cicadulina* sp. leafhoppers (Reynaud and Peterschmitt, 1992; Storey, 1933, 1938), unlike

Cucumber mosaic virus (CMV) that is transmitted in a non-persistent, non circulative manner by aphids (Gallitelli, 2000).

1.3 Plant viruses and host preference

Animal life on the planet Earth was only made possible because plants evolved approximately 470 million years ago which, together with the oceans generated sufficient ambient aquatic and atmospheric oxygen levels to support animal life (Kenrick & Crane, 1997).

Plants infecting viruses have diverse genome types (Hull & Hull, 2014a). Interestingly, the prevalence of plant virus species with (+)-sense single-stranded ribonucleic acid (ssRNA) genomes far exceeds that found in bacteria and fungi which tend to have far more viruses with double-stranded deoxyribonucleic acid (dsDNA) and double-stranded ribonucleic acid (dsRNA) genomes (Hull & Hull, 2014c).

1.4 The *Geminiviridae* Family

Geminiviruses are economically important pathogens that persistently pose a devastating widespread agricultural threat to monocot and dicot crops especially in the tropical and sub-tropical climates (Fauquet & Stanley, 2005; Kraberger *et al.*, 2017; Malik *et al*., 2011). Despite having small single-stranded DNA genomes that typically encode only four to six proteins, geminiviruses have high mutation and recombination rates which provide them with evolutionary flexibility needed to continually evolve novel pathogenic variants (Martin *et al*., 2011; Preiss & Jeske, 2003).

Geminivirus infections of plants can be either symptomatic or asymptomatic. Symptoms range from curled leaves, yellow veins, bright yellow mosaics, chlorosis, chlorotic streaks and varying degrees of stunting (Guadie *et al*., 2019; Legg & Fauquet, 2004; Moffat, 1999; Pande *et al*., 2012). Even though closely related viruses can produce vastly different symptoms, symptom-based diagnoses of geminivirus infections is still the way in which most geminivirus epidemics are tracked in the field. Symptoms can also provide useful insights into geminivirus-induced host pathology and have been used on several occasions together with sampling locations for further virulence-based classification schemes (Martin & Rybicki, 2002).

Recent advancements in diagnostic methods and DNA sequencing have both created new research areas for plant virologists, and have increased the pace at which geminivirus full-genome sequences are being generated. These advances are, however, exposing weaknesses in the systems used to taxonomically classify virus full genome sequences. There are currently nine recognized genera within the family *Geminiviridae*: *Mastrevirus*, *Curtovirus*, *Topocuvirus*, *Begomovirus*, *Eragrovirus*, *Turncurtovirus*, *Becurtovirus*, *Capulavirus* and *Grablovirus.* These genera are based on virus insect vector species, host ranges, genome organizations and genome-wide pairwise sequence similarities (Fauquet *et al*., 2003; Muhire *et al*., 2013; Varsani *et al*., 2014a; Varsani *et al*., 2017). The genome component organizations of representative viruses in each of the nine genera together with viruses that have yet to be assigned to genera are shown in Figure 1.2 and Table 1.1 (Varsani, Navas-Castillo, *et al*., 2014a).

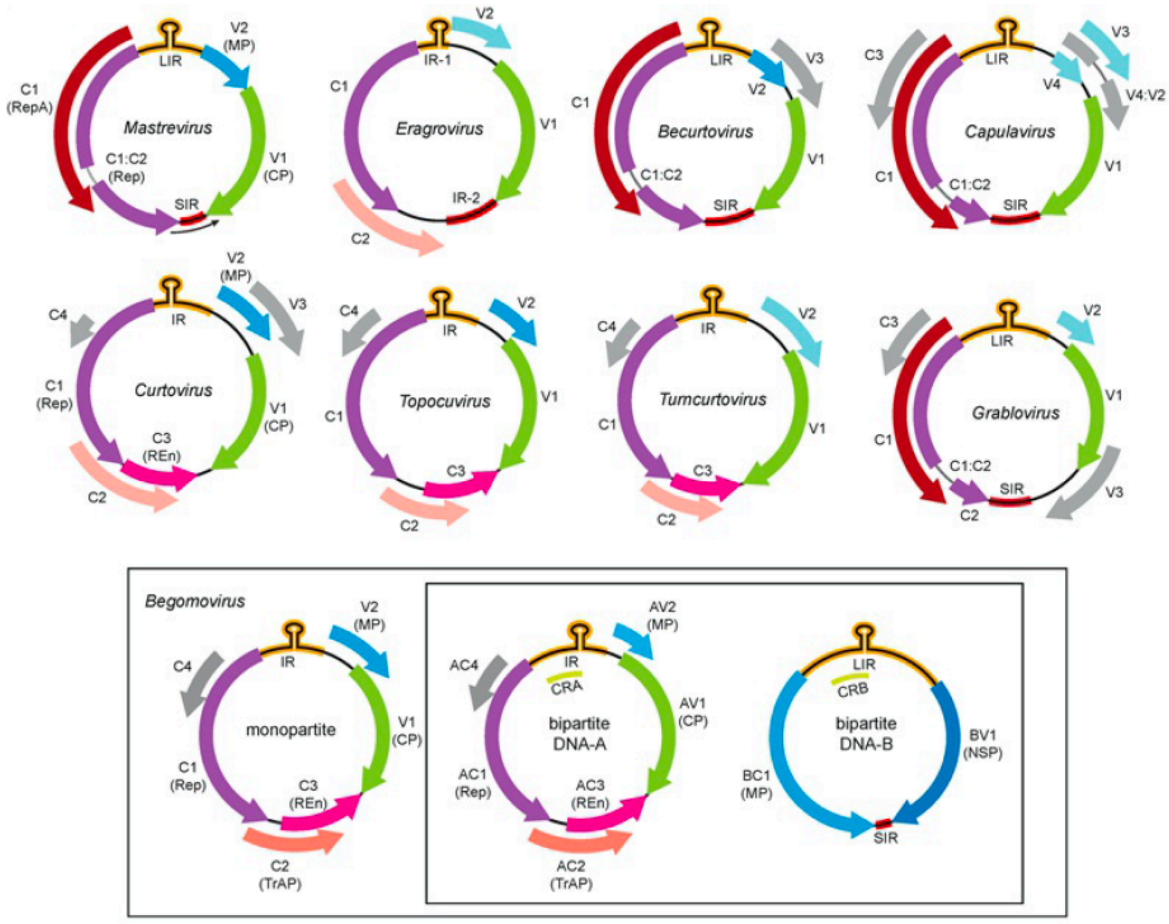


Figure 1.2: Genome organizations in the geminivirus genera from Zerbini *et al*. (2017). Open reading frames (ORFs) are illustrated as being encoded on the virion-sense (V) or complementary-sense (C) strand, and protein products are coloured. The stem-loop containing the conserved 5′-TAAGATTCC-3′ sequence in the long intergenic region (LIR) is shown. CRA and CRB, common regions A and B; coat protein, CP; intergenic region, IR; movement protein, MP; nuclear shuttle protein, NSP; replication enhancer protein, Ren; replication associated protein, Rep; short intergenic region, SIR; transcriptional activator protein, TrAP.

Table 1.1: Summarized classifications and properties of the *Geminiviridae* (Zerbini *et al*., 2017).

Genus	No. of approved exemplar members	Examples of representative species	Genome size (nt) / Arrangement	Insect vector	Host range	Species sequence identity
Mastrevirus	41	*Maize streak virus* (MSV) *Maize streak Réunion virus* (MSRV)	2684-2701/monopartite 2882/monopartite	Leafhopper Leafhopper	Poaceae Poaceae	<78%
Curtovirus	3	*Beet curly top virus* (BCTV) *Hoseradish curly top virus* (HrCTV)	2933/monopartite 3080/monopartite	Leafhopper Leafhopper	Dicot plants Horseradish	<77%
Topocuvirus	1	*Tomato pseudo-curly top virus* (TPCTP)	2861/monopartite	Treehopper	Solanaceae	N/A
Begomovirus	425	*Bean golden mosaic virus* (BGMV) *Tomato yellow curl virus* (TYLCV)	A: 2644; B: 2609/bipartite 2743-2790/monopartite	Whitefly Whitefly	Leguminosae Solanaceae	<91%
Eragrovirus	1	*Eragrostis curvula streak virus* (ECSV)	2752-2754/monopartite	Leafhopper	Poaceae	>94%
Turncutovirus	3	*Sesame curly top virus* (seCTV) *Turnip curly top virus* (TCT)	2964/monopartite 2981/monopartite	Leafhopper Leafhopper	Brassicaceae Brassicaceae	>80%
Becurtovirus	3	*Beet curly top Iran virus* (BCTIV) *Exomis microphylla latent virus* (EmLV)	2839-2845/monopartite 2974/monopartite	Leafhopper Leafhopper	Dicot plants Dicot plants	>80%
Capulavirus	4	*Alfalfa leaf curl virus* (ALCV) *Euphobia caput-medusae latent virus* (EcmLV)	2745/monopartite 2678/monopartite	Aphid Aphid	Fabaceae Solanaceae	>78%
Grablovirus	3	*Grapevine red blotch virus* (GRBV) *Prunus latent virus* (PrLV)	3206/monopartite 3174/monopartite	Treehopper Treehopper	Grapevine Stone fruit tree	<80%

The *Mastrevirus, Becurtovirus,* and *Capulavirus* genera have short and long intergenic regions and express *Rep* from spliced complementary strand transcripts. In addition, begomoviruses and curtoviruses have a gene completely embedded within *Rep* (ORF C3) (Varsani *et al*., 2017).

Geminivirus genomes encode up to eight genes that are bidirectionally transcribed during replication (Varsani *et al*., 2014b). While the complementary sense strand encodes the *Rep* gene, the capsid protein (*Cp*) is encoded on the virion-sense strand and conserved across characterized geminiviruses (Fauquet *et al*., 2005; Monsalve *et al*., 2014). Other genes present in begomoviruses, curtoviruses, topocuviruses and turncurtoviruses are very similar to the replication enhancer (*ren*) and silencing suppressor (*ss)*, symptom determinant (*sd*) or transcription activator (*trap*) genes found in eragroviruses (Varsani *et al*., 2009; Varsani *et al*. 2014). Despite commonly falling in the same genome region downstream of the virion strand transcription initiation area, geminivirus movement protein (*Mp*) genes are so diverse that there is no detectable amino acid similarity in *Mp* sequences of these viruses in different genera (Fauquet & Stanley, 2005).

Another distinctive feature of the mastreviruses, eragroviruses and becurtoviruses is the presence of two intergenic regions (IRs). The long intergenic region (LIR) containing the virion-strand origin of replication in mastreviruses is the smaller of the two intergenic regions in known eragroviruses, while the short intergenic region (SIR) contains the transcription termination signals of the bidirectionally transcribed genes (Varsani *et al*., 2009; Varsani *et al*., 2017).

1.5 Evolution of virulence in economically important mastreviruses

Recombination seems to have played a prominent role in the evolution of the different geminivirus genera, in some cases potentially creating new genes (Martin *et al*., 2011; Saleem *et al*., 2016; Varsani *et al*., 2008). The evolution of species within the geminivirus genera has also involved frequent recombination but appears to be primarily driven by high mutation rates (Duffy & Holmes, 2008; Van Der Walt *et al*., 2008).

High mutation and recombination rates do not, however, always translate into persistently rapid evolution (Hull & Hull, 2014b; Monjane *et al*., 2014). Put simply, mutation and genetic recombination create genetically diverse populations of viruses, some of which will likely be better adapted to whatever new environmental conditions the members of a virus species might encounter. If new environmental conditions arise, these better adapted variants will predominate and diversify to yield new populations of genetically diverse variants that are poised to confront additional new conditions. However, if new environmental conditions do not arise then, in most virus populations that are already well adapted to the prevailing environment, almost none of the novel genetic variants generated by recombination and mutation will be materially fitter than any others and evolution will not progress in a directed way such that mutants and recombinants that persist in the population will do so almost entirely by chance: a process called genetic drift (Garcia-Arenal, Fraile, & Malpica, 2003; Rousseau *et al*., 2018).

Such considerations are relevant because one of the most important environmental changes that will ever be encountered by a virus is the infection of a host species that it has never previously encountered (Acosta-Leal *et al*., 2011; Jones, 2009). Being pathogens, a major

consequence of the evolutionary adaptation of a virus to a new host environment will be that the degree of harm it inflicts on the new host will likely initially be suboptimal with respect to ensuring the highest probability of onward virus transmission (Figure 1.3) (Alizon *et al*., 2009; Anderson & May, 1982). Here, I will refer to the degree of harm that the virus inflicts on its host as "virulence". In general it appears as though more resistant hosts tend to favour the evolution of more virulent pathogens whereas, more sensitive hosts tend to favour the evolution of less virulent pathogens (Thrall & Burdon, 2003). The genetic causes of altered virulence are, however, frequently very complex (Payne, 2017a) with increased/decreased virulence generally occurring as a collateral consequence of selection almost entirely favouring the maximization of virus transmission: a factor that could result in highly virulent viruses evolving in sensitive hosts and very mild viruses evolving in resistant hosts.

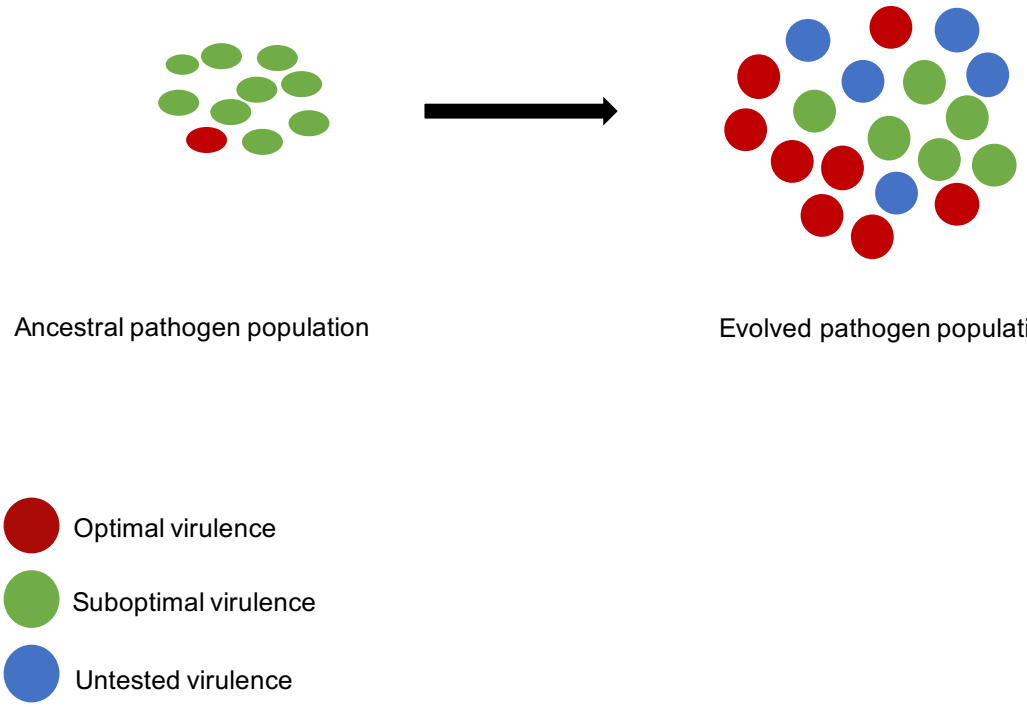


Figure 1.3: The evolution of virulence in virus populations. Pathogen population over time will diverge following expansion and diversifying evolution.

1.6 *Maize streak virus*

Maize streak virus is the type member of the genus *Mastrevirus*, and family *Geminiviridae*, Members of this family have single-stranded DNA genomes comprising one or two circular molecules that are encapsidated in a distinctive germinate particles of 18 nm by 30 nm quasi icosahedral shapes (Chen *et al*., 2015; Lefeuvre *et al.*, 2016; Martin *et al*., 2008 and Zhang *et al*., 2001). All known mastreviruses have only a single genomic molecule of between 2600-2800 nt in length (Buck, 1999; Kreuze *et al*., 2009). The genus *Mastrevirus* according to >41 described species including: *Chickpea chlorotic dwarf virus* (CpCDV), *Bean yellow dwarf virus* (BeYDV), *Chloris striate mosaic virus* (CSMV), *Digitaria streak virus* (DSV), *Tobacco yellow dwarf virus* (TYDV) and *Wheat dwarf virus* (WDV) (ICTV, 2019).

Being a major pathogen of economic importance throughout sub-Saharan Africa and the adjacent Indian Ocean Islands, MSV seriously constrains yields of the region's most important food crop (Varsani *et al*., 2008). Laboratory tests have shown that MSV can also infect other grass families such as barley, rye, and wheat (Autrey, 1983; Storey, 1930; Storey, 1936; Rose, 1978).

All MSV variants/species share >75% genome-wide nucleotide sequence similarity with other MSV species. Within the MSV species, however are eleven known strains, named strain A through strain K according the order of discovery. Variants within each of the strains sharing >91% genome wide nucleotide sequence similarity with other variants within those strains. Within each of the strains different subtypes exist with variants in each subtype sharing >98% genome-wide nucleotide sequence similarity with all other variants within that subtype. Variants within a subtype tend to share similar degrees of virulence in any given host and have

well defined geographical sub-ranges within the global MSV geographical range (Fauquet *et al*., 2008; Martin *et al*., 2001).

Based on intra-strain recombination patterns, Owor *et al.* (2007) further classified variants within the MSV-A strain, specifically the MSV-A_1 subtype grouping, into recombinant lineage groupings. Of the eleven MSV strains only MSV-A is known to cause severe maize streak disease (MSD) in maize (Martin *et al*., 2001). This particular strain is therefore of immense economic relevance. MSV-B and -C are other strains with no major economic importance since they mostly cause MSD in wild grasses species (Varsani *et al*., 2008).

MSV-A is believed to have only emerged as a serious pathogen of maize in approximately 1850 following its emergence via recombination between predominantly grass-infecting MSV-strains (MSV-B, MSV-F and/or MSV-G) (Harkins *et al.*, 2009; Varsani *et al*., 2008).

1.6.1 *Maize streak Réunion virus*

Another maize infecting mastrevirus, *Maize streak Réunion virus* (MSRV) was first reported from Réunion Island in the Indian Ocean (Pande *et al*., 2012) and in China (Chen *et al.*, 2015). Oluwafemi *et al*. (2014) also recovered the first wild grass infecting MSRV from Nigeria with > 97% sequence similarity to the maize-infecting R**é**union isolate. MSRV isolates share < 50% genome-wide nucleotide sequence identity with other known mastreviruses. Importantly, MSRV has been found in maize plants showing severe MSD symptoms which suggests that it may in some way contributing to maize losses to MSD. Further, a related virus, with genome-wide sequence similarity to MSRV of between 70-71% has been found in maize plants displaying MSD symptoms in Ethiopia (Guadie *et al*., 2019). MSRV, just like MSV, appears

to be widely distributed throughout sub-Saharan Africa has frequently been found associated in mixed, infections of grasses with MSV strains such as MSV-B (Claverie *et al*., 2019; Oyeniran *et al*., 2021). From these findings, It is therefore plausible that MSRV and MSV may be synergistically interacting within the context of severe MSD.

1.6.2 Genome organization, replication, and life-cycle of MSV and MSRV

The MSV and MSRV genomes are relatively small (~2.7 kb for MSV, and ~2.8 kb for MSRV) and encode only four genes (Boulton *et al*., 1989; Fauquet *et al*., 2005; Pande *et al*., 2012). Despite the MSV genome being one of the smallest virus genomes known, regulation and interactions of its genes are inherently complex because of the distinct pathways and biological activities its genes and products undergo to establish infections. Specific genome regions also interact with vital host factors to regulate both the life cycle of the virus and the reproduction cycles of infected cells (Shepherd *et al*., 2010). For instance, the virus-encoded replication associated protein (*Rep*) is multifunctional, initiating and facilitating virus replication and regulating virus and host cell transcription (Collin *et al*., 1996; Hefferon, Moon, & Fan, 2006; Hofer *et al*., 1992).

1.6.3 Replication

Geminiviruses have rolling circle replication (RCR) mechanism and recombination-dependent replication (RDR) mechanisms (Jeske *et al*., 2001). MSV, like other geminiviruses, undergoes double-stranded circular transitional replication leaving mini-chromosomes in the infected cell nuclei (Abouzid *et al*., 1988; Pilartz & Jeske, 2003).

The first stage of RCR involves the conversion of genomic circular ssDNA [(c)ssDNA] to covalently closed supercoiled circular dsDNA [(ccc)dsDNA] transitional or replicative form

intermediates (RFI): a process that is accomplished via host targeted, DNA-primed (RNA-primed in begomoviruses) complementary (minus) strand synthesis (Jeske *et al*., 2001). Amplification of the RFI by RCR happens in the second stage, during which the viral RFI serves as a template for transcription and additional DNA replication steps. It is quite important to note that viral RFI serving as transcription templates leads to early expression of *Rep* and *RepA*. A replication initiation site in the 'loop' of a stem-loop structure in the mastrevirus LIR is the site of the virion strand origin of MSV, and presumably MSRV, replication (Fontes, 1994).

Interaction with specific repeated DNA sequences close to the virion strand origin of replication position will prompt *Rep* to introduce a sequence-specific endo-nucleolytic cut in the V- sense RFI genomic strand for transcription purposes (Nikovics *et al*., 2001). In WDV, *Rep* binds with low affinity at the stem of the stem-loop making up O- complexes for low-affinity binding, while high affinity binding occurs at ~140 nucleotides upstream of the virion strand origin in the proximity of the complementary strand gene promoter: a binding complex referred to as a C- complex (Hull & Hull, 2014c; Schalk *et al*., 1989). Formation of O- complex alone is believed to be sufficient for carrying out sequence-specific binding at the loop (Castellano *et al*., 1999; Gutierrez *et al*., 2004). It is also plausible that C- and O- complex bound *Rep* molecules interact to form a higher order complex at the origin (Gutierrez, 1999; Gutierrez, 2000).

The initiation of RCR strand-nicking reaction is achieved in a series of molecular steps involving the nucleophilic attack of a conserved tyrosine residue in a particular motif of *Rep* by the OH group, to the phosphodiester bond between the last T and A residue of the nonanucleotide original plus strand sequence, TAATATT|AC (Heyraud-Nitschke *et al*., 1995; Settlage *et al.*, 2001; Stanley, 1995).

Following the initiation reaction, the virion strand sequence is replicated by host DNA polymerases. Once synthesis of the new virion strand is completed, the parent plus strand is moved from the negative template by the putative helicase activity of *Rep*. Depending on activated replication stage cycle, the newly released positive strand may be channelled back into the replication pool or converted to another RFI, or accumulate as (c)ssDNA destined for encapsidation.

1.6.4 Life-cycle

Infection starts when an insect vector deposits an encapsidated ssDNA genome of MSV or MSRV in a nucleated plant cell (Fondong, 2013; Roossinck, 2010; Shepherd *et al*., 2007). Viral particles left in the leaf-associated phloem tissues will be conveyed through the sieve tubes to parts of the plant far from the feeding site: parts including the photosynthetic sink tissues surrounding the apical meristem that are preferred sites of mastrevirus replication (Peterschmitt *et al*., 1992; Shepherd *et al*., 2010; Storey, 1933). Actively dividing host cells such as those surrounding the apical meristem are not absolutely required for MSV replication to occur (Peele *et al*., 2001), but for an infection to become properly systemic and yield the streak symptoms that characterize MSD, the virus must infect the actively dividing cells that surround the apical meristem which eventually give rise to leaves and the individual tissues therein from which MSV (and presumably MSRV too) must be onwardly transmitted (Hanley-Bowdoin, Settlage, & Robertson, 2004; Lucy *et al*., 1996).

Upon entry into a suitable cell, the virion first uncoats and then releases the ssDNA to the nucleus for replication. In MSV, it appears that partially uncoated ssDNA are moved into the cells by the coat protein (*CP*) which has both DNA-binding capability, and an accompanying nuclear localization signal (Davies *et al*., 1997; Guerrero *et al*., 2020; Owor *et al*., 2007). Inside the nucleus, the ssDNA viral genome is converted to (ccc) dsDNA (RFI) as stated earlier in the

previous section. There are usually no changes to host cell gene expression since vital cofactors and nick repair enzymes associated with the negative strand synthesis are constitutively expressed in plant cells (Palmer & Rybicki, 1998). Once RFI transformation is achieved, the next important step would be to express *Rep* and *RepA*. *Rep* is required to start rolling circle replication (RCR), and *RepA* is essential for inducing the expression of co-factors and host enzymes required for the completion of RCR (Hefferon *et al*., 2006; Lazarowitz, Pinder, Damsteegt, & Rogers, 1989). Promoters of some host genes required for replication, directly or indirectly activated by both *Rep* and *RepA*, may also alter the cell cycle regulatory systems and induce host genes required for virus replication (Hefferon *et al*., 2006; Lazarowitz *et al*., 1989).

Interestingly, the MSV complementary or C- sense promoter in maize suspension cells is most active in the early S phase of the cell cycle (before histone H4 transcription begins), whereas the *CP* promoter, on the other hand, has significantly higher activity in early G2 (Muñoz-Martín *et al*., 2003; Nikovics *et al*., 2001). The observable expression timing difference from the promoters correlates with MSV gene product functions. For example, higher C- sense promoter expression by early S- phase could produce *RepA*-mediated distortion of the host cell cycle by cleaving retinoblastoma-related (RBR) protein, releasing the regulators for transitioning from G1 to the S- phase of the cycle (Davies *et al*., 1997; Ruschhaupt *et al*., 2013). Further, *Rep* is pivotal to starting RCR during the S- phase. Comparatively, *CP* is needed for DNA encapsidation and systemic infection (Boulton, 2002; Davies *et al*., 1997; Guerrero *et al*., 2020; Shepherd *et al*., 2007), but not in the early stages of replication (Nikovics *et al*., 2001).

CP expression is usually timed in a manner that ensures ssDNA is converted early during an infection and it's usually removed late in the infection (Guerrero *et al*., 2020). It is noteworthy

that the C- sense promoter becoming active again in the late G2 phase may directly relate to the fact that Rep or *RepA* interfere with G2 phase progression, constraining the infected cells to S- phase and, in this manner, also prevent host cells from mounting an effective response to infection while maximizing virus replication chances (Guerrero *et al*., 2020; Gutierrez, 2000; Nikovics *et al*., 2001).

Rep binding near the C- sense TATA box (C complex) in MSV has been linked with the interference of *Rep* transcription initiation, resulting in a negative feedback mechanism (Argüello-Astorga & Ruiz-Medrano, 2001). Even with the *Rep* gene copy number increasing exponentially during genome replication, *Rep* expression remains fairly constant. (Horváth *et al*., 1998; Kong *et al*., 2000). Chen *et al*. (2011) opined *RepA* also functions as a down-regulator of viral replication. *RepA* seems important to all stages of infection, from early preparation of cellular environment for replication, blocking *Rep* in the later stage of the cycle, to activating *CP* appearance in the latest infection stages. *RepA* mediated regulatory mechanisms may be relevant at the transcriptional, post-transcriptional/intron splicing, or even after translation (Boulton, 2002; Hofer *et al.*, 1992) stages of gene expression. Post-translational activities may include further processing involving, for example, the *RepA* accumulation state (Collin *et al*., 1996). Figure 1.4 summarizes the MSV infection process.

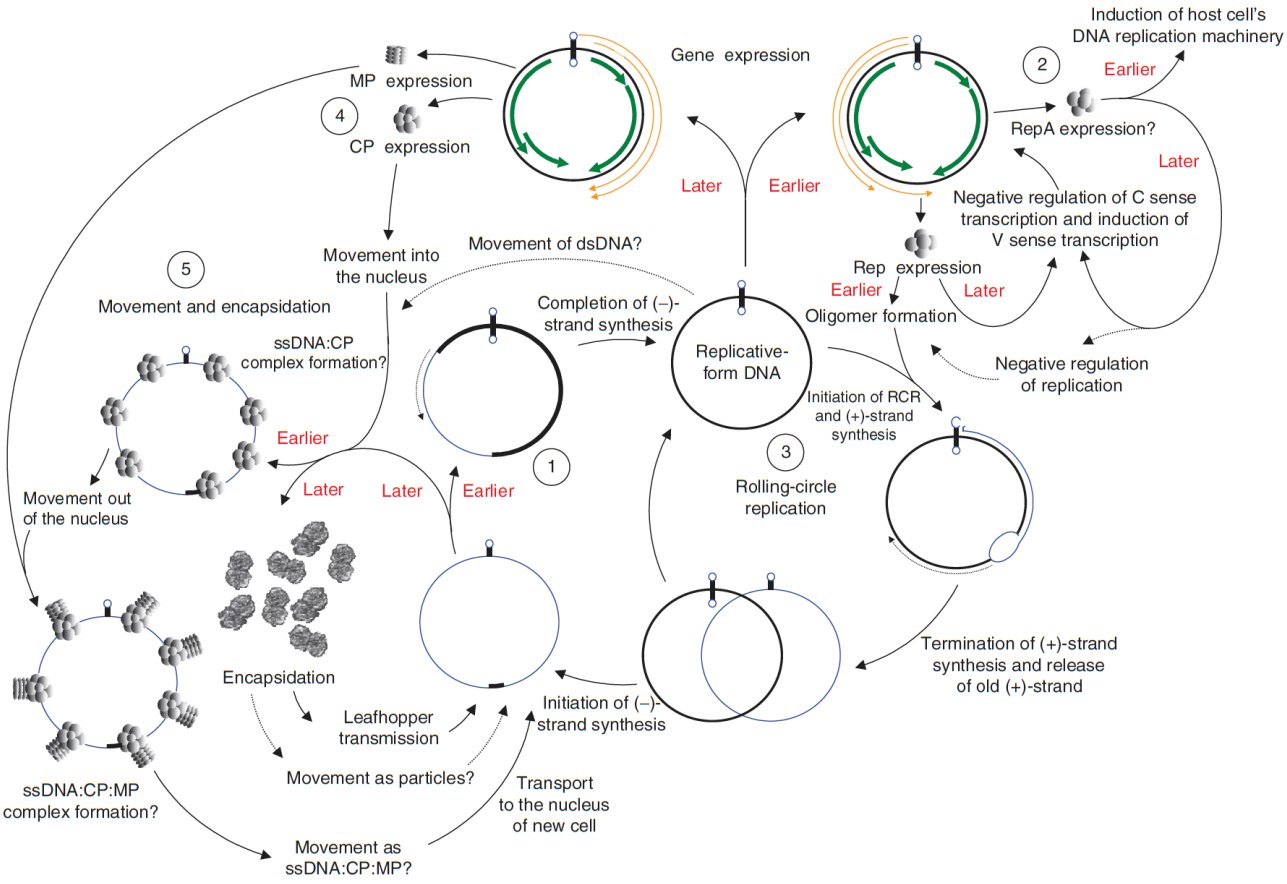


Figure 1.4: MSV infection process overview from Martin *et al*. (2008). Early during an infection, a dsDNA replicative form intermediate (RFI) forms (1). This is followed by *RepA* expression and the induction of a cellular conducive to DNA replication (2). Early *Rep* expression and RCR initiation then occurs (3). Expression of MP and CP occurs at a later stage, following genome amplification and probable Rep and/or RepA induction of the V- sense promoter (4) Movement and/or encapsidation then takes place (5). Cell to cell or systemic movements within plants phloem of unencapsidated, dsDNA, and encapsidated ssDNA. Blue, bold black, and orange lines represent ssDNA, dsDNA, and RNA respectively.

1.7 Epidemiology

Approximately 300 years following the introduction of maize to Africa by Portuguese traders in the 16th century (McCann, 2001), the economically important maize adapted MSV-A strain emerged and began constraining the production of maize on the continent (Harkins *et al*., 2009). MSV-A is able to cause symptomatic infections in over 80 wild-grass species (Damsteegt, 1983), and it is like the progenitors of MSV-A were adapted to infecting grasses in the genus *Digitaria* sp (Varsani *et al.*, 2008). MSV-A can also infect and cause streak disease in other exotic species such as barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), rye (*Secale cereal*), sugarcane (*Saccharum officinarum*), and oats (*Avena sativa*) albeit with minor economic implications (Soto, 1982). Maize streak disease yield losses in susceptible maize plants are tightly linked to the time of viral acquisition. Bosque-Pérez, Olojede, & Buddenhagen, (1998) posited that the younger the plant, the higher the susceptibility to MSV, and that seedlings inoculated at the earlier leaf stages may likely experience higher disease severities and yield losses.

Within field spread of MSV can be linked to several factors such as (i) early planting season usually characterized with poor and irregular rainfall that provide breeding grounds for the virus (Dabrowski & Cwikla, 1991; Shepherd *et al*., 2010) (ii) the population density of MSV-A-infected wild grasses that are in close proximity to maize plants (Autrey, 1983) (iii) the presence of high percentages of MSV transmitting leafhoppers and (iv) external environmental factors that drive long-range leafhopper migrations (Rose, 1978).

Insect vectors play a crucial role MSV spread. For instance, of the nine leafhopper species in the genus *Cicadulina* that are capable of transmitting MSV (Bosque-Pérez *et al*., 1999; Magenya *et al*., 2009; Page *et al.*, 1999), *Cicadulina mbila* and *Cicadulina storeyi* fly the

farthest and are able to transmit MSV more efficiently than other species (Charles, 2014; Dabrowski & Cwikla, 1991; Fajinmi *et al*., 2012). Leafhopper species compositions, sizes and distributions vary across Africa and are strongly impacted by seasons, geography and farm practices that control the availability of their foliar diets (Martin & Shepherd, 2009; Moreno, 2009).

1.8 Hypothesis

Of the eleven known MSV strains, only MSV-A causes economically significant damage. However, Willment *et al*. (2001); Varsani *et al*. (2008) and Harkins *et al*. (2009) had posited that MSV has the potential to evolve increased virulence such that mild strains might adapt to infect and cause severe MSD in maize. Consistent with this hypothesis is the recent discovery of multiple MSV-C isolates infecting maize plants displaying MSD symptoms in Kenya (Kraberger *et al*., 2017), the isolation of MSRV from maize plants displaying severe MSD symptoms in La Réunion and the discovery of frequent mixed infections involving MSRV and MSV-B isolates in La Réunion (Claverie *et al*., 2018).

Right now, most MSV resistant maize genotypes have only been selected for resistance to MSV-A. If MSV strains such as MSV-C or different mastrevirus species such as MSRV are in the process of adapting to infecting maize, it is as yet unclear whether MSV-A resistant maize genotypes will also be resistant to these other potentially important pathogens (Pande *et al*., 2012). The question that comes to mind is: What if other MSV strains or mastrevirus species are also in the process of adapting to infect maize? This question is relevant for the various MSV strains since MSV-C, has already been found within maize plants displaying severe MSD (Kraberger *et al*., 2017). Based on these findings, it will be very interesting to determine just how well these other isolates, or their close relatives are presently adapting to infecting maize.

Having ranked the strains in order of their adaptation to infecting maize, it would then be of interest to identify, within a rigorous statistical framework, how evolution since the emergence of MSV-A has shaped the epidemiology, expanding geographical distributions, and virulence of this economically important strain.

Chapter 2: How Virulent are Emerging Maize-Infecting Mastreviruses?

2.1 Abstract

Maize streak disease (MSD), is one of the most significant biotic constraints on the production of Africa's most important cereal crop. Until recently, the only virus known to cause severe MSD was the A-strain of *Maize streak virus* (MSV-A); a mastrevirus in the family *Geminiviridae*. However, over the past decade, two other mastreviruses, MSV-C and *Maize streak Réunion virus* (MSRV), have been repeatedly found in the absence of MSV-A within maize plants displaying severe MSD symptoms. Here I report on infectious clones of MSV-C and MSRV and test their ability to cause severe MSD symptoms. Although cloned MSV-C and MSRV genomes could cause systemic symptomatic infections in MSD sensitive maize genotypes, these infections yielded substantially milder symptoms than those observed in the field. The MSV-C and MSRV isolates that I have examined are therefore unlikely to cause severe MSD on their own. Furthermore, mixed infections of MSRV and MSV-C with other mild MSV strains also consistently yielded mild MSD symptoms. It is however noteworthy that MSRV produces distinctive striate symptoms in maize that are similar in pattern, albeit not in severity, to those seen in the field, showing that this virus may contribute to the severe MSD symptoms seen in the field. Therefore, despite not fulfilling Koch's postulates for MSV-C and MSRV as causal agents of severe MSD, it is impossible to exclude the possibility that these viruses could be contributing to currently emerging maize diseases.

2.2 Introduction

Maize streak virus (MSV; genus *Mastrevirus*, family *Geminiviridae*) is a single-stranded DNA virus responsible for causing maize streak disease (MSD); one of the most serious constraints on maize production in sub-Saharan Africa. The most striking symptoms of MSD include chlorotic streaks along leaf veins, plant stunting and malformed cobs that frequently yield no seeds (Martin & Shepherd, 2009; Shepherd *et al*., 2010). In this context, MSD causes significant economic losses to African maize farmers and remains a persistent threat to the food security of some world's most vulnerable communities (Charles, 2014; Martin & Shepherd, 2009).

There are 11 known MSV strains - named MSV-A through K - of which only the A-strain causes severe infection in maize. The rest mainly infect wild grasses and only elicit mild infections in maize (Martin & Shepherd, 2009; Shepherd *et al*., 2010;). However, maize plants displaying MSD symptoms have been reported that, rather than containing detectable MSV-A variants, contained either MSV-C (Kraberger *et al*., 2017) or the distantly related mastrevirus, *Maize streak Réunion virus* (MSRV) (Claverie *et al*., 2019; Pande *et al*., 2012). This has prompted speculation that MSV-C and MSRV might be emerging threats to maize cultivation (Kraberger *et al*., 2017; Pande *et al*., 2012).

An emerging pathogen should be capable of entering the cells of a new host species, replicating within these cells, spreading to uninfected cells, and accessing intra- or extra-cellular sites within the host organism from which transmission to new host individuals can occur. Often, achieving these milestones will require that an emerging pathogen causes severe symptoms within its new host (Engering, Hogerwerf, & Slingenbergh, 2013). For the severity of disease

symptoms that arise during coinfections by different geminiviruses, regardless of whether either of these are well adapted to the coinfected hosts, can also be impacted by synergism or antagonism between the coinfecting viruses: either of which can yield biological traits that could be acted on by natural selection (Fondong *et al*., 2000). Given enough time, however, emergent pathogens, whether in isolation or within the context of a disease complex, will likely adapt to their new hosts through the evolution of disease symptoms that become more or less severe based on how they impact the probability of the virus being onwardly transmitted (Alizon *et al*., 2009; Monjane *et al*., 2020).

Although the first infectious clone of a MSV-C was described in 2001 (Schnippenkoetter *et al*., 2001) and was found to produce only mild symptoms in maize, no studies have been undertaken to investigate the symptoms produced in maize by the specific MSV-C isolates that have been found associated with severe MSD. MSRV was first discovered on Réunion Island in 2012 (Kraberger *et al*., 2017) after which closely related MSRV variants were also found associated with severe MSD in Nigeria, China and Ethiopia (Oluwafemi *et al*., 2014; Chen *et al*., 2015; Guadie *et al*., 2019). As with MSV-C, there have been no studies into the symptomatology in maize of any cloned MSRV isolates irrespective of whether they were originally associated with severe MSD or not.

Here I experimentally investigate the severity of MSD caused by a MSV-C clone from a Kenyan maize plant displaying MSD symptoms, and a MSRV clone obtained from a Nigerian *Setaria barbata* plant that is 97.5% identical to a MSRV isolate from a maize plant in Reunion that displayed severe MSD in the absence of a detectable MSV-A infection (Pande *et al*., 2012). Although it would have been better to characterise a MSRV clone taken from a severely

symptomatic maize plant, the MSRV clone that I analysed is the only infectious MSRV clone that I am aware of.

2.3 Materials and methods

2.3.1 *Maize streak virus* isolates and maize hosts

MSV-C (GenBank accession #KM230011) (Kraberger *et al*., 2017), MSRV (GenBank accession #KJ437669) (Oluwafemi *et al.*, 2014) and a South African MSV-B isolate (GenBank accession #KM230017 sampled in 2009 from severe MSD displaying *Ehrharta erecta*) (Kraberger *et al*., 2017) in *Rhizobium radiobacter* as in Monjane *et al*. (2014) and Martin, Willment, & Rybicki, (1999). Three-day-old seedlings of both a MSV sensitive maize genotype (hereafter referred to as the S-maize genotype; Golden Bantam; Sakata Seed Southern Africa), and a moderately MSV resistant maize genotype (hereafter referred to as the M-maize genotype; Popcorn kernel; MBO, South Africa) were inoculated. My inclusion of a MSV-B isolate in this investigation was prompted by a recent metagenomics study of mastrevirus diversity in Réunion which discovered that MSRV in the field frequently occurs in coinfections with MSV-B (Claverie *et al*., 2019). MSV-B is a MSV strain that generally elicits only mild infections in MSV-sensitive maize genotypes (Martin *et al*., 2001; Willment *et al*., 2002).

2.3.2 Construction of *Rhizobium*-infectious clones

Viruses were made infectious using a slightly modified version of methods described by Martin *et al*. (1999). Full-length monomeric isolates genomes cloned into PGEM3ZF(+) (Promega, USA) were linearized by digestion of 1-2 µg DNA of each clone with *Aat*II or *Sca*I for 3 hours followed by a partial digest with a one-tenth unit of *BamH*I for 12 minutes. Tubes were quickly transferred to -80 ^{o}C and then loaded on ice to 0.7 % agarose gel and allowed to

run at 25 V for 480 minutes. Two DNA bands or fragments containing monomer units of the virus and its vector were identified, excised, and gel purified using a Qiagen gel purification kit (QIAGEN USA) and ligated together using standard ligation procedures. Dimers of MSV-C and MSV-B were restricted with full *Xba*l (3 hours) and partial *EcoR*I (one-tenth, 1 minute) because of the presence of *EcoR*I site in these genomes. Full *Xba*l and *EcoR*I digests were achieved for MSRV. Restricted dimers were cloned into analogously digested *EcoR*I and *Xba*l sites of pBI121 (Inqaba Biotech South Africa). *Rhizobium radiobacter*, (formerly *Agrobacterium tumefaciens* C58Ci strain) (Koncz & Schell, 1986) was made competent using standard method and was transformed with pBI121 containing dimer clones as described by Monjane *et al*. (2014)

2.3.3 Rhizobium-inoculation of seedlings

Three days before inoculation, maize seedlings were germinated in damp vermiculite and incubated at 30^{o}C. *Rhizobium* culture from frozen stock was also inoculated in Luria-Bertani (LB) broth (Miller's, Thermofisher), supplemented with rifampicin (100 μg/ml), kanamycin (50 μg/ml), and gentamicin (25 μg/ml) incubated at 30^{o}C and 250 r.p.m till optical density (OD) 600 of between 0.5 - 1.0 was achieved. A plain expression vector (pBI121, Inqaba Biotech South Africa) and a MSV-A isolate (GenBank accession #AF329881) (Martin *et al*., 2001) variant served as negative and positive controls respectively. After confirming the OD, 3 ml of the culture was centrifuged for 3 minutes at 3000 r.p.m and the supernatant discarded. Approximately 1.5 ml of plain LB broth was added, and the tubes tapped gently to resuspend the cells. Cells were spun down before resuspending in 200 μl plain LB broth.

The maize seedlings were cleaned and placed in moist trays covered with a moist paper towel. Using a modified method of Martin *et al*. (1999). The seedlings were randomly separated into three replicate groups of between 36 to 40. A 25 µl Hamilton syringe (Hamilton, Bonaduz, Switzerland) sterilized with bleach, 70% ethanol and rinsed with sterile water was used to introduce *Rhizobium radiobacter* with MSV whose tandem dimer being tested has been inserted between the borders of its T-DNA to the maize seedlings. Specifically, this involved injecting approximately 5 µl of a *Rhizobium radiobacter* culture 1 mm below the coleoptilar node. Mixed *Rhizobium radiobacter* cultures of equal proportions with similar OD range of 0.92 to 1.01 (to ensure the inoculants had equal viral loads) of these viruses were also tested, and infected plants accessed to determine what the impact of coinfection would be on symptom severity. In addition, a *Rhizobium*-infectious clone of an MSV-B isolate (GenBank accession #KM230017) (Kraberger *et al*., 2017), was tested alone and in coinfections with the MSRV infectious clone. Inoculated seedlings were grown in a plant room with a 16:8 hour light: dark cycle for approximately 45-50 days.

2.3.4 Symptom analysis

Symptom analysis was accomplished according to the methods described by Martin & Rybicki, (1999). The plants were monitored for MSD symptoms starting from the third day post-inoculation and each day thereafter for seven weeks. Infection frequencies were taken as the proportion of inoculated plants surviving the *Rhizobium*-inoculation process that displayed infection symptoms 13 days post-inoculation. From 14 days post-inoculation, leaf sections (the second quarter from the base) were sampled weekly from each of the most recent fully emerged leaves. These sections were used to estimate chlorotic areas and intensities of chlorosis using image analysis software as in Monjane *et al*. (2020) and Martin, (2019). Leaf sections were

scanned against a blue background and images used were resolved at 300 dots per inch. A coin was scanned together with the leaves to provide a size standard. The procedure was repeated for leaves three, four, five and six as soon as these leaves were fully out. Images captured were analysed using a modified version (Martin, 2019) of an image analysis programme developed by Martin & Rybicki, (1998) that computes the percentage of leaf area covered by chlorotic lesions in plants showing symptoms. From the collated data, chlorotic area and the intensity of chlorosis were computed for leaves 2-6 for each isolate and plotted on a graph.

2.3.5 Total DNA extraction and recovery of full-length mastrevirus genome

Plants showing foliar streak symptoms and apparently healthy ones were cut and maintained at -20^{o}C. Total genomic DNA was extracted from leaf material of each sample using an Extract-N-Amp^{TM} Plant kit (Sigma-Aldrich, USA) as described by Shepherd *et al*. (2008) according to the manufacturer's instructions. Circular viral DNA was amplified using an Illustra TempliPhi Amplification Kit (GE Healthcare, USA). Full viral genomes were isolated using restriction digest. For each restriction digest reaction, 2.0 µL of rolling circle amplification (RCA) DNA was digested using *BamHI* to yield unit length ~2.7 kb genomes which were ligated to analogously digested Puc19 (New England Biolabs) and sequenced at Macrogen South Korea by primer walking for confirmation purposes only.

2.4 Results and discussion

2.4.1 Infection frequencies in S- and M- maize genotypes

I observed what appeared to be symptoms in 10/34 and 6/42 infected S- and M-maize genotype plants inoculated with MSV-C respectively. For MSRV, distinctive striate symptoms were

observed in 9/31 and 7/37 S- and M-maize plants respectively (Figure 2.1). MSRV and MSV-C also had substantially lower infection frequencies (9/31, and 10/34 respectively) than did MSV-A inoculated S-maize genotype plants (36/37; Figure 2.1) and MSV-B (25/30; Figure 2.1). In general, observed infection frequencies for the viruses were much lower in M- maize genotype than in S- maize genotype. I confirmed that these plants were indeed symptomatically infected with MSV-C and MSRV following Sanger sequencing full viral genomes isolated by rolling circle amplification from symptomatic leaves.

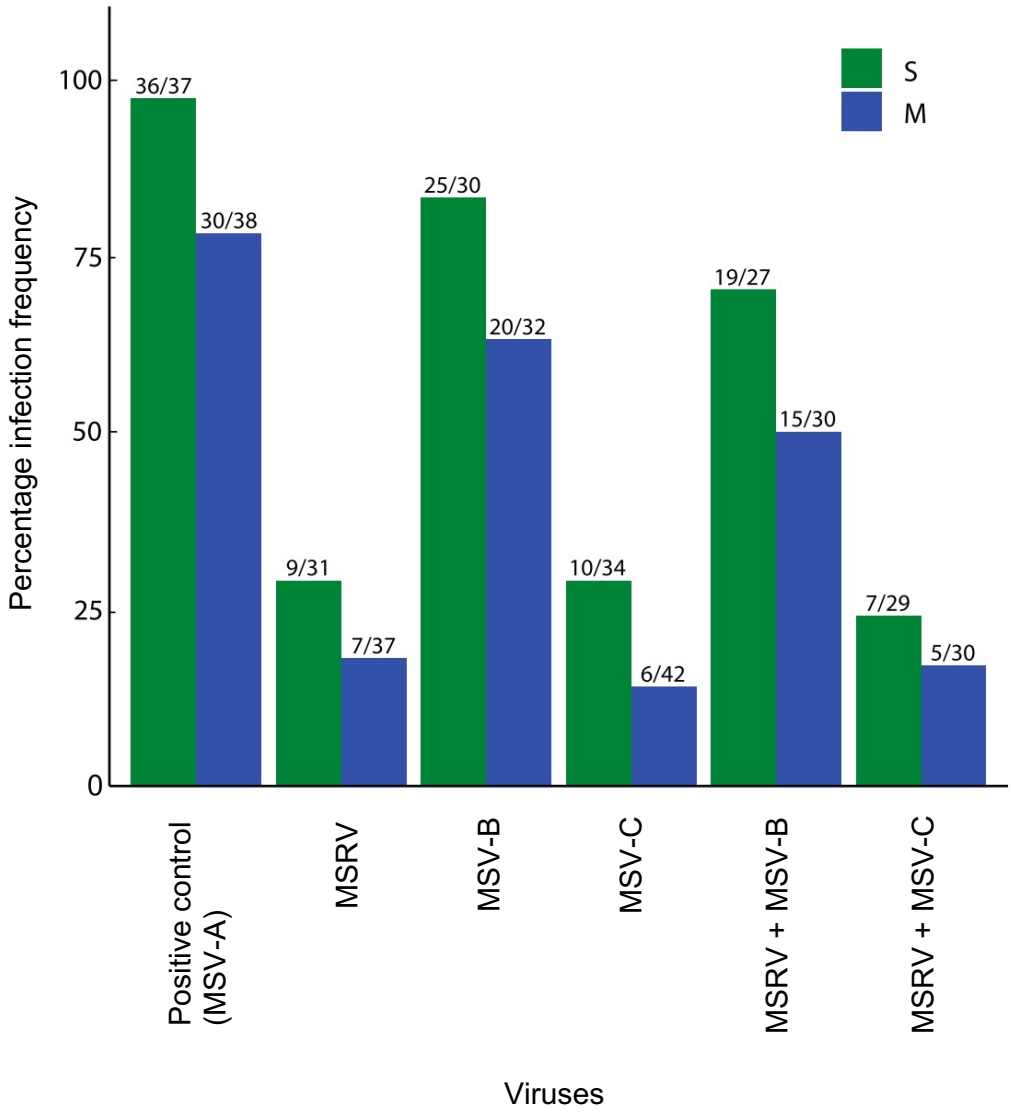


Figure 2.1: Virus infection frequencies in S and M maize genotypes. Fractions above the bars indicate the proportions of all plants surviving the *Rhizobium*-inoculation process that developed symptoms.

2.4.2 Infection symptom and streak morphology

Observed symptom severities induced by MSV-C and MSRV were much lower than those induced by the MSV-A isolate in both the S- and M- maize genotypes (Figure 2.2). The MSV-C isolate produced short, narrow pale-yellow chlorotic streaks that became less discernible on successive emerging leaves (Figure 2.2). The MSRV isolate produced distinctive striate symptoms in maize that were similar in pattern, albeit not in severity, to those seen in the field personal observation (Pande *et al*., 2012) and are also similar to those produced by a distantly related MSRV-like mastrevirus recently discovered in Ethiopia (Guadie *et al*., 2019) (Figure 2.2). Streak morphology can give insights into the virulence status of a plant virus and are usually indicative of underlying chloroplast destruction and harm to the host (Page *et al*., 1999; Charles, 2014). Chlorotic streaks can be consistently extensive and yellowish, which in more severe cases, will turn white and can be as wide as the entire leaf and unbroken for several centimetres (Van Der Walt *et al*., 2008) as observed with the positive control (MSV-A) in Figure 2.2.

2.4.3 Chlorotic area and intensity of chlorosis: MSV-C, MSRV and mixtures of MSV-C and MSRV

The percentage leaf areas that were covered by chlorotic streaks (the chlorotic area) was consistently lower than 20% for MSV-C and MSRV (Fig. 2.3 i & iii) in both the S- and M-maize genotypes. However, the intensity of chlorosis induced by these viruses was comparable to those of the MSV-A isolate in both maize genotypes. In the S-maize genotype, MSRV and mixtures of MSV-C with MSRV yielded intensities of chlorosis within chlorotic streaks on

leaves five and six that matched or surpassed those induced by the MSV-A isolate (Fig. 2.3 ii & iv).

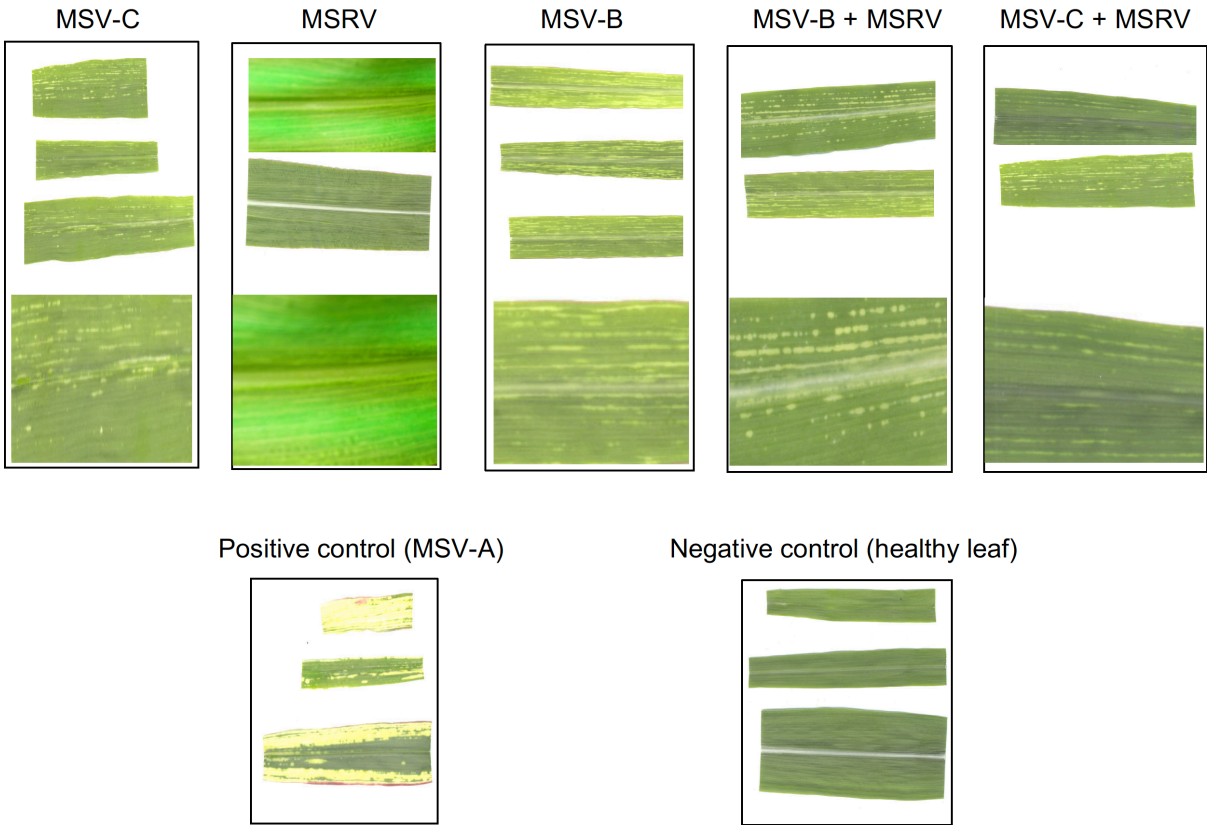


Figure 2.2: Examples of infection symptoms caused by analysed infectious virus clones in the S-maize genotype, their magnified views and controls.

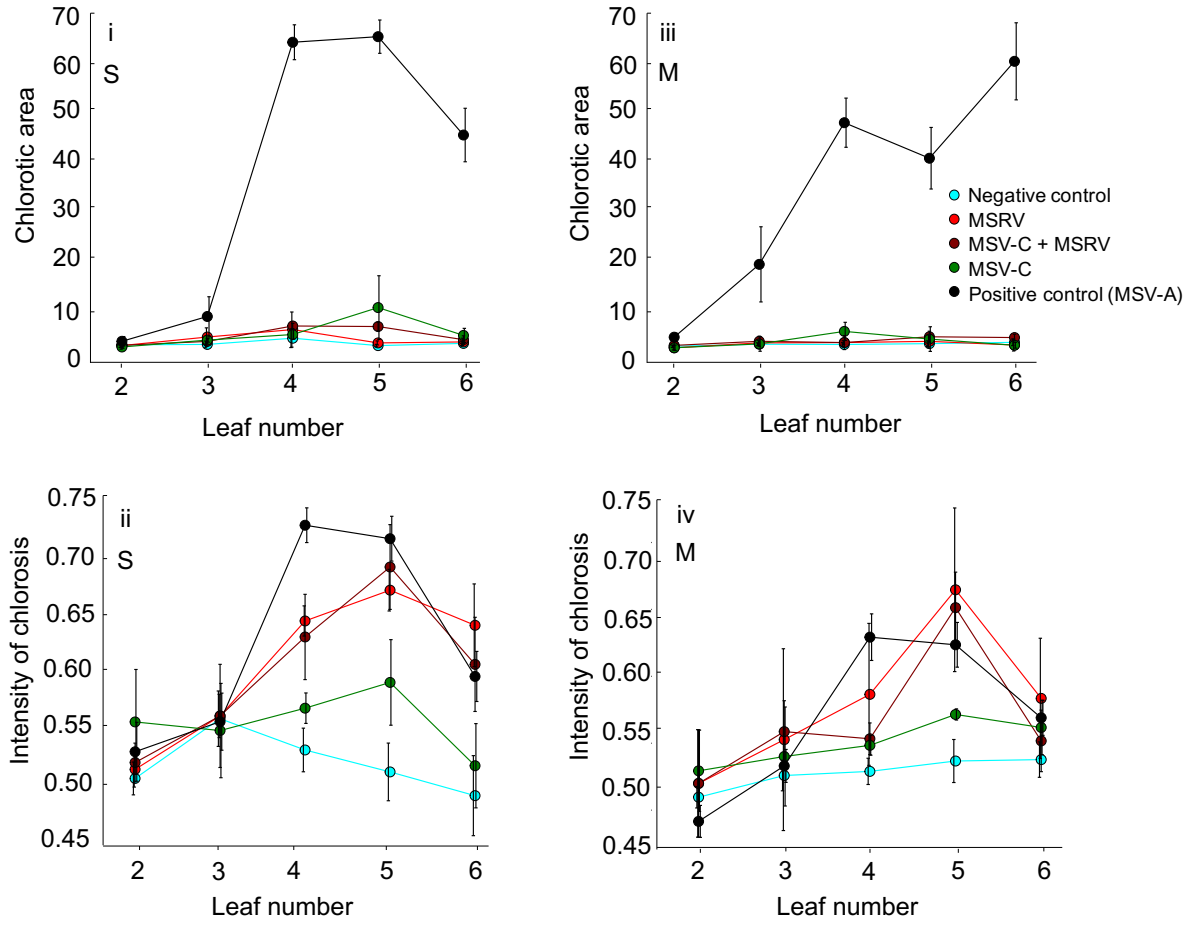


Figure 2.3: Chlorotic area and chlorotic intensity symptoms induced by infectious virus clones in maize. Percentage chlorotic areas (i and iii) and intensities of chlorosis (ii and iv) induced following successful infections of S and M maize genotypes with MSV-C, MSRV and mixtures of MSV-C and MSRV. Error bars represent the 95 % confidence interval of the mean.

2.4.4 Chlorotic area and intensity of chlorosis: MSV-B, MSRV and mixtures of MSV-B and MSRV

In S- maize genotype inoculated with MSV-B, the chlorotic area gradually increased towards leaf 6 and covering about 20% of the leaves. Chlorotic areas observed for MSRV and the mixtures of MSV-B and MSRV were higher than that of M- maize genotype. Statistically significant chlorotic area increases ($P = 0.032$) relative to MSRV on its own were also observed in the S-genotype for mixtures of MSV-B and MSRV (Figure 2.4 i & iii). However, the same cannot be said of the M- maize in which the chlorotic areas induced by the viruses fell under 20%. Chlorotic intensities ranges were like that of MSV-C and comparable to that of MSV-A (Figure 2.4 ii & iv). It is noteworthy that in both S- and M-maize plants displaying symptoms following mixed infections of MSV-B and MSRV, intensities of chlorosis were on average both similar to those observed in plants infected with MSRV alone, and significantly higher than those observed in plants infected with MSV-B alone (Figure 2.4 ii & iv).

While chlorotic areas induced by an MSV isolate indicate the numbers of chlorenchyma cells which the virus isolate is able to infect (because the virus is only found within chlorotic lesions; Lucy *et al*., 1996), chlorotic areas are likely also a proxy measure for virus transmissibility in that viruses can only be transmitted by their insect vectors from chlorotic lesions (Peterschmitt *et al*., 1992). Chlorotic intensity, on the other hand indicates the degree of chloroplast destruction that occurs in infected cells (Pinner *et al*., 1993; Monjane *et al*., 2014a). By modulating the colour of symptoms, chlorotic intensity might also directly impact transmission probabilities because insect vectors are differentially attracted to green and yellow leaves (Hodge & Powell, 2008; Moreno, 2009; Storey, 1938)

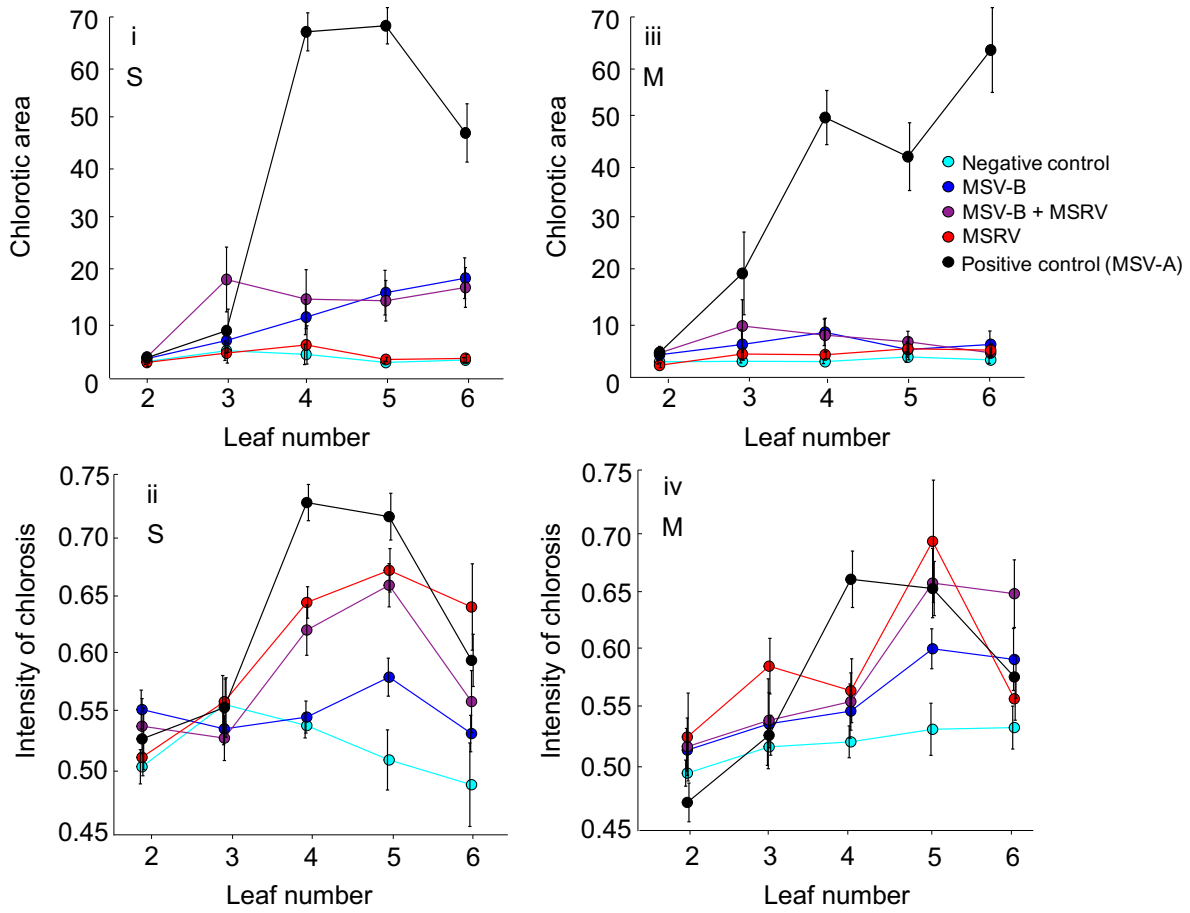


Figure 2.4: Chlorotic area and chlorotic intensity symptoms induced by infectious virus clones in maize. Percentage chlorotic areas (i and iii) and intensities of chlorosis (ii and iv) induced following successful infections of S and M maize genotypes with MSV-B, MSRV, and mixtures of MSV-B and MSRV. Error bars represent the 95 % confidence interval of the mean.

The observations that (i) MSV-B and MSRV isolates frequently occur together within mixed infections in the field, (ii) that MSRV appears to determine the intensity of chlorosis in mixed laboratory infections with both MSV-B and MSV-C, (iii) and that MSV-C and MSV-B appear to determine the chlorotic areas induced during mixed laboratory infections with MSRV, represent circumstantial evidence that synergism likely occurs during mixed infections of MSV-B or C with MSRV. The plausibility of such a synergistic relationship is further supported by the fact that such relationships are known to occur in other geminiviruses (Fondong *et al*., 2000; Acosta-Leal *et al*., 2011).

It has recently been suggested that, over the 100 years following the emergence of the MSV-A strain as a serious maize pathogen (Monjane *et al*., 2020), adaptation of the virus to infecting maize involved the concomitant induction of increased chlorotic areas and decreased chlorotic intensities. It is noteworthy that the chlorotic intensities induced in maize by MSRV alone and in mixed infections with MSV-C or MSV-B are, on some leaves at least, comparable to those induced by the MSV-A isolate that we analysed. This suggests that for MSRV the possibility exists for this virus to amplify the amount of chloroplast destruction that occurs during mixed infections of maize with MSV variants which could, in turn, increase the amount of harm that otherwise mild MSV variants would inflict on maize plants.

To the best of my knowledge, this is the first study to characterise the ability of MSRV to infect maize via *Rhizobium*-inoculation with infectious viral genome clones. Given both that I have only examined individual clones, and that the MSRV clone examined was not itself taken from a plant displaying severe MSD symptoms, we must be cautious to not over-generalize the meaning of these findings. In this regard it should be emphasized that Koch's postulate remains unfulfilled in that a causative relationship between severe MSD and MSRV or MSV-C

infections was not established. I have nevertheless demonstrated that these viruses can elicit mild symptomatic infections in maize. As such it is plausible that spill-over infections of maize in the absence of detectable MSV-A infections is to be expected.

2.5 Conclusion

Although neither of the examined virus clones are solely capable of causing severe MSD, it also remains plausible that natural genetic variants that are closely related to these clones could be playing a role in the emergence of novel MSD-like diseases in the context of mixed infections with other mastreviruses.

2.6 Authors' contributions and acknowledgements

Main author's contribution

I conducted the experiment, generated the data, and wrote R- programming scripts I used to analyse the data. I also wrote ~80% of the manuscript.

Co-authors' contributions

1 Darren Martin had key roles in conceiving this project and supervising it, wrote about 20% of the manuscript and critically edited the manuscripts and figures.

2 Arvind Varsani had key roles in conceiving this project. He sent the virus clones for this project and continue to provide timeous, critical support with sequencing my data. He also edited the manuscripts and figures.

3 Lara Donaldson gave me access to her laboratory and critically edited manuscripts and figures.

4 Penelope Hartnady and Aderito Monjane assisted with useful guidance and tips needed for the experiment and edited the manuscripts and figures.

5 Sohini Claverie, Pierre Lefeuvre and Jean-Michel Lett contributed in conceiving approaches and also edited the manuscripts and figures.

Acknowledgements

I am thankful to everyone in Joann Passmore and Lara Donaldson laboratories and to the laboratory staff in the Molecular and Cell Biology department of UCT.

Chapter 3: Movement of the A-Strain *Maize streak virus* In and Out of Madagascar

3.1 Abstract

The A-strain *Maize streak virus* (MSV-A) is the only known cause of maize streak disease (MSD). This disease remains a persistent constraint on maize production throughout sub-Saharan Africa and its adjacent Indian Ocean islands. Regardless of the underlying drivers of MSV-A dissemination, the routes and rates of MSV-A movements across its geographical range can be inferred by using MSV genome sequence data and the times when and places where sequences were sampled. Such information is useful both for tracing the geographical origins of the viral variants that trigger sporadic MSD epidemics in particular regions, and, during epidemics, for identifying geographical regions where viruses persist within reservoirs and contribute to longer epidemics. Here, I focus on applying such analyses to determine when, where and from where MSV-A arrived on the island of Madagascar. Specifically, I use model-based phylogeographic analyses of 524 full MSV-A genome sequences, including 56 newly determined genomes from Madagascar, to reconstruct the most plausible movements that delivered MSV-A to Madagascar. I found substantial support for at least eight independent movements of MSV-A variants from East Africa to Madagascar that occurred between approximately 1990 (95% highest probability density interval [HPD], 1986 to 1995) and 2003 (95% HPD, 2000 to 2005). Conversely, I found only marginal evidence of a single instance of MSV-A moving out of Madagascar: to the Comoros Islands (95% HPD between 2003 and 2006). Whereas I inferred that across their geographical range MSV-A variants are disseminating at a median rate of 38.9 km/year (95% HPD 34.0 to 44.4), following their arrival on Madagascar, MSV-A variants have been moving at a median rate of 47.6 km/year (95% HPD 36.05 to 61.70). Human mediated factors are likely major facilitators of both sporadic long-range MSV-A movements between mainland-Africa and Madagascar and, perhaps also, short to medium range movements on the island.

3.2 Introduction

The impact of maize streak disease (MSD) on maize production is a serious threat to sub-Saharan Africa (Charles, 2014; Shepherd *et al.*, 2010) since maize is an economically important staple food in most parts of the continent. MSD epidemics occurring at a frequency of once in every five to 20 years in any given region continue to sporadically devastate maize production and remain a persistent potential trigger for future regional famines (Martin & Shepherd, 2009; Shepherd *et al.*, 2010). The epidemiological underpinnings of sporadic MSD outbreaks are complex in that they arise as a consequence of poorly understood interactions between the viruses that cause MSD - predominantly the A strain *Maize streak virus*, MSV-A (Martin & Shepherd, 2009) - the nine leafhopper species that transmit the virus (Magenya, Mueke, & Omwega, 2009; Page *et al*., 1999), and the over 120 grass species that host the virus: which besides maize include over 100 uncultivated grass species (Damsteegt, 1983).

MSV is the type member of the genus *Mastrevirus* of the family *Geminiviridae*. It occurs throughout continental Africa and on surrounding islands of the Indian and Atlantic Oceans (Charles, 2014; Martin & Shepherd, 2009). As with other mastreviruses, it has a ~2.7 kb monopartite, single-stranded DNA (ssDNA), circular genome that encodes four genes including a movement protein (MP) and a coat protein (CP) gene encoded on the virion sense genome stand, and a replication-associated protein (Rep) and a RepA gene that are expressed from the complementary sense genome strand (Martin, Willment, & Rybicki, 1999; Martin & Rybicki, 2002).

Studies comparing the genetic diversity and geographical distributions of MSV-A with both other MSV strains, and other African mastrevirus species, have indicated that MSV-A appears

to have an unusually high dispersal rate (Varsani *et al*., 2009, 2008). This higher dispersal rate might be due to MSV-A having a broader host range than many of its relatives, but it is most likely at least partially attributable to the fact that it is more likely than its close relatives to be transferred over long distances by humans (Martin & Shepherd, 2009; Varsani *et al*., 2009). In addition, compared to other related MSV strains or mastrevirus species, the widespread cultivation of maize in Africa likely provides MSV-A with a bigger, and more evenly distributed population of susceptible hosts: a factor that would be expected to impact the rate at which MSV disperses via natural insect-vector transmission (CABI; EPPO, 1997; Fajinmi *et al*., 2012).

Understanding the movement dynamics of MSV-A across Africa is crucial for the development of strategies to prevent the spread of sporadic localized MSD outbreaks to the rest of the continent. The best way to track past movements of MSV-A is via the analysis of genomic sequence data in the context of the locations and times when the genomic sequences were sampled. With this data in hand it is possible to both infer the geographical origins of MSV-A variants that are responsible for outbreaks and estimate the rates or frequencies of MSV-A movements between origin and outbreak locations (Firth *et al*., 2009; Lemey *et al*., 2009).

Although a previously published MSV-A phylogeographic analysis (Monjane *et al*., 2011) has described the emergence of MSV-A as a maize pathogen in southern Africa in approximately the 1870s and its subsequent trans-continental dissemination, these analyses did not include samples from southwest African countries such as Angola and the Democratic Republic of Congo or from northern African countries such as Sudan, Ethiopia and Egypt. They also did not include samples from many of the islands off the east coasts of Africa including the largest, Madagascar. Given that MSD is known to be endemic on Madagascar (CABI; EPPO, 1997;

Charles, 2014), it remains plausible that Madagascar could have played, and might still be playing, a hitherto unappreciated role in the evolution and dissemination of novel MSV-A variants.

Here, I examined 524 full MSV-A genome sequences (following the removal of non-MSV-A and irrelevant MSV-A sequences from West and Central Africa) that were sampled from southern Africa, East Africa and various Indian ocean islands between 1979 and 2010, together with 56 newly determined MSV-A genome sequences from Madagascar (53) and the Comoros (3) that were sampled between 2009 and 2010, and used these to estimate the numbers, timings and origins of MSV-A movements between Madagascar and continental Africa. Besides revealing additional evidence that human activities are facilitating the rapid long distance dispersal of MSV-A variants, I further show that estimated local rates of virus dissemination within Madagascar are slightly higher than those estimated for continental Africa.

3.3 Materials and methods

3.3.1 Madagascan isolates

A total of 56 MSV-A isolates were sampled between 2009 and 2010 at 12 different locations on Madagascar (n=53) and The Comoros (n=3) from maize plants displaying characteristic MSD symptoms.

3.3.2 Genome cloning and sequencing

Full-length viral genomes were amplified from total plant DNA extracts using phi29 DNA polymerase (TempliPhi; GE Healthcare) as described by Shepherd *et al*. (2008). Amplified

genome concatemers were digested with either *Kpn*1 or *BamH*I to yield linearized viral genomes (~2.7 kb) which were ligated to similarly linearized pGEMZf+ (Promega Biotech). Using primers described previously in Shepherd *et al.* (2008) and Owor *et al*. (2007), both strands of each full length genome were sequenced (Macrogen Inc, South Korea). Viral genome sequences were assembled and edited using DNAman (version 5.2.9; Lynnon Bio-soft, Canada).

3.4 Dataset preparation

3.4.1 Recombination analyses

The 56 MSV-A isolates from Madagascar and the Comoros were aligned with an additional 630 MSV-A sequences and 182 non-MSV-A sequences (representing strains of MSV-B through -K) that were publicly available from GenBank using the computer program, Muscle (Edgar, 2004). The obtained alignment was further edited by visual inspection using the computer program IMPALE (Khoosal, 2015). This MSV full genome alignment was analysed for recombination using recombination detection program (RDP) version 4.46 (Martin *et al*., 2015) to obtain a recombinant-free dataset that was subsequently used for downstream phylogenetic inferences. Screening for recombinants was accomplished using the default RDP settings and removing all evidence of recombination within non-MSV-A strains by identifying recombinant sequence fragments and removing these by replacing the corresponding sequence tracts with the standard "gap" characters, "-" without altering the alignment of the retained nucleotides.

The rationale for excluding West African sequences from the phylogeographic analyses was that within phylogenetic trees, these all formed sub-clades within larger clades of East African

isolates and none were on branches of the phylogenetic trees that were close to those on which Madagascan sequences fell: a fact indicating that there is absolutely no evidence of MSV-A movements between Madagascar and West Africa. Therefore to enable timeous convergence of Bayesian phylogeographic inferences, all of the irrelevant West African sequences were removed from subsequent phylogeographic analyses.

3.4.2 Nucleotide substitution model test

jModeltest version 2.1.10 (Darriba *et al*., 2012) implemented on the CIPRES server at http://phylo.org (Miller, Pfeiffer, & Schwartz, 2010) was used to estimate the best-fitting nucleotide substitution model for the MSV-A dataset.

3.4.3 Phylogenetic analysis

A maximum likelihood (ML) tree was constructed for the MSV-A dataset using IQ-TREE version 1.6.12 (Nguyen *et al*., 2015). Branch supports in the tree were determined using 2000 replicates of non-parametric Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT; Guindon *et al*., 2010; Shimodaira & Hasegawa, 1999) as well as 5000 ultrafast bootstrap replicates (Hoang *et al*., 2018). From branch support analyses, a 95% majority-rule consensus tree was constructed based on the 5000 bootstrap trees which was subsequently visualized using FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and the ggtree R package (Yu *et al*., 2017) using custom written R scripts.

3.4.4 Test for temporal signal

I assessed the presence of temporal signal in the MSV-A dataset using TempEst version 1.5.3 (Rambaut *et al*., 2016) on the ML phylogeny inferred with IQ-TREE to estimate the correlation coefficient (r) and the determination coefficients (R^2) of the linear regression. The *p*-values were calculated using the approach of Murray *et al*. (2016) and based on 1000 random, clustered permutations of the sequence sampling dates.

3.4.5 Phylogeographic analyses

Unlike with continuous models of spatio-temporal diffusion, discrete phylogeographic movement models do not rely on homogenous Brownian motion or a relaxed random walk process being accurate descriptors of virus movement (Lemey *et al*., 2009). Therefore when analysing virus movements over long distances (such as between continental Africa and Madagascar) discrete phylogeographic models are preferred since viruses are expected to jump in one big move from one location to the next rather than to steadily diffuse between the locations.

Discrete phylogeography models can also provide robust Bayes factor (BF) statistical support for well supported epidemiological linkages between sampling locations (Lemey *et al*., 2009). As has been described for MSV (Harkins *et al*., 2009; Monjane *et al*., 2011), along with a discrete phylogeography model, the computer program, Bayesian evolutionary analysis sampling of trees (BEAST) version 1.10.4 (Suchard *et al*., 2018) was used to estimate the MSV-A nucleotide substitution rate (per site per year; employing the best fitting nucleotide

substitution model, a strict molecular clock model, and a constant population size model), within replicate BEAST runs involving a Markov chain Monte Carlo (MCMC) length of 10^8.

To achieve sufficient mixing of the MCMC, all analyses ran up to a point at which effective sample size (ESS) estimates for all model parameters relating to the dispersal substitution rate process were above 200 (Drummond & Rambaut, 2007). Using log-combiner version 1.10.4, log and tree files that were associated with similar tree likelihood values from separate independent replicate runs were combined after discarding the first 10% of each sample of trees as burn-in. The resulting log files were further inspected in Tracer version 1.7.1 (Rambaut *et al*., 2018) and a Maximum Clade Credibility (MCC) tree was annotated using Tree Annotator version 1.10.4 and visualized using Figtree version 1.4.4.

Using the MCC tree annotated with the most probable location states of ancestral viral lineages, I visualized the MSV-A movements with an emphasis on movements in and out of Madagascar using the program, spatial phylogeographic reconstruction of evolutionary dynamics (SPREAD3; version 0.9.7.1; https://rega.kuleuven.be/cev/ecv/software/SpreaD3; (Bielejec *et al*., 2016). I calculated the Bayes factor (BF) supports for MSV-A movements between pairs of locations from the transition rates log file using a Bayesian stochastic search variable selection (BSSVS) procedure under the symmetric, reversible substitution model (Lemey *et al.*, 2009).

3.4.6 Continuous phylogeographic and *post hoc* analyses

To look at the movement dynamics of MSV-A spread within Madagascar, continuous diffusion model (Lemey *et al*., 2010) implemented in BEAST 1.10.4 (Suchard *et al*., 2018) was used to

model a spatially explicit phylogeographic reconstruction of the MSV-A dispersal history in Madagascar and the Comoros. Specifically, I used the strict Brownian diffusion model to generate posterior tree distributions within which internal nodes of trees were associated with geographic coordinates. A Brownian random walk was used to model the among-branch heterogeneity in diffusion velocity (Lemey *et al*., 2010). Continuous phylogeographic reconstructions were then accomplished for these (n = 524) sequences. Markov chains were setup to run for 10^8 generations sampling every 0.4 million states. Like the discrete phylogeographic inference, MCMC stationary convergence properties were assessed with Tracer after discarding 10% of sampled trees as burn-in. A MCC tree annotated with continuous geographical coordinate trait data was visualized with the R package, Seraphim (Dellicour *et al*., 2016). To estimate dispersal statistics, I used functions available in Seraphim to extract spatiotemporal information from 1,000 posterior, post-burn-in, trees and estimated the weighted lineage dispersal velocity, where d_i and t_i are respectively the geographic distance travelled (great-circle distance in km) and the time elapsed (in years) respectively:

$$^{V}\text{weighted} = \sum_{i=1}^{n} di / \sum_{i=1}^{n} ti$$

3.5 Results and discussion

3.5.1 Field sampling regions

Between 2009 and 2010, maize plants displaying symptoms characteristics of maize streak disease were sampled from 12 locations across Madagascar and Comoros (Figure 3.1). Sixty percent of the Madagascan sampling locations fell within active maize growing regions. Maize being a staple food in the southern parts of Madagascar is mostly cultivated in the high plateaus (Bjarnason, 1986). Overall, these maize growing regions have similar tropical, climatic conditions. Maize plants in the eastern part of the island generally experience higher MSD

loads, possibly due to climatic conditions on this side of the mountain being more conducive to the spread of MSD (Peel, Finlayson, & Mcmahon, 2007) (see Supplementary Table 1).

Full genome sequences were determined from clones of 56 sampled MSV-A isolates previously sampled by Jean-Michel Lett (CIRAD, Reunion), cloned by Arvind Varsani (University of Cape Town) and sequenced by Macrogen Inc. (South Korea). These genomes were all clearly MSV-A isolates based on relatedness to previously determined MSV-A and, among MSV-A isolates they seemed to be most closely related to those in the MSV-A_1 subtype.

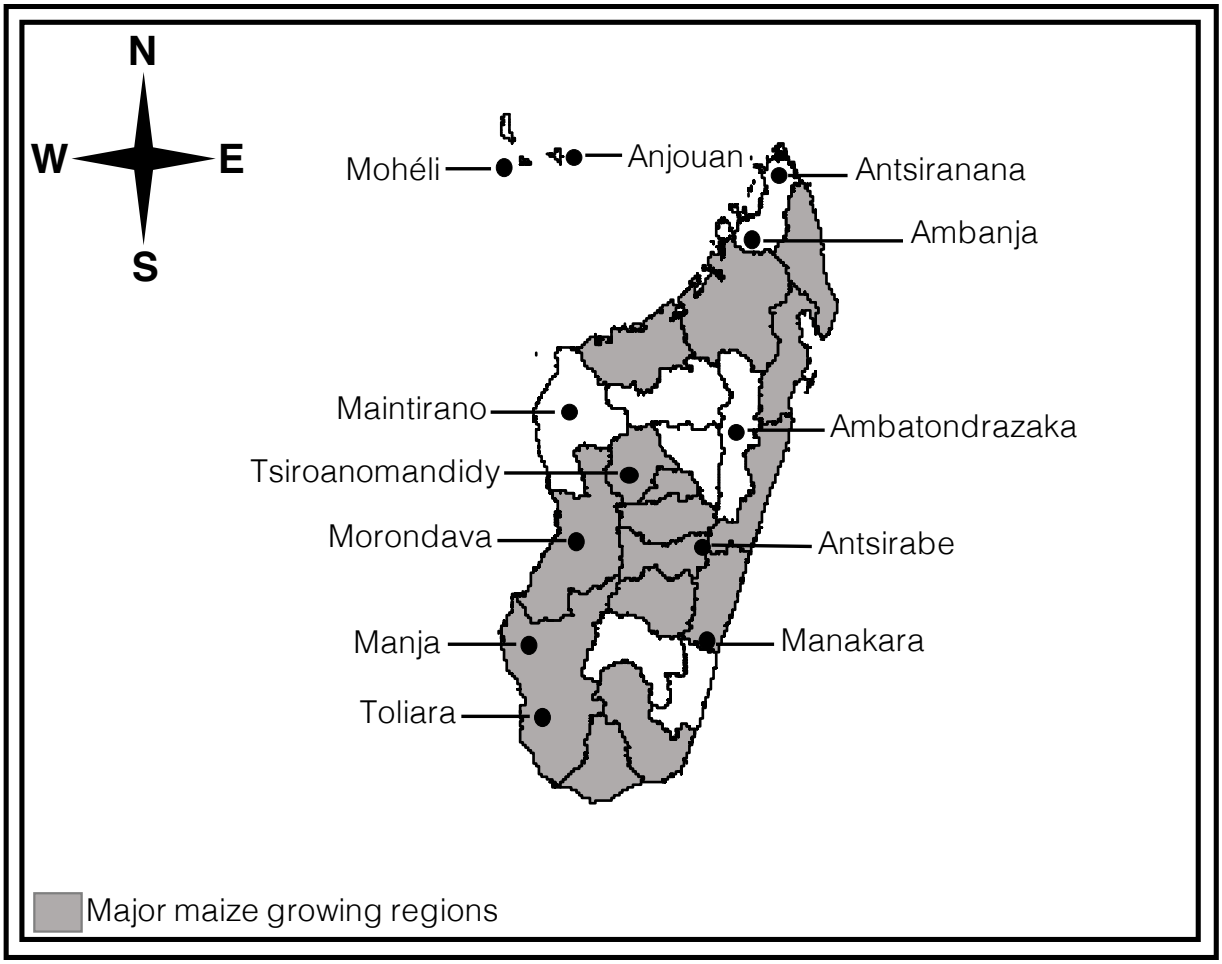


Figure 3.1: Maize streak virus sampling sites across Madagascar and the Comoros

3.5.2 Phylogenetic evidence of multiple MSV introductions to Madagascar from Africa

Maximum likelihood tree was constructed to illustrate the evolutionary relationships between the novel Madagascan and Comoros MSV-A isolates with MSV-A variants sampled in Africa and the islands of Réunion and Mauritius. This tree was also used for both selecting the best-fitting nucleotide substitution model and testing the temporal signal of the MSV-A dataset. The best fitting DNA substitution model for the full genome MSV-A alignment was GTR+I+G4: which is consistent with previous MSV-A phylogeographic studies (Harkins *et al*., 2009; Monjane *et al*., 2011). Here, GTR represents the empirical DNA model of Tavare, (1986), which permits unequal rates for all six of the possible reversible nucleotide substitutions and unequal base frequencies, with I representing the proportion of invariant sites and G4 the discrete Gamma model of Yang, (1994) with four rate categories (proportion of sites: relative rate = 0.1568: 0.05669 for category 1; 0.1568: 0.4136 for category 2; 0.1568: 1.321 for category 3; and 0.1568: 4.586 for category 4). The ML tree constructed from the MSV-A sequences is presented in Figure 3.2. Likewise, a ML tree indicating the sampling year of the represented MSV-A isolates is given in Supplementary Figure 1.

MSV-A has been classified into subtypes with isolates sharing >98% sequence similarity generally being assigned to the same subtype. These subtypes are mostly designated for convenience in describing the relationships of the sequences but they are also reflective of the severity of symptoms produced by different groups of MSV-A isolates (Martin *et al*., 2001) and the spatial distributions throughout Africa of the isolates (Harkins *et al*., 2009; Monjane *et al*., 2011)

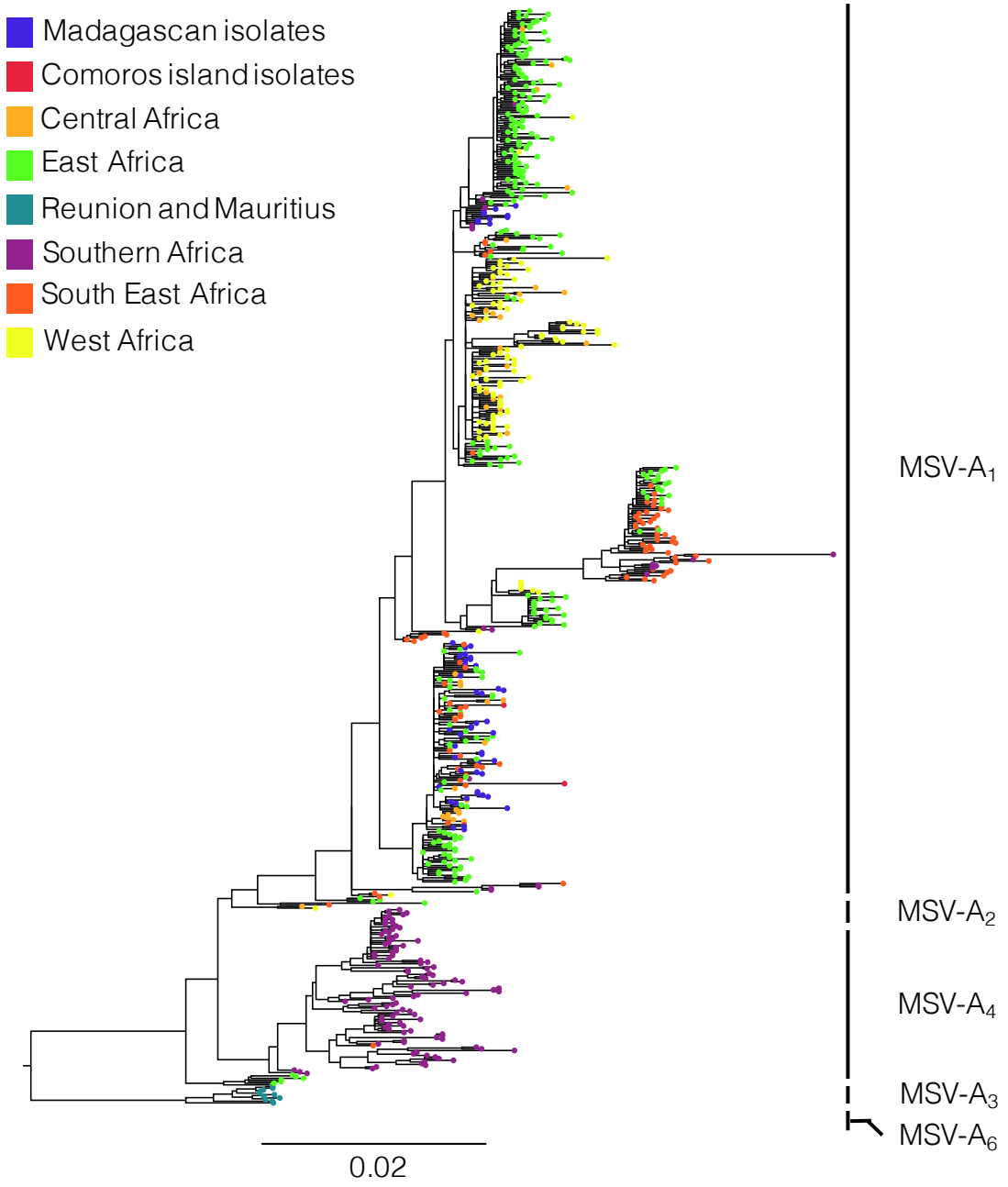


Figure 3.2: MSV-A maximum likelihood phylogeny. The phylogeny (model GTR+I+G_4) was rooted using the outgroup Reunion and Mauritius isolates and coloured according to regions of sampling origin. Clade labels indicate the MSV-A subtypes. The bar indicates the number of nucleotide substitutions per site.

Martin *et al*. (2001); Owor *et al*. (2007) and Monjane *et al*. (2011) further classified MSV-A isolates in subtypes A_1 and A_4 based on patterns of past intra- and inter-strain recombination, into 19 MSV-A recombinant lineages, 14 of which are comprised predominantly of MSV-A_1 derived sequences. Based on this classification it was determined that the Madagascan isolates all belong to the MSV-A_1 subtype and, further that collectively they represent three distinct recombinant lineages: XI, IV and V; all of which have been previously detected in East Africa (see Supplementary Table 1).

These recombinant lineages all descended from intra-MSV-A strain recombinants (i.e viruses in the MSV-A strain that arose through recombination between two MSV-A viruses). Such recombinants are relatively common with, for example, more than half of isolates sampled from Uganda in 2005 (Owor *et al*., 2007), belonged to recombinant lineages similar to those to which the Madagascan viruses belong.

Also, the discovery that the Madagascan MSV-A viruses appear to be examples of recombinant lineages first discovered in Uganda is consistent with other evidence suggesting that East Africa in general, and the Great Lakes region in particular, plays a key role in both the genesis of new MSV-A variants, and in their dissemination to other parts of the continent (Monjane e*t a*l., 2011). East Africa is most likely the source of MSV-A_1 lineages that has been found throughout continental Africa- and my discovery here that this subtype is also present on Madagascar further confirms its importance as the predominant cause of severe MSD throughout the geographical range of MSV.

3.5.3 Sufficient temporal signal

Prior to inferring time-calibrated phylogenetic trees within the Bayesian genealogical inference framework for a rapidly evolving virus like MSV, it is important to test for sufficient temporal signal present in the dataset, which is the divergence accumulating between samples over successive sampling years. Here, the evolutionary rate is an estimable parameter that is equivalent to the slope of divergence through time that enables rescaling of the tree from units of genetic distance to time. These analyses indicated that the sequences did indeed have a significant degree of temporal signal (R^2 = 0.111, p-value < 0.001), with a correlation coefficient of 0.333 for a regression of sampling dates against root to tip divergence estimated from the TempEst program (Rambaut *et al*., 2016). This degree of temporal signal together with heterochronous sampling (i.e. MSV-A samples having been collected over ~35 years) indicated that the data was suitable for phylogenetic reconstructions under molecular clock models (see Supplementary Figure 2).

3.5.4 Estimating the MSV-A substitution rate

The rate of substitution indicates how fast or slow a pathogen, a gene, genome or population accumulates novel nucleotide substitutions. I estimated that the mean substitution rate of MSV-A since the most recent common ancestor of all analysed MSV-A isolates was 4.8687×10^{-4} substitutions per genome site per year (95% highest probability density interval [HPD] 4.0924×10^{-4} to 5.5181×10^{-4}). This estimate lies within the confidence intervals of previous estimates of MSV-A substitution rates determined both from similar model-based analyses of MSV-A genome sequences sampled from nature (Harkins *et al*., 2009; Monjane *et al*., 2011; Pande *et*

al., 2017) and in evolution experiments lasting between one and six years (Harkins *et al*., 2009a).

3.5.5 Where, when and how many times has MSV-A been introduced into Madagascar?

To determine where MSV-A was introduced into Madagascar, I analysed the full genome dataset (n = 524) including isolates from East Africa, southern Africa, Madagascar, Reunion, Mauritius and the Comoros. My analyses incorporated discrete sampling year and sampling country metadata to respectively inform strict molecular clock and discrete phylogeography models which would enable the inference of where geographically ancestral sequences occurred and when over the past decades ancestral sequences moved to Madagascar from the regions where viruses were sampled. The time-scaled phylogeographic history of MSV-A representing its movements between these regions was summarised as a maximum clade credibility (MCC) tree (Figure 3.3). All eight of the tree nodes that represented ancestral sequences of monophyletic Madagascan MSV-A clades are at the tips of branches with posterior probability supports > 0.7 (see arrows in Figure 3.3). The most probable locations from which the ancestral viruses entered Madagascar were all East African (i.e. Kenya, Uganda, Tanzania or Ethiopia) with posterior probability supports ranging between 0.71 and 1.0. (Table 3.1)

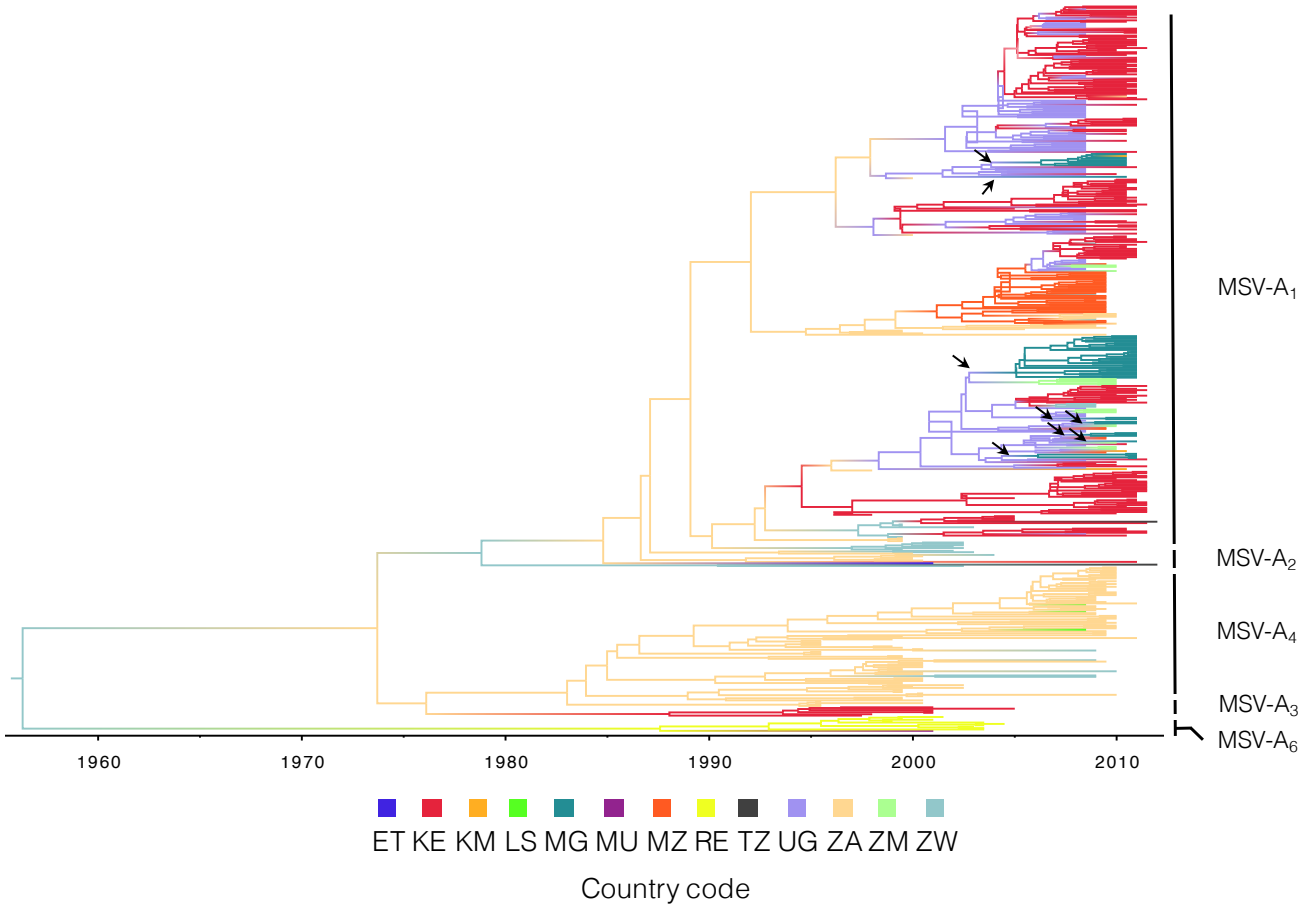


Figure 3.3: MSV-A maximum clade credibility (MCC) phylogeny generated under the discrete phylogeographic diffusion model. Branches are coloured according to either the most probable country where ancestral sequences represented by the branches existed (for the internal branches), or by the country where isolates were sampled (for the terminal branch tips). Clade labels indicate the MSV-A subtypes. Colour changes along branches are indicative of ancestral virus movements or transitions between countries. Branches representing movements between East Africa and Madagascar are indicated by black arrows. Two-letter country codes are as follows: Ethiopia [ET], Kenya [KE], Comoros Island [KM], Lesotho [LS], Madagascar [MG], Mauritius [MU], Mozambique [MZ], Reunion Island [RE], Tanzania [TZ], Uganda [UG], South Africa [ZA], Zambia [ZM], and Zimbabwe [ZW].

Table 3.1: **MSV-A movements in and out of Madagascar**

Movement	Posterior probability	Year (95% HPD)
Inward (East Africa to Madagascar)		
1	0.98	1990 (1986 to 1995)
2	1.00	1993 (1991 to 1994)
3	0.96	1995 (1993 to 1996)
4	1.00	1996 (1991 to 2000)
5	0.71	2000 (1996 to 2003)
6	1.00	2001 (1998 to 2004)
7	0.82	2002 (1998 to 2005)
8	0.72	2003 (2000 to 2005)
Outward (Madagascar to Comoros Islands)		
1	1.00	2005 (2003 to 2006)

The first detectable movement from East Africa to Madagascar of any of the MSV-A sequences analysed here was estimated to have occurred in 1990 (95% HPD, 1986 to 1995) whereas the most recent of the detectable movements from East Africa, was inferred to have occurred in approximately 2003 (95% HPD, 2000 to 2005).

When considering the best supported MSV-A movement pathways, which included five within eastern Africa, southern Africa and the Indian Ocean islands (Figure 3.4 A & B), the most important in the context of our study was the East Africa to Madagascar pathway which had an overwhelming degree of Bayes factor support (BF >10; Figure 3.4 B). Only one MSV-A movement out of Madagascar (to the Comoros islands) was detected (PPS = 1.0). This movement occurred in approximately 2005 (95% HPD, 2003 to 2006).

Frequent MSV-A movements from East Africa to southern Africa and West Africa have been previously inferred (Monjane *et al*., 2011), with the region of Uganda bordering Lake Victoria being a prominent hotspot of MSV-A dissemination. Another previously identified factor associated with the dissemination of MSV-A across continental Africa over the past 50 years has been the genesis in, and spread from, the great lakes region of recombinant MSV-A lineages (Monjane *et al*., 2011; Owor *et al*., 2007). The discovery here that, in 2010, Madagascan MSV-A populations were dominated by multiple recombinant lineages that had all originated in East Africa further emphasises the important impact that this region has on MSV epidemiology across the geographical range of the virus.

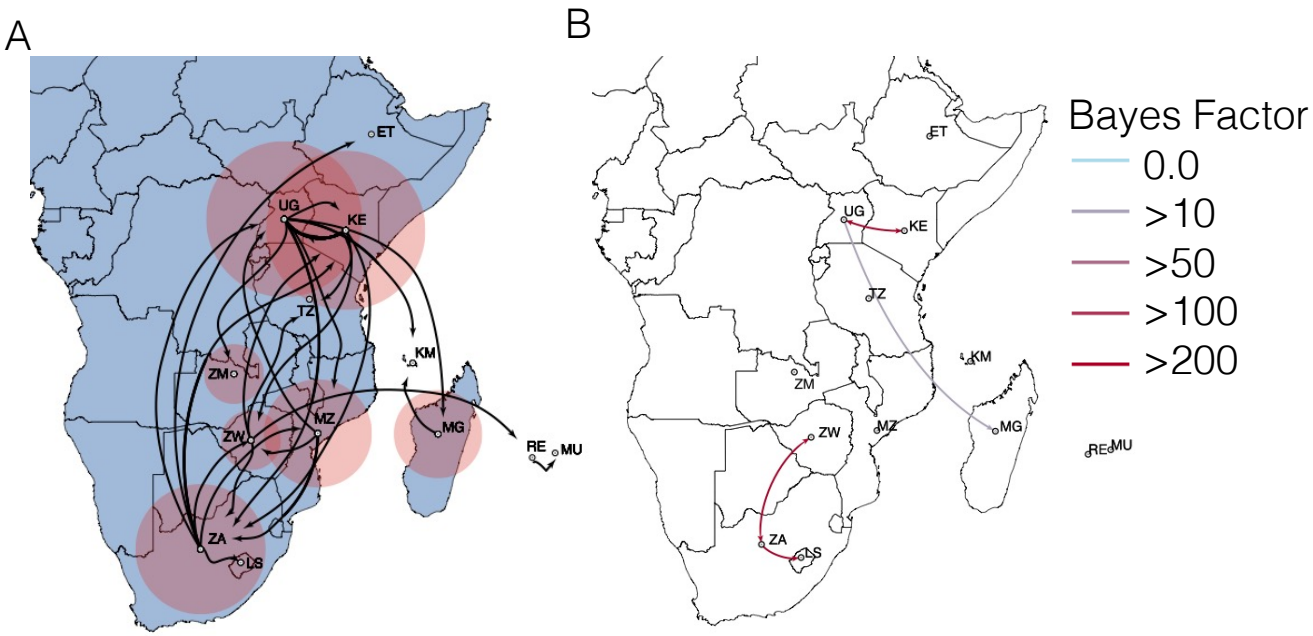


Figure 3.4: Geospatial representations of MSV-A movements in and out of Madagascar. The countries of origin are labelled with their International Organization for Standardization (ISO) two-letter codes as follows: Ethiopia [ET], Uganda [UG], Kenya [KE], Tanzania [TZ], Zambia [ZM], Comoros Island [KM], Mozambique [MZ], Madagascar [MG], Reunion Island [RE], Mauritius [MU], Zimbabwe [ZW], South Africa [ZA] and Lesotho [LS]. (A) Extremely well supported MSV-A migration routes within eastern, Southern Africa and the Indian Ocean inferred with the discrete phylogeographic diffusion model. Arrows indicate movement directions. (B) Bayes factor supports for the four best supported movement pathways as inferred using a Bayesian Stochastic Search Variable Selection (BSSVS) procedure under the symmetric substitution model.

Although the MSV-A variants studied here first entered Madagascar in approximately 1990 it is very unlikely that these lineages were the first to make it to Madagascar. Although MSD was first reported on Madagascar in 1974 (Autrey, 1983; Bock, 1974; Soto, Buddenhagen, & Asnani, 1982) it is plausible that it may have been present on the island for decades before these reports. It is similarly, difficult to say when MSD first reached the Comoros islands and may have been present there for as long as it has been present on Madagascar. Therefore although I inferred a MSV-A movement from Madagascar to the Comoros in approximately 2002 (95% HPD 1997 to 2005) it is very unlikely that this was the first introduction of MSV-A to the island chain.

Natural disasters such as adverse weather conditions, hurricanes, droughts are quite common in the Comoros, as well as inadequate land mass for farming has meant that the island chain tends to frequently import goods from their trade partners (Comoros, 2014; Congress, 1994). Food imports constitute 32-70% of total imports from their foreign partners including from Madagascar (Mundi, 2019a; World Integrated Trade Solution, 2018). Rice imports alone, for instance, accounted for 30% of food imports between 1980 and 1987 (Congress, 1994); while food and vegetable imports from Madagascar, went from 9.01% to 2.28% of total imports between 1990 and 2010 (World Integrated Trade Solution, 2018). In 2002, the percentage vegetable imports shares for Kenya, South Africa and Madagascar were 1.09%, 2.19% and 4.54% respectively. Following the very first implementation of trade liberalizations by Madagascar in 1982 (WTO, 2001), Madagascar further participated in several other regional trade agreements with the member states of East African Community (EAC) and the Indian Ocean Commission (IOC) which may have played a key roles in determining the frequency with which viruses such as MSV were inadvertently transported to and from Madagascar (Adar, 2011; Wangia, Wangia, & Groote, 2002).

Madagascar experienced rapid import growth in the 1990s through 2020 which was probably linked to increased air and sea traffic from mainland Africa and other parts of the world (USDA, 2020; World Integrated Trade Solution, 2018): air traffic specifically increased very rapidly from 1992, achieving an all-time peak in 2001 (UNdata, 2019). These years agree with this phylogeographic inference, correlate well with the Madagascan 'boom corn' years (Maret, 2007), and might played a major role in the movement of novel MSV-A_1 variants to, and its establishment on, Madagascar.

Given that MSV is not seed transmissible it is unlikely that MSV was transported to the island within grain. The virus would most likely have been introduced via live viruliferous leafhoppers, infected plant floral parts in shipping containers or airplanes, or via MSV infected cuttings of vegetatively propagated MSV host species such as sugarcane (Antwerpen *et al*., 2011; Storey, 1930). It appears East Africa might have played a key role in previous movements to the Comoros and considering the fact that these Islands and their insect fauna are closely related to those found in Africa and Madagascar (Peake, 1971), one cannot completely discount the possibility of trans-oceanic spread by wind-blown viruliferous insects. Although likely exceedingly rare, wind-blown leafhoppers are occasionally transported from Africa and Madagascar.

Despite no definitive evidence for insect transmitted viruses having been transferred in this way during modern times, there is some circumstantial evidence for a geminivirus (Creamer, 2020; Gilbertson *et al*., 2019; Maramorosch, 1980) and other insect transmitted viruses that such long journeys are possible (Radcliffe & Lagnaoui, 2007; Thresh, 1983). Brewbaker, (1978) also posited that wind-borne *Maize mosaic virus* (MMV) infected *Peregrinus maidis* leafhoppers from the Caribbean invaded a state of south-eastern Mexico after flying 1200 km

over the ocean to cause an epidemic in the late 8th century. Given that many Madagascan insects appear to be more taxonomically related to African insects than they are to those from Asia or Australia (Cassola, 2003; Donnelly & Parr, 2003), geographical proximity or nearness might be a determining factor for insect movement across the ocean (Monaghan *et al*., 2005). The Madagascan insect fauna might plausibly have ultimately involved the trans-oceanic transport to the island over millions of years for most of its endemic insect groups (Trewick, 2000). Model-based analyses of differences in the proportions of wind-blown insect-immigrants from Africa and Madagascar accounting for prevailing wind directions indicated that insect were twice as likely to be blown across the ocean from Madagascar to the Comoros as they were to be blown to the islands from mainland Africa (Peake, 1971).

Therefore, although it is possible that viruliferous insect vectors could sometimes get caught in strong updrafts and be carried by airstreams for long-enough distances to transport them from mainland Africa to Madagascar and/or the Comoros islands, the frequencies of such events are unlikely to be anywhere near high enough to account for the eight independent introductions of MSV-A variants to the island between 1990 and 2003 that I have detected. It is far more likely, therefore, that rather than being blown there by winds, MSV-A is being regularly transported to the islands from Africa within viruliferous leafhoppers carried by boats and/or aeroplanes.

Curiously, one of the most important MSV-vector species, *Cicadulina mbila* was apparently absent from Madagascar prior to 1995 (Marchand *et al*., 1995), which would suggest either that leafhoppers transferred to the islands are unable to effectively colonize the islands, or that MSV-A is transferred to the island within infected material rather than within insects. Since MSV cannot be transmitted by seed (Varsani *et al*., 2008), this material might reasonably

include immature maize cobs transported within MSV-infected leaf sheaths (CABI, 2019; Varsani *et al*., 2008), or MSV-infected cuttings of plants such as sugarcane that are known to be susceptible to MSV infection (Antwerpen *et al*., 2011; Storey, 1930).

3.5.6 The spread of MSV-A within Madagascar

In addition to examining MSV-A imports to, and exports from, Madagascar I also did an analysis of MSV-A movements within Madagascar. To achieve this, rather than doing a discrete phylogeographic analysis where MSV-A sequences might be analysed based on their country or state of origin, I carried out a continuous phylogeographic analysis where the MSV-A sequences were analysed based on the GPS coordinates where they were sampled. The mean substitution rate inferred with this continuous analysis was 4.269×10^{-4} (95% HPD 3.6186×10^{-4} to 4.9627×10^{-4}) which is slightly lower than the 4.8687×10^{-4} (95% HPD 4.0924×10^{-4} to 5.5181×10^{-4}) rate inferred using the discrete analysis. Accordingly, the inferred root date for the most recent common ancestor of all the MSV-A sequences used for this analysis was 1882 (95% HPD 1861 to 1901); a date which is earlier than the date of the MSV-A most recent common ancestor (MRCA), 1900 (95% HPD 1882 to 1916), inferred using the discrete analysis.

Consistent with the results of the discrete phylogeographic analyses, spatially explicit, continuous phylogeographic reconstructions of MSV-A dispersal within Madagascar over the last decade, (Figure 3.5) revealed at least five introductions from mainland Africa occurring in 1990 (95% HPD 1988 to 1993), 1991(95% HPD 1983 to 1998), 1995 (95% HPD 1992 to 1999), 1995 (95% HPD 1991 to 1999) and in 1999 (95% HPD 1994 to 2002). An introduction of MSV-A to the Comoros from Madagascar was inferred to have occurred in 1999 (95% HPD

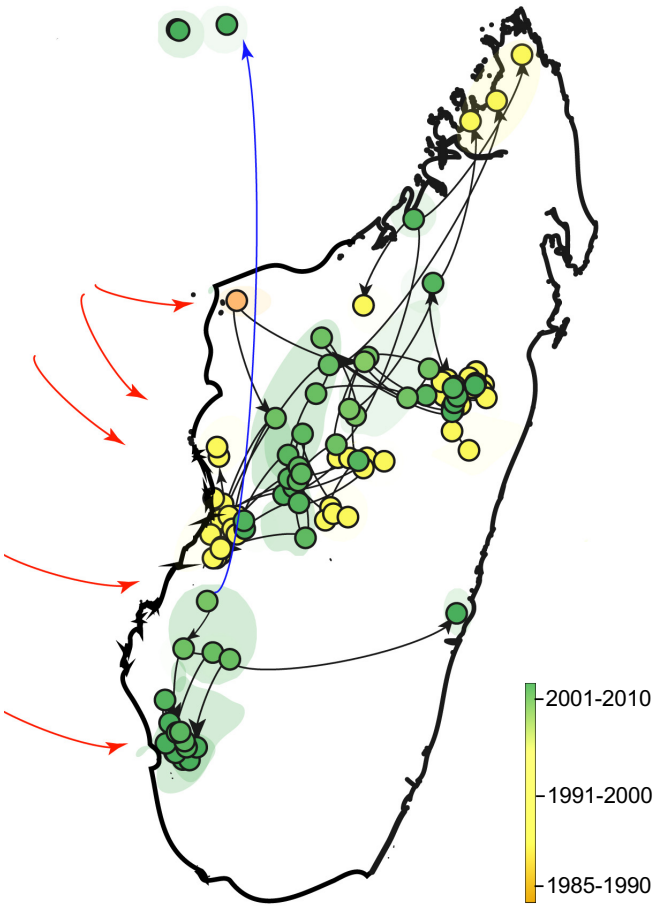


Figure 3.5: Spatial phylogeographic reconstruction of MSV-A dispersal history within Madagascar and the Comoros. The maximum clade credibility (MCC) tree and overall 80% highest posterior density (HPD) regions representing the uncertainty related to the phylogeographic inference. MCC tree nodes were coloured according to their time of occurrence. Arrows indicate the direction of movements. Red and blue arrows represent in and out movements respectively.

1994 to 2002), approximately six years earlier than that inferred with discrete phylogeographic analysis. Although fewer independent MSV-A introductions to Madagascar were inferred by this continuous analysis than were inferred by the discrete analysis, the estimated timings of all five of the introductions inferred by the continuous analysis fell within the estimated time window inferred for the eight introductions inferred by the discrete analysis.

This continuous phylogeographic analysis also revealed that following the introduction of MSV-A to Madagascar the virus appears to have disseminated at a rate of approximately 47.6 km/year (95% HPD 36.05 to 61.70). This dispersal rate is a little faster than the approximately 38.9 km/year (95% HPD 34.0 to 44.4) that I and others (Monjane *et al*., 2011) estimate for MSV-A dispersal rates on mainland Africa. Caution must however be taken against over-interpreting the meaning of this inference since dispersal rate estimates may be strongly impacted by sample sizes and the strengths of temporal signal within the analysed nucleotide sequence datasets (Lemey *et al*., 2010).

The continuous phylogeographic analysis also yielded some clues as to the regions where the analysed MSV-A lineages may have entered Madagascar. Specifically, the analysis provide clues as to the roles played by ports of entry during the introduction of MSV-A onto the island. Key sampling regions in the western and southern parts of Madagascar included major ocean ports where active imports and exports of good were occurring throughout the timeframe when MSV-A imports were inferred to have occurred using my continuous phylogeographic analysis.

Further, these ports of entry are actively linked to ports of export in East Africa. Specifically, air and seaports in Toliara (Southwest Madagascar) played a crucial role in the ‘boom corn’

years of 1980s-90s in Madagascar (USAID, 2019; Wetterdienst, 2017) and the locations of these ports correlate spatially with my inference of the region in Madagascar where one of the inferred MSV-A imports occurred. I can reasonably speculate following entry of MSV-A at these western Madagascan ports the virus spread to other regions.

3.6 Conclusion

These inferred MSV-A movements into Madagascar over the past 30 years, all appear to have occurred from East Africa. Factors that strongly influence MSV-A movements from East Africa to Madagascar are likely more a consequence of human and socio-political mediated factors than natural, ecological events given the low probabilities of viruliferous insect vectors having been blown across the ocean between mainland Africa and Madagascar.

The current coronavirus pandemic has illustrated that island nations are ideally positioned to prevent the importation of pathogens. Key measures that might limit pathogen importations in the future include bio-sanitation rules such as restrictions on the importation of cuttings of MSV-susceptible plant species such as sugarcane until these are proven to be MSV-free. Steps can also be taken to kill leafhoppers within shipping containers and aeroplane cabins. Further, the limitation of illegal imports might also help to curb the transmission of MSV-A and other plant diseases to the island.

3.7 Authors' contributions and acknowledgements

Main author's contribution

I carried out all of the phylogenetic and phylogeographic analyses (IQ-Tree, Tempest, Beast, and Seraphim) and wrote R scripts used for data visualization. Lastly, I wrote ~80% of the manuscript.

Co-authors' contributions

1. Darren Martin inspired this project and supervised it. He wrote about 20% of the manuscript and provided critical advice on editing the text and figures.
2. Arvind Varsani inspired this project, supervised it and carried out the cloning and sequencing of all the Madagascan samples.
3. Gordon Harkins provided the MSV-A dataset and supervision.
4. Jean-Michel Lett carried out the sampling survey in Madagascar

Acknowledgements

I thank Mark Miller for providing access to CIPRES servers at www.phylo.org and useful tips for Beast analysis.

Chapter 4 : Movement of *Maize streak virus*-A In and Out of Ethiopia and Rwanda

4.1 Abstract

The A-strain of *Maize streak virus* (MSV-A) causes maize streak disease (MSD), a major biotic threat to maize production in sub-Saharan Africa. Throughout this region, intermittent MSV-A epidemics every three to ten years are a major contributor to the economic and food insecurity of farmers who in some years can experience negligible losses to MSD, but in others can experience complete crop failures. Following the emergence of MSV-A in southern Africa in the early 1900s, phylogeographic reconstruction studies have consistently posited that MSV-A gradually disseminated to the other parts of Africa and was also transferred to the adjacent Indian Ocean island of La Réunion. Apart from southern Africa playing a key role in this dissemination, these studies have also indicated that East Africa is an important hub of trans-continental MSV-A movements. Further, they have also indicated that East Africa is a major hotspot of MSV-A diversification. Despite this, little is known about the MSV-A populations of East African countries such as Rwanda and Ethiopia. Here, I use a model based phylogeographic approach to analyse 792 full genome sequences including newly determined genomes from Ethiopia (n=58) and Rwanda (n=31) to reconstruct the movements of MSV-A into and out of Ethiopia and Rwanda. I found substantial statistical support for ten MSV-A movements from Kenya to Ethiopia; the first of these in approximately 1985 (95% Highest Posterior Density [HPD], 1979 to 1988) while the most recent in 2010 (95% HPD, 2009 to 2012). Four MSV-A movements into Rwanda were also all from other East African countries (two from Ethiopia, two from Kenya); all occurring between 2007 (95% HPD, 2005 to 2009) and 2010 (95% HPD, 2009 to 2011). While these movement patterns are consistent with natural dispersal of MSV-A throughout East Africa by viruliferous leafhoppers, it cannot be excluded that human activities might also be facilitating movements of MSV-A throughout the region.

4.2 Introduction

Maize streak virus strain A (MSV-A) is a mastrevirus in the Family *Geminiviridae*, and is the causal agent of maize streak disease (MSD). In sub-Saharan Africa this disease is of uttermost economic importance to large, small, and subsistence scale farmers because in epidemic years the virus can cause complete crop failures (Charles, 2014; Martin & Shepherd, 2009; Shepherd *et al*., 2010). Infected maize plants display characteristic chlorotic streaks on their leaves, are severely stunted, have low yields and occasionally die (Martin *et al*., 2001; Martin, Willment, & Rybicki, 1999; Pinner *et al*., 1988). Most farmers in sub-Saharan Africa lack the resources to use insecticides (which control the insect vector of MSV) or to perchance, and sow MSV-resistant hybrid seed. This, coupled with limited access to climatic mitigation strategies for minimizing the impacts of droughts and flooding, means that maize farmers in low-income sub-Saharan African countries are twice as likely to experience near total maize crop losses to MSD than farmers in middle income African countries (Alegbejo & Banwo, 2005; Bosque-Pérez, 2000; Thresh, 2004).

The epidemiology and spread of MSV-A is complex and is strongly dependent on the behaviours and population dynamics of the *Cicadulina* sp. that transmit it (Lefeuvre *et al*., 2019; Magenya, Mueke, & Omwega, 2008; Reynaud & Petererschmitt, 1992; Shepherd *et al*., 2010). The average reported dissemination rate of MSV-A in Africa is ~39 kilometres a year which is broadly consistent with the primary mode of MSV-A movement being via natural insect movements: although occasional long-distance movements either across the continent or to adjacent islands of the Indian Ocean strongly suggests that human activities are also likely at least occasional drivers of MSV-A dissemination.

Human mediated factors such as trade and immigration policies have played key roles in the economies of African countries by ensuring free movements of goods and more liberal commerce (Adar, 2011; Barka, 2012; Wangia *et al*., 2002). These processes can also affect the rates and frequencies with which plant viruses such as MSV-A move from one place to another and between countries either within viruliferous leafhoppers carried on/in road and rail traffic, or within infected plant material such as sugarcane cuttings (although not seeds since the virus is not seed transmissible).

Ethiopia in East Africa is a large country with suitable climatic conditions for cultivating temperate and tropical crops such as maize, millet, telf, cotton, wheat, potatoes and enset (Abraham, 2019; Deribe *et al*., 2012; Mesfin, Den Hollander, & Markham, 1991). Within the country, maize is grown primarily in the southern, southwestern and western regions (Bjarnason, 1986; Dawit *et al*., 2014; Guadie *et al*., 2019). With respect to other cultivated cereals dating back to the 1980s, maize has been the most cultivated per hectare and the country has had a historical dependence on it for food and economic purposes. As a consequence of this, Ethiopia has since experienced on average annual increase in maize production of 7.60% between 1980 and 2020 (Bjarnason, 1986; Knoema, 2020a; Mesfin *et al*., 1991).

Rwanda has also recorded increases in maize output between 1980 and 2020 despite fluctuations in recent years (Knoema, 2020b). Maize was second most cultivated cereal group after sorghum and fifth in overall cultivation (Bjarnason, 1986; Batirbaev *et al*., 2013). In Rwanda, maize is cultivated in all rural regions in the country (Batirbaev *et al*., 2013). Along with recent increases in the intensity of maize farming in Rwanda and elsewhere in East Africa has come increases in the prevalence of maize diseases such as maize lethal necrosis (MLN) and MSD (Asiimwe *et al.*, 2020; Redinbaugh & Stewart, 2018). Rwanda, like Ethiopia and

other East African countries, has experienced intermittent MSD epidemics since the 1980s (Asiimwe *et al*., 2020; Martin & Shepherd, 2009; Mesfin *et al*., 1991; Owor *et al*., 2007; Pande *et al*., 2017).

MSV-A likely initially evolved in southern Africa over a century ago, (this region is its original cradle), and spread elsewhere throughout the continent at a rate that appears to have been quicker than the movement rates of both other non-MSV-A strains, and other non-MSV African streak virus species (Harkins *et al*., 2009; Monjane *et al*., 2011). During the emergence and spread of MSV throughout continental Africa from the 1920s through to the 1970s, at least five MSV-A subtypes-MSV-A_1, -A_2, -A_3, -A_4, and -A_6 evolved: each differing in virulence, and geographical range (Harkins *et al*., 2009; Martin *et al*., 2001; Monjane *et al*., 2011; Owor *et al*., 2007). Since the 1970s East Africa has become both the primary hub of trans-continental MSV-A movements, and this region is also the primary hotspot of MSV-A diversification (i.e. it is the present cradle of the virus (Monjane *et al*., 2011; Pande *et al*., 2017; Varsani *et al*., 2008). Within East Africa the regions of Uganda and Kenya around Lake Victoria appear to be the primary sources of new MSV-A variants whereas the rift valley and coastal regions of Kenya are sinks of MSV-A diversity (Pande *et al*., 2017). However, due to sparse sampling and sequencing of MSV-A variants, little is known about the roles of other East African countries such as Rwanda and Ethiopia in the genesis and dissemination of new MSV-A variants

Here, I used nucleotide sequences from a recent survey in Ethiopia and processed in this study together with those available in GenBank from Ethiopia and Rwanda to reconstruct the movement dynamics of MSV-A in and out of these countries.

4.3 Materials and methods

4.3.1 Virus sampling

A total of 84 MSV-A isolates were sampled during the 2019 planting season in Ethiopia from maize plants showing streak symptoms by Daniel Ketsela as part of an epidemiological survey.

4.3.2 Cloning and sequencing of complete MSV genomes

Full-length viral genomes and sub-genomics were amplified from total plant DNA extracts using phi29 DNA polymerase (TempliPhi; GE Healthcare) as described by Shepherd *et al*. (2008). Amplified genome concatemers were digested with *BamH*I to yield linearized viral genomes (~2.7 kb) which were ligated to similarly linearized pUC19 (New England Biolabs). Using primers described previously in Owor *et al*. (2007) and Shepherd *et al*. (2008), both strands of the full length genome and the sub-genomics were sequenced at Macrogen, South Korea. Viral genome sequences were assembled and edited using Geneious Prime (version 2020.2.4; Biomatters Limited, New Zealand).

4.4 Dataset preparation

4.4.1 Phylogenetic and recombination analyses

The 84 MSV-A isolates from Ethiopia were aligned with an additional 665 MSV-A sequences including 31 new MSV-A sequences from Rwanda, and 182 non-MSV-A sequences (representing strains of MSV-B through -K) that were publicly available in GenBank with the Muscle computer program (Edgar, 2004). The obtained alignment was further edited by visual inspection in Aliview (Larsson, 2014). This MSV full genome alignment was analysed for

recombination using RDP version 4.46 (Martin *et al*., 2015) to obtain a recombinant-free, and sub-genomics inclusive MSV-A only dataset (n = 792) that was subsequently used for downstream phylogenetic inference. Screening for recombinants was accomplished using the default RDP settings and removing all evidence of recombination within the alignment by identifying recombinant sequence fragments and removing these by replacing these sequence tracts with the standard "gap" characters, "-" without any changes to the alignment of the retained nucleotides.

4.4.2 Nucleotide substitution model test

jModeltest version 2.1.10 (Darriba *et al*., 2012) implemented on the CIPRES server at http://phylo.org (Miller *et al*., 2010) was used to determine the best-fitting nucleotide substitution model for the MSV-A dataset.

4.4.3 Phylogenetic analysis

Maximum likelihood (ML) tree was constructed for the MSV-A dataset using IQ-TREE version 1.6.12 (Nguyen *et al*., 2015). Branch supports in the tree were determined using 2000 replicates of non-parametric Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT) (Guindon *et al*., 2010; Shimodaira & Hasegawa, 1999) as well as 5000 ultrafast bootstrap replicates (Hoang *et al*., 2018). A 95% majority-rule consensus tree was constructed based on the 5000 bootstrap trees and this was subsequently visualized using FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/)

4.4.4 Test for temporal signal

I tested for the presence of temporal signal in the MSV-A dataset for phylogeographic analysis using TempEst version 1.5.3 (Rambaut *et al*., 2016) on the ML phylogeny inferred with IQ-TREE to estimate the correlation coefficient (r) and the determination coefficients (R^2) of the linear regression. The *p*-values were calculated using the approach of Murray *et al*. (2016) and based on 1000 random, clustered permutations of the sequence sampling dates.

4.4.5 Phylogeographic analyses

Discrete phylogeographic models, being quite useful for routine monitoring of outbreaks, can also serve as key response measures involving an ongoing sampling and genome sequencing study in which pathogen spread is modelled between distant locations with a measure of Bayes factor (BF) statistical support between these locations (Lemey *et al*., 2009). Due to the unavailability of associated GPS coordinate data at the time of analysis, only discrete phylogeographic analysis was accomplished. As has been described for MSV (Harkins *et al*., 2009; Monjane *et al*., 2011), along with a discrete phylogeography model, BEAST version 1.10.4 (Suchard *et al*., 2018) was used to estimate the MSV-A nucleotide substitution rate (per site per year; employing the best fitting nucleotide substitution model, a strict molecular clock model, and a constant population size model), within replicate BEAST runs involving a Markov chain Monte Carlo (MCMC) length of 10^9.

To achieve sufficient mixing of the MCMC, all analyses ran up to a point at which all effective sample size (ESS) estimates for all model parameters relating to the dispersal substitution rate process were above 200 (Drummond & Rambaut, 2007). Using log-combiner version 1.10.4,

I combined log and tree files that were associated with similar tree likelihood values from separate independent replicate runs after discarding the first 10% of each sample of trees as burn-in. The resulting log files were further inspected in Tracer version 1.7.1 (Rambaut *et al*., 2018) and a Maximum Clade Credibility (MCC) tree was annotated using Tree Annotator version 1.10.4 and visualized using customized R scripts

The MCC tree annotated with the most probable location states of ancestral viral lineages was used to visualize the MSV-A movements with an emphasis on movements in and out of Ethiopia and Rwanda using the program, spatial phylogeographic reconstruction of evolutionary dynamics (SPREAD3; version 0.9.7.1; https://rega.kuleuven.be/cev/ecv/software/SpreaD3; (Bielejec *et al*., 2016). Bayes factor (BF) supports for MSV-A movements between pairs of locations were calculated from the transition rates log file using a Bayesian stochastic search variable selection (BSSVS) procedure under the asymmetric, non-reversible substitution model (Lemey *et al.*, 2009).

4.5 Results and discussion

4.5.1 New Ethiopian isolates belong in subtype A_1 and recombinant lineage V

ML trees are useful for depicting relationships between old MSV-A isolates sampled prior to 2015 and the more recently sampled isolates from Ethiopia and Rwanda. Further, for economically important MSV-A isolates, such trees are useful for classifying newly sequenced genomes into subtype and recombinant lineages. These trees are also useful for inspecting alignments meant for downstream molecular-clock based phylogeographic analyses to determine whether the analysed sequence have sufficient temporal signal to warrant using a molecular clock model.

Here, I used the best fitting DNA substitution model GTR+I+G4 for the MSV-A (n=792) alignment as previously posited by Harkins *et al*. (2009) and Monjane *et al*. (2011). GTR is the empirical DNA model of Tavare (1986), which allows unequal rates for all six of the possible reversible nucleotide substitutions and unequal base frequencies, with I representing the proportion of invariant sites and G4 the discrete Gamma model of Yang (1994) with four rate categories. The best majority-rule consensus ML tree inferred from the MSV-A dataset is presented in Figure 4.1.

Based on degree of virulence, host range and geographical clustering (Martin *et al*., 2001; Monjane *et al*., 2011) MSV-A has been previously grouped into subtypes and recombinant lineages. Subtypes share greater than 98% sequence similarity (Fauquet *et al*., 2008; Martin *et al*., 2001). For MSV-A, five subtypes (MSV-A_1, MSV-A_2, MSV-A_3, MSV-A_4 and MSV-A_6) have been identified.

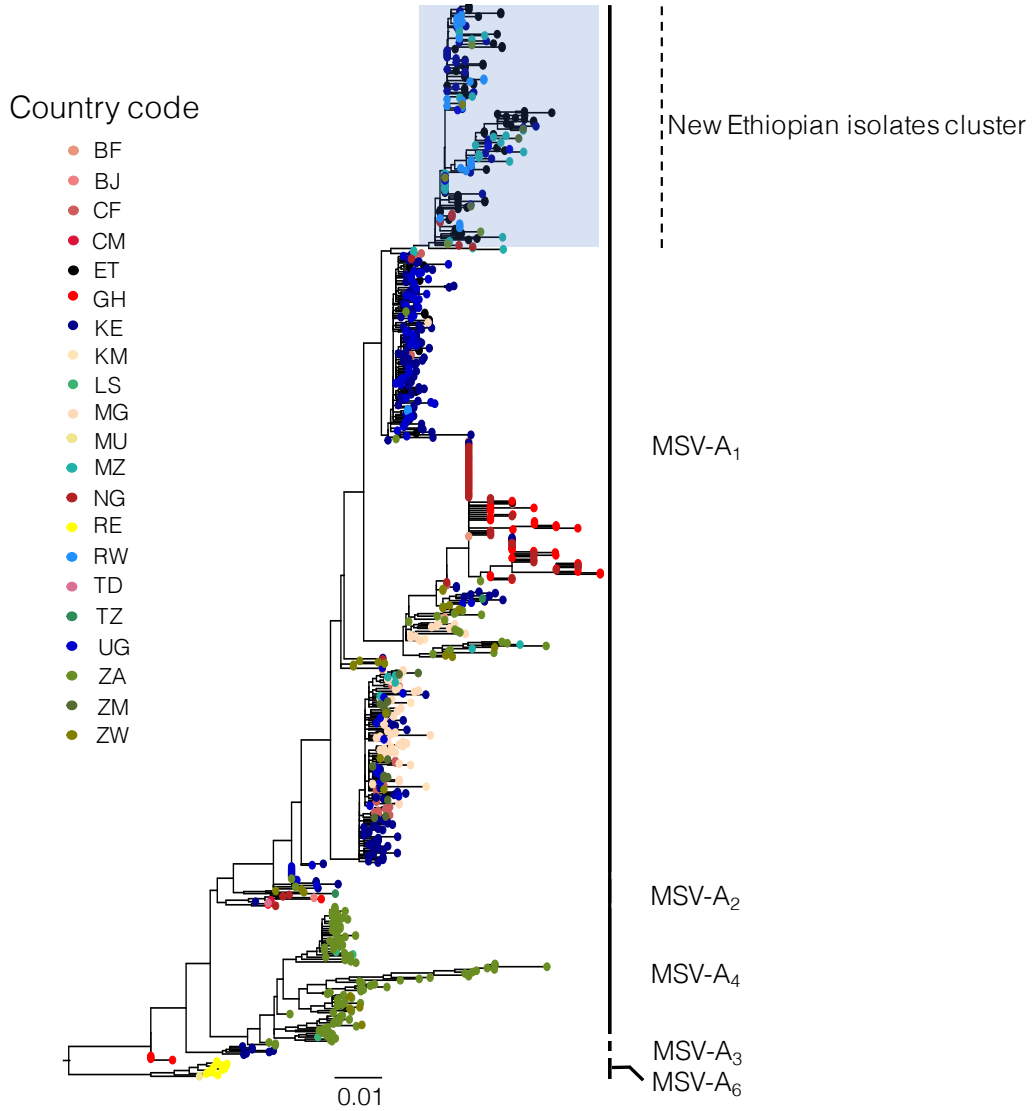


Figure 4.1: MSV-A maximum likelihood phylogeny. The phylogeny (model GTR+I+G_4) was rooted using the Reunion and Mauritius isolates as outgroups and coloured according to countries of sampling origin and labelled with their International Organization for Standardization (ISO) two-letter codes as follows: Burkina Faso [BF], Benin [BJ], Central African Republic [CF], Ethiopia [ET], Ghana [GH], Kenya [KE], Comoros [KM], Lesotho [LS], Madagascar [MG], Mauritius [MU], Mozambique [MZ], Nigeria [NG], Reunion Island [RE], Rwanda [RW], Chad Republic [TD], Tanzania [TZ], Uganda [UG], South Africa [ZA], Zambia [ZM], and Zimbabwe [ZW]. The bar indicates the number of nucleotide substitutions per site.

Monjane *et al*. (2011) has grouped MSV-A variants into 24 recombinant lineages, 16 (MSV-A_1I to MSV-A_1XVI) of which contained predominantly MSV-A_1 sequences. Of all the MSV-A subtypes, it appears that MSV-A_1 has the widest distribution throughout continental Africa and is also found on Madagascar and the Comoros islands (Martin *et al*., 2001; Monjane *et al*., 2011). The Ethiopian isolates in the present study all belong to recombinant lineage V of the MSV-A_1 subtype (Supplementary Table 2). They share about 98% sequence similarity with MSV-A_1 isolates from Kenya (Pande *et al*., 2017) and Rwanda (Asiimwe *et al*., 2020).

4.5.2 Temporal signal analysis

My results showed that the MSV-A whole genome sequences have a significant degree of temporal signal (R^2 = 0.101, p-value < 0.001), with a correlation coefficient of 0.505 for a regression of sampling dates against root to tip divergence estimated (as determined using TempEst; Rambaut *et al.*, 2016). Before any time-calibrated phylogenetic tree inference within the Bayesian genealogical framework, especially for a fast-evolving virus such as MSV, it is advisable to test for suitable temporal signal in the dataset. Put simply, this tests whether the divergence accumulated in a sequences that have been sampled at different times correlates with the times when the sequences were sampled. (see Supplementary Figure 2).

4.5.3 Substitution rate and evolution

The estimated mean substitution rate of MSV-A since the most recent common ancestor of all analysed MSV-A isolates was 5.7419×10^{-4} substitutions per genome site per year (95% highest probability density interval [HPD] 5.1174×10^{-4} to 6.3899×10^{-4}). MSV, like RNA viruses, has a higher substitution rate than most other DNA viruses and thus has the potential to evolve

almost as rapidly as RNA viruses. It appears that such high substitution rates are conserved across the *Geminiviridae* (Siobain Duffy & Holmes, 2009; Lefeuvre *et al*., 2011; Pinto *et al*., 2021). Overall, these estimates fall within the confidence intervals of previous estimates of MSV-A substitution rates determined both from similar model based analyses of MSV-A genome sequences sampled from nature (Harkins *et al*., 2009; Monjane *et al*., 2011; Pande *et al*., 2017)

4.5.4 MSV-A introductions into Ethiopia and Rwanda; from where, when, and how many times?

I analysed the MSV-A dataset for phylogeographic inferences of where, when and how many times the present MSV-A ancestral sequences moved into Ethiopia and Rwanda from the regions where viruses were sampled, using the discrete sampling year, sampling country metadata and strict molecular clock. The time scaled phylogeographic history of MSV-A representing its movements between these regions is summarised as a maximum clade credibility (MCC) tree (Figure 4.2). Out of the ten tree nodes that represented ancestral sequences with immediate descendants that moved into Ethiopia, nine had posterior probability support of 1.00 and one had 0.99 support. The geographical locations of these "origin" ancestors were all inferred to be Kenya; all with location state posterior probability supports (PPS) of between 0.99 and 1(Table 4.1).

The first detectable movement into Ethiopia was estimated to have occurred in 1985 (95% HPD, 1979 to 1988) with a posterior probability support (PPS for Kenya) of 1.00. The most recent of the detectable movement from also Kenya, with a PPS of 1.00, was inferred to have occurred in approximately 2010 (95% HPD, 2009 to 2012). I identified a total of four well

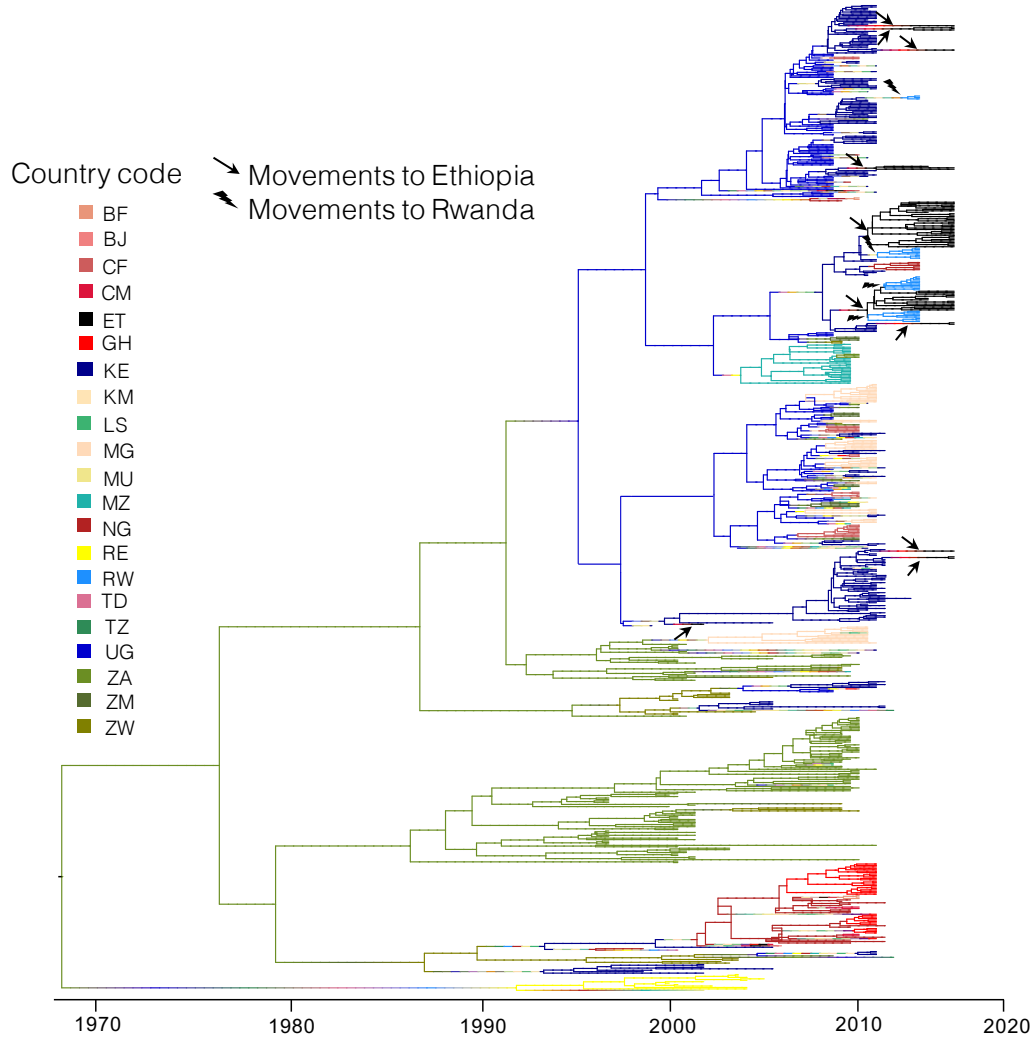


Figure 4.2: MSV-A maximum clade credibility (MCC) phylogeny generated under the discrete phylogeographic diffusion model. Branches are coloured according to either the most probable country where ancestral sequences represented by the branches existed or by the country where isolates were sampled. Colour changes along branches are indicative of ancestral virus movements or transitions between countries. The countries of origin are labelled with their International Organization for Standardization (ISO) two-letter codes as follows: Burkina Faso [BF], Benin [BJ], Central African Republic [CF], Ethiopia [ET], Ghana [GH], Kenya [KE], Comoros [KM], Lesotho [LS], Madagascar [MG], Mauritius [MU], Mozambique [MZ], Nigeria [NG], Reunion Island [RE], Rwanda [RW], Chad Republic [TD], Tanzania [TZ], Uganda [UG], South Africa [ZA], Zambia [ZM], and Zimbabwe [ZW].

Table 4.1: Estimated MSV-A movements into Ethiopia and Rwanda

Movement	Country (Posterior probability)	Year (95% HPD)
Inward movement to Ethiopia		
1	Kenya (1.00)	1985 (1979 to 1988)
2	Kenya (1.00)	2004 (2003 to 2005)
3	Kenya (1.00)	2005 (2002 to 2006)
4	Kenya (1.00)	2005 (2003 to 2006)
5	Kenya (0.99)	2005 (2001 to 2008)
6	Kenya (1.00)	2007 (2005 to 2009)
7	Kenya (1.00)	2008 (2004 to 2009)
8	Kenya (1.00)	2009 (2007 to 2010)
9	Kenya (1.00)	2009 (2007 to 2011)
10	Kenya (1.00)	2010 (2009 to 2012)
Outward Ethiopia/Inward Rwanda		
1	Kenya (1.00)	2007 (2005 to 2009)
2	Kenya (1.00)	2007 (2005 to 2008)
3	Ethiopia (1.00)	2008 (2004 to 2009)
4	Ethiopia (1.00)	2010 (2009 to 2011)

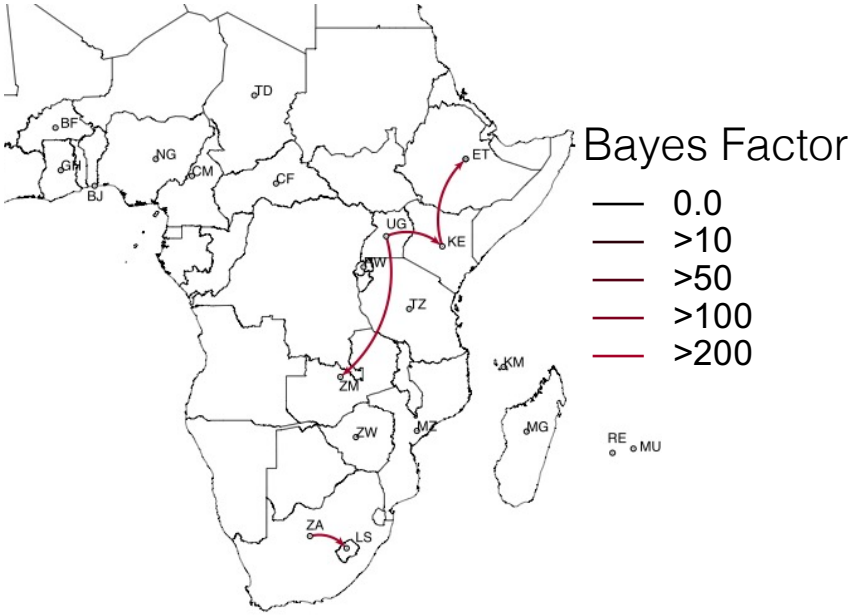


Figure 4.3: Geospatial representations of Bayes factor supports for the best supported movement pathways as inferred using a Bayesian Stochastic Search Variable Selection (BSSVS) procedure under the asymmetric, non-reversible substitution model within analysed locations. Arrows indicate movement directions. Red and blue arrows indicate in and out country movements respectively. The countries of origin are labelled with their International Organization for Standardization (ISO) two-letter codes as follows: Ethiopia [ET], Uganda [UG], Kenya [KE], Tanzania [TZ], Zambia [ZM], Comoros Island [KM], Mozambique [MZ], Madagascar [MG], Reunion Island [RE], Mauritius [MU], Zimbabwe [ZW], South Africa [ZA] and Lesotho [LS].

supported (i.e., with a BF of > 10) MSV-A movement pathways within continental Africa (Figure 4.3). Most importantly here was the BF support of > 100 for MSV-A movements into Ethiopia from Kenya which indicates overwhelming statistical support for either Kenya or Uganda being the origin of the MSV-A lineages with most widespread dissemination (Figure 4.3).

The first detectable MSV-A movement into Rwanda occurred in 2007 (95% HPD, 2005 to 2009) from Kenya with PPS of 1.00. The most recent movement was from Ethiopia in 2010 (95% HPD, 2009 to 2011) also with PPS of 1.00 (Table 4.1).

The present analyses have shown the most plausible time of MSV-A movement into Ethiopia and Rwanda. Although it is unlikely that these movements represent the very first MSV-A movement into these countries, they nevertheless give inferential insights into both plausible, evidence based, movement timelines, and the timeframes over which pathogen diversity turn-over occurs within a particular African country.

For instance, the first inferred movement of MSV-A between Kenya and Ethiopia in ~1985 , preceded a serious MSD outbreak across maize growing regions of Ethiopia in 1986 (Briddon *et al*, 1994; Mesfin *et al*., 1991; Pinner *et al*., 1988): a report that was also the first to explicitly indicate that MSV was endemic in Ethiopia since before the 1980s and has been constraining maize yields in that country for decades before it was first reported there in the 1980s. Although It is difficult to say with a high degree of accuracy when and from where maize was introduced into Ethiopia, it is likely that maize first came into Ethiopia via East Africa in the late sixteenth or early seventeenth century and that Ethiopian varieties likely originated in the Americas and the Caribbean (Africa & Maize, 1968; Berger, 1962; Bjarnason, 1986; Huffnagel, 1961).

The establishment of African Free Trade Zones (AFTZ) as well as liberal trade policies that have promoted more human movement between member states of the East African Community (EAC) (Adar, 2011; Wangia *et al*., 2002) may have had key role in MSV-A spread within the East African countries of Ethiopia and Rwanda. For instance, between 2005 and 2009, Ethiopia's vegetable imports from both Kenya and Uganda (WITS, 2009a, 2009b) had almost doubled to 77.59% from 46.05%. Since MSV cannot be transmitted within seeds, human-mediated means of transmission might include transportation of viruses within infected corn sheaths, floral parts, cuttings (of for example sugarcane which MSV-A also infects) (Antwerpen *et al*., 2011) and seedlings (of maize or other grass species that MSV-A is known to infect; CABI, 2019; Martin & Shepherd, 2009; Varsani *et al*., 2008).

Cross border smugglings between Kenya and Ethiopia are common and could also significantly impact virus disseminations and overall economic stability of the region (Murdoch, 2016; Turi, 2020). Rapid increased in movement of goods, humans, and services by road and air in this period also significantly correlate with the inferred MSV-A movements for these countries (Mundi, 2019c, 2019b). Despite tighter trade and border policies with concerted efforts on technological leverage for border protections in recent years in Rwanda, (IOM, 2015), significant informal trades in green maize between Rwanda, Democratic Republic of Congo (DRC), Uganda and other countries bordering it are also common and indicative of challenges inherent in free trade agreements with almost little to non-existent work plan for proper informal trade routes monitoring (Alusala, 2010; Batirbaev *et al*., 2013). Smugglers are twice as likely to have neither the understanding of how a virus such as MSV-A spreads, nor the proper bio-sanitation rules that can limit unwanted viruliferous insects, or MSV-A infected sugarcane cuttings, given they are mainly subsistent, green (maize that is consumed shortly after harvest) traders.

The first report of MSV in Rwanda was in 1988 (Pinner *et al*., 1988), around the time of MSV epidemics in Ethiopia: possibly indicating that the scale and geographical range of the Ethiopian outbreak may have encompassed much of East Africa. This isolate, an MSV-A came from a *Setaria* plant growing in a maize field, further gives clear insight into how wild grasses that serve as MSV-A natural host can impact its transmission, and from time to time maintain an endemic status within a region (Varsani *et al*., 2008).

Further, the virus can be transported via viruliferous leafhoppers that travel fairly long distances within the cabins of automobiles, trains or aeroplanes; and also by wind across borders (Brewbaker, 1978; Thresh, 1983). The latter is fairly easy to achieve for insects that migrate short distances of few meters to kilometres within habitat patches across borders, is even more plausible here. Given that these insects have low flying speed of usually less than 3 m/s, their navigation is expected to be assisted by an external wind force, and depending on the wind speed (Creamer, 2020; Gilbertson *et al*., 2019; Maramorosch, 1980; Drake, 1995), viruliferous insect vector dispersal and coverage rates can significantly increase over a short period of time.

The demographic and movement dynamics of the nine *Cicadulina* leafhopper species that transit MSV (Asanzi, Bosque-Perez, & Nault, 1995; Bosque-Pérez, Buddenhagen, & Bosque-Perez·, 1999; Okoth, & Dabrowski, 1987), mainly found in the sub-Saharan Africa are the main determinants of where MSV-A occurs (Bosque-Pérez e*t al*., 1999; Rose, 1974). These leafhopper insect vectors, and the indigenous grasses upon which they feed serve as intermediate hosts and reservoirs for the virus. The leafhopper vectors themselves can also be described as mobile, virus reservoirs in that once MSV-A is acquired by these insects they can transmit it for the remainder of their lifespans. MSV-A infected grasses which have much

longer lifespan than insects would, however, be the main means by which reservoirs of MSV-A are maintained between maize growing seasons (Lefeuvre *et al*., 2019; Mesfin *et al*., 1991; Page *et al.*, 1999). Variations in species compositions, and the sizes of these plant populations across continental Africa must also strongly influence MSV disease dynamics. Further, insect (and therefore also MSV-A) migrations are also likely strongly influenced by seasonal and geographical availability of the plants that leafhoppers feed on (Okoth, & Dabrowski, 1987).

4.6 Conclusion

Economic, socio-political, ecological and geographical proximity factors are likely all at least partially responsible for MSV-A movements between Ethiopia, Rwanda and other East African countries. Natural leafhopper mediated movements are also more plausible given the higher possibility of viruliferous leafhoppers migrations across borders. Further, unintended movements of either viruliferous insect vectors or infected plant materials in green maize and sugarcane cuttings trades via porous land borders can also play a role in MSV-A disseminations to these countries. More regulatory frameworks should be put in place across land borders, and proper phytosanitary and bio-sanitation measures will further prevent MSV-A spread.

4.7 Authors' contributions and acknowledgements

Main author's contribution

I cloned and sequenced the newly determined MSV-A genome from Ethiopia, ran all the tools for phylogenetic and phylogeographic analyses (IQ-Tree, Tempest, and Beast) and wrote R scripts used for data visualization. Lastly, I wrote ~80% of the manuscript.

Co-authors' contributions

1. Darren Martin inspired this project and supervised it. He wrote about 20% of the manuscript and provided critical text and figure editing.
2. Arvind Varsani inspired this project, supervised it, and helped with sequencing the data.
3. Gordon Harkins provided supervision.

Acknowledgements

I thank Mark Miller for providing access to CIPRES servers at www.phylo.org and useful tips for Beast analysis.

Chapter 5: Ancestral Lineage Specific Symptom Evolution in the A_1- Subtype *Maize streak virus* is Host Adaptive

5.1 Abstract

Efforts to unravel the events closely following the emergence of the A-strain of *Maize streak virus* (MSV-A) as a maize pathogen could yield useful insights into how successful plant pathogens evolve. Specifically, tracking the symptom evolution of a highly successful MSV-A lineage in relation to the genetic evolution of that lineage could reveal useful insights into how the biology of emergent pathogens adapts over decades-long periods to optimize their infection of a new host species. I therefore examined the ancestral sequences of the model MSV-A isolate, MSV-A-ZW-MatA_1994: the first characterized member of the MSV-A_1 subtype. MSV-A_1 is presently the most commonly found of all presently known MSV-A subtypes and also has the broadest geographical distribution: including all of Africa and the island of Madagascar. Here, a direct biological characterization of chemically resynthesised ancestral MSV-A genomes was used to study the symptom phenotypes of several direct ancestors of MSV-A-ZW-MatA_1994; including the most recent common ancestor of all the currently known MSV-A isolates, which is believed to have existed in ~1900 shortly after the emergence of MSV-A as a severe maize pathogen. I showed that whereas a symptom type such as chlorotic leaf area increased in intensity through the ~90 years of evolution that yielded MSV-A-ZW-MatA_1994, other symptom types either varied in a less concerted way or remained largely unchanged. These data indicate that MSV-A has likely evolved over the last century to maximize the numbers of leaf cells that it replicates within while ensuring that it does not inflict excessive damage on the plants that it infects.

5.2 Introduction

Following the introduction of maize to West Africa in the 1500s by Portuguese traders (Mccann, 2001), its long journey to prominence as a major staple food crop in the region, and the intensification of its cultivation between the mid 1600s to 1800s were not without attendant challenges (Charles, 2014; Harkins *et al*., 2009; Shepherd *et al*., 2010). Maize streak disease (MSD) caused by *Maize streak virus* (MSV; Genus: *Mastrevirus*; Family: *Geminiviridae*) is one of over 50 pathogens that infect and cause diseases in maize plant, and persists as one of the major biotic constraints on maize production in Africa (Redinbaugh & Zambrano, 2014).

The emergence of a MSD causing MSV variants in southern Africa in approximately the late 1800s (Harkins *et al*., 2009), and its subsequent spread to other parts of the continent, has today yielded a major threat to food security in the sub-Saharan Africa and its adjacent Indian Ocean islands (Asiimwe *et al*., 2020; Bediako *et al*., 2017; Guadie *et al*., 2019; Monjane *et al*., 2011; Pande *et al*., 2012, 2017). Of the eleven MSV strains (named A through K), the maize adapted MSV-A strain is the one that is predominantly responsible for maize yield losses: especially among impoverished farmers in sub-Saharan Africa who have limited access to improved MSV resistant maize varieties and insecticides (Martin & Rybicki, 2002; Martin, Willment, & Rybicki, 1999; Martin & Shepherd, 2009).

The emergence of MSV-A as a maize pathogen is an example of how the introduction of exotic plant species into a new environment can precipitate the host-jumping and emergence in the exotic hosts of indigenous viruses that normally infect only indigenous plants in that environment (Monjane *et al*., 2020; Varsani *et al*., 2008). Recombination likely played a key role in the host-jumping and emergence of MSV-A, similar to what has been observed in RNA

viruses (Fargette *et al*., 2006; Shackelton *et al*., 2005; Varsani *et al*., 2008; Walt *et al*., 2009). Specifically, MSV-A is a recombinant of two other MSV strains: -B and a currently unsampled strain most closely related to strains -F or -G (Varsani *et al*., 2008). The potential for recombination to facilitate the rapid evolution of host adaptation in MSV has been demonstrated *in vitro* (Monjane *et al*., 2014).

Leafhopper insects in the Family *Cicadellidae* including *Cicadulina mbila* and *Cicadulina storey* among others, can transmit MSV (Asanzi, Bosque-Perez, & Nault, 1995; Bigirwa *et al*., 1995; Dabrowski, 1987; Mccann, 2001; Okoth, & Dabrowski, 1987). While the host range of MSV includes over 100 diverse grass species that *Cicadulina* sp. feed on, the geographical range of the virus also reflects that of its insect transmission vector. Leafhoppers are piercing and sucking feeders that use both mechanical force, and enzymatic digestion to penetrate plant cell walls to access nutrient rich sieve tubes, parenchyma and collenchyma leaf cells (Bosque-Pérez *et al*., 1999). MSV virions present within cellular contents ingested by a leafhopper pass from the gut to the salivary glands via the haemolymph (Bosque-Pérez, Buddenhagen, & Bosque-Perez, 1999; Mesfin & Bosque-Perez·, 1998; Stores, 1925). Leafhopper transmission of MSV is defined as persistent, circulative, and non-propagative such that a leafhopper that has ingested MSV particles can potentially transmit the virus for the rest of its life (Lefeuvre *et al*., 2019; Marchand *et al*., 1995; Martin & Shepherd, 2009). Hence, these insect vectors play a key role in MSV-A dissemination and are central to its evolution.

Studies of symptoms produced by MSV in infected plants are useful and can reveal many details about the evolution, of MSV disease severity and virulence (Martin *et al*., 2001). The evolution over decades-long time periods of symptoms produced by viruses such as MSV can illuminate why some viral lineages persists and others go extinct. Based on whether they have

lethal or sub-lethal effects on their hosts, symptoms produced by MSV-A can be classified as more or less harmful (Martin & Rybicki, 2002; Martin *et al*., 2001; Monjane *et al*., 2020). Symptoms that are more reflective of harm to plant hosts include: (i) chlorotic intensity (ii) leaf deformation and (iii) leaf stunting. Symptoms that are more reflective of benefits to the virus include chlorotic areas produced on symptomatic leaves since these are both a direct reflection of how well MSV colonizes host cells, and an indicator of the potential of the virus to be successfully transmitted (Monjane *et al*., 2020).

Symptom severity studies in a virus such as MSV-A can therefore be quite useful for studies of how viruses evolve to balance benefits for themselves and harm that they inflict on their hosts. Specifically, much of the adaptation of a virus is constrained by a requirement that it does not harm its host so excessively that it reduces its probability of onward transmission.

The emergence of MSV-A and the previous attempts to understand the likely host adaptive events following its initial emergence as a maize pathogen could provide useful information on how pathogens with broad host ranges such as MSV-A evolve. Symptoms produced in maize by a single unbroken lineage of MSV-A variants since the emergence of MSV-A as a maize pathogen in the early 1900s until the sampling of the MSV-A model, MSV-A-ZW-MatA_1994, in 1994 would be particularly useful for determining how virulence has evolved over time. Currently, the best supported theories suggest that host adaptation of a pathogen such as MSV-A could involve infections becoming less harmful to maize if this maximizes their transmission probability and, ultimately, the global prevalence of the virus (Cressler *et al*., 2016; Read, 1994).

Here I use a direct biological characterization of chemically resynthesised direct ancestors of MSV-A-ZW-MatA_1994 to study how the symptom phenotypes of MSV-A_1 subtype evolved during the 1900s.

5.3 Materials and methods

5.3.1 Recombination analyses and inference of ancestral MSV sequences

A MSV-A full genome sequence dataset containing 59 isolates for which symptoms were assessed and quantified together with an additional 630 MSV-A sequences, and 182 sequences of MSV strains B through K was constructed with MUSCLE (Edgar, 2004) using default settings and edited by eye with IMPALE (Khoosal, 2015). The resulting alignment containing 871 MSV full genome sequences was screened for recombination using seven different recombination detection approaches implemented in RDP4.46 (Martin *et al*., 2015).

The first screening step involved using the default RDP4.46 settings to identify and remove all evidence of inter-strain recombination from all the MSV strains other than MSV-A. All recombinant sequence fragments were removed and the associated nucleotide characters in the alignment replaced with the standard 'gap' character, '-'. In the second step, a p-value cut-off of 10^{-6} for without multiple testing p-value correction with all other RDP4.46 settings left as default. The conservative nature of the Bonferroni uncorrected p-value cut off is because the ancestral sequence reconstructions were specifically focused on a small subset of all detectable recombinant events that would significantly impact the inference of particular ancestral MSV-A sequences that would later be chemically synthesised to determine the symptoms associated with these ancestral MSV-A sequences.

Ancestral sequence inference from recombination-free MSV-A datasets was accomplished using RDP4.46 and MrBayes 3.2 (Ronquist *et al*., 2012) as previously described by Monjane *et al*. (2020). Here, my interest lies in investigating the progression of symptoms from the most recent common ancestor (MRCA) of all the currently sampled MSV-A isolates, hereafter referred to as 'A0', and a further three ancestral sequences leading to MSV-A-ZW-MatA_1994.

5.4 Direct inference of ancestral virus symptom phenotypes

5.4.1 Rhizobium inoculation and symptom quantification

In previous studies carried out in the Molecular cell biology department at the University of Cape Town, MSV field isolates representing the known diversity of MSV-A isolates (n = 59) were selected based on the broadest possible sampling times (ranging from 1979 to 2007) and were used to make, and test the symptoms associated with, Rhizobium-infectious clones. The symptoms associated with these 59 isolates are broadly representative of the major MSV-A clades that were found in eastern, western and southern Africa between 1979 and 2007. Descriptions of the construction of these infectious clones can be found in Martin *et al*. (2001), Boulton *et al*. (1989) and Monjane *et al.* (2020).

Symptoms produced by these isolates were quantified by Monjane *et al.* (2020) using three different maize genotypes with varying degrees of MSV sensitivity: (i) Golden Bantam which is sensitive to MSV (and will hereafter be referred to as the S genotype), (ii) STAR174 which is moderately sensitive to MSV (hereafter referred to as the M genotype), and (iii) PAN77 which is highly resistant to MSV (hereafter referred to as the R genotype). The MSV resistance properties of these three maize genotypes also represent those of commercial cultivars in use

throughout Africa dating back to the 1980s (Martin *et al*., 1999). Between 36 and 72 three-day old seedlings were inoculated in up to five separate experiments with *Rhizobium radiobacter* (formerly *Agrobacterium tumefaciens*) infectious constructs of the MSV-A isolates as in Martin *et al*. (2001).

Each experiment had a control of between 12 and 16 plants inoculated with only *R. radiobacter* containing plain expression vector (pBI121, Inqaba Biotech South Africa). These controls served as standards for calculating leaf stunting in infected plants as well as for detecting cross-contamination. Inoculated plants were grown for 35 days with 16 hr of light per day at between 20 and 25 degrees centigrade in a biosafety level one plant growth room. Leaves four, five and six were harvested at post inoculation days 21, 28 and 35 respectively by carefully collecting cuttings of the second quarter from the base of fully emerged maize leaves. Harvested leaf segments were subjected to automated symptom quantification by image analysis. Leaf segments were scanned at 300dpi using a HP Scanjet 559OP flatbed scanner (Hewlett Packard USA), and Windows 32 bit bitmap images were analysed with the custom-written image analysis program, DIA (Martin, 2019).

5.4.2 DIA symptom quantification for identified leaf segments

Leaf images were captured against a blue background with a circular size standard. The image analysis program, DIA, firstly identified and counted the pixel numbers in the size standard and thereafter identified the leaf segments as other ‘non-blue’ objects in the image that were above 500 pixels in size. The leaf segment area relative to the standard size was calculated by dividing the pixel number within the leaf segment by the pixel number within the size standard, and the unit of leaf area measurement adjusted to millimetre square (mm^2) using the known

size-standard area. The leaf segment areas determined from each *Rhizobium*-inoculation experiment for a given maize host genotype were individually divided by the average leaf areas determined for the control plants of the same genotype from the same experiment. This number was then singly subtracted from one in order to yield the final stunting score (such that higher scores would mean more stunting, while lower scores would mean reduced or lower stunting) for each leaf segment.

Degrees of leaf deformation were measured as the deviation of the leaf segment from being a perfect rectangle by fitting the smallest possible rectangle around the leaf segment and then calculating the proportion of pixels inside this rectangle that fell outside the leaf segment (such that higher values would indicate higher degrees of deformation).

The leaf area covered by chlorotic streaks or lesions was determined by initially finding the ten most distinct pixel colour groups in individual leaf images and then identifying which of the pixel colour groups denote chlorotic lesions. The program determined the key pixel colour groups by sampling 1000 pixels at random from the leaf segment image, computing the Euclidean distance between the red, green and blue colour score for every pixel pair and then constructing a UPGMA dendrogram from these scores followed by the identification of the pixels mapping to each of the ten most basal clades of this dendrogram. Each pixel in the leaf image was assigned to one of the ten groups subject to the Euclidean distance between the pixel's red, green, and blue scores and the mean of the red, green, and blue scores of the ten-pixel groups.

Following all pixel assignments to groups, the average red, green, and blue colour scores for each of the ten groups were estimated. To define the pixel groups for chlorotic lesions, the ten-

pixel colour groups were partitioned into the two sets that maximized the mean red + blue colour scores of the sets. The pixel sets with the highest mean red + blue scores was adjudged as the 'chlorotic lesion set' and those pixel sets with the lower scores as the 'non-chlorotic leaf area set'. The proportion of pixels allocated to the chlorotic lesion set divided by the total number of pixels in the leaf segment image represent the proportion of the leaf area covered by chlorotic lesions. The average intensity of chlorosis, which is a measure of the degree of chloroplast destruction within infected leaf cells, was calculated by estimating the average of the red, green and blue colour scores of every pixel in the chlorotic lesion set and dividing by 255 such that pure white = 1 and pure black = 0).

5.5 Results and discussion

5.5.1 Changes over time in chlorotic areas and intensity of chlorosis from A0 through to MSV-A-ZW-MatA_1994

I examined chlorotic area and intensity of chlorosis symptoms in the lineage of viruses leading from A0, the most recent common ancestor of all sampled MSV-A variants, through MSV-A_1, the most recent common ancestor of all MSV-A variants sampled on mainland Africa, through MSV-A_2, the most recent common ancestor of all subtype MSV-A_1 variants, through MSV-A6, the most recent common ancestor of the sampled variants in the most prevalent MSV-A_1 sub-lineage to MSV-A-ZW-MatA_1994, the first discovered MSV-A_1 variant.

Chlorotic area and intensity of chlorosis symptoms are interesting because whereas chlorotic areas are a measure of how effectively MSV variants colonize the chlorenchyma cells of maize leaves from which leafhoppers acquire the virus for onward transmission (and are therefore a reasonably direct measures of variant fitness), chlorotic intensities indicate the degree to which

viruses impede photosynthesis via damage to chloroplasts within infected chlorenchyma cells (and therefore are a reasonably direct measure of the harm that variants incur on infected plants).

It was observed that chlorotic areas tended to increase in all maize genotypes along the lineage of MSV variants beginning with A0 in ~1900 and culminating MSV-A-ZW-MatA_1994 in 1994 (Figure 5.1). The chlorotic area symptoms induced by the variants in this lineage were similar between the S- and M-maize (note the overlapping 95% confidence intervals of the means for the purple and orange plots in Figure 5.1) but were, as expected, consistently higher in these genotypes for individual variants than those observed in the R-maize (purple plots in Figure 5.1).

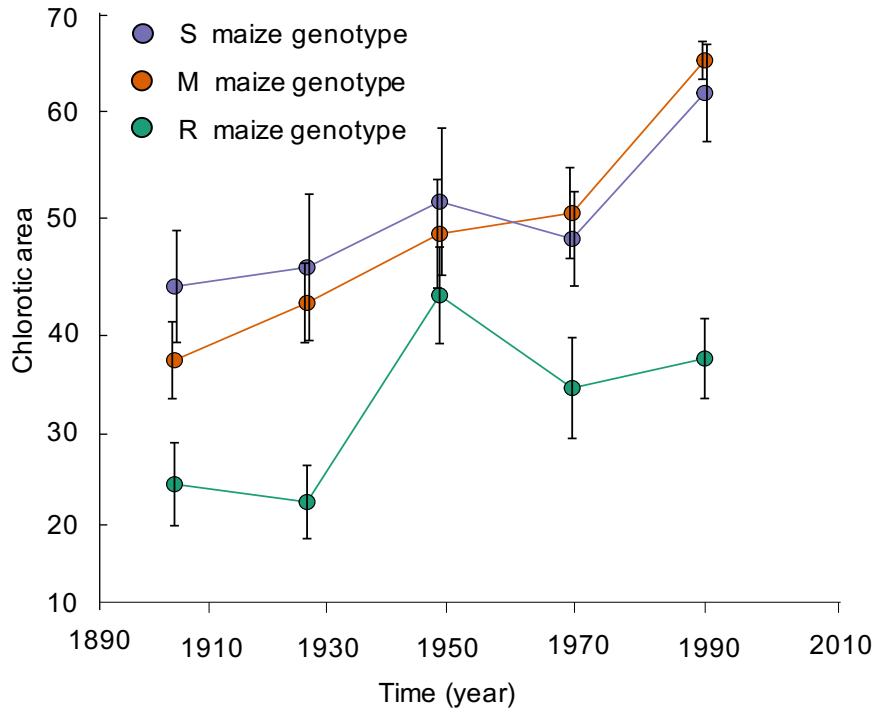


Figure 5.1: Changes in the observed chlorotic areas induced over time by viruses in the lineage of MSV-A variants beginning with the most recent common ancestor of all sampled MSV-A variants, A0 in ~1900 through to the MSV-A-ZW-MatA_1994 (the first MSV-A_1 variant ever discovered) in 1994. Included here are mean symptom measurements for individual synthesised direct ancestors of MSV-A-ZW-MatA_1994 that likely existed in ~1975 (A6), ~1955 (A2) and ~1930 (A1). Symptom measurements were made in sensitive (S), moderately resistant (M) and resistant (R) maize genotypes. Error bars represent 95% confidence intervals of means.

For intensity of chlorosis symptoms on the other hand (Figure 5.2) the patterns of variation between lineages was not as straightforward. In all maize genotypes a sharp decrease in chlorotic intensity was noted between the A0 and A1 ancestors suggesting that the degree of chloroplast destruction within infected cells that was induced by the earlier MSV-A variants may have been maladaptive. Although selection may have favoured a decrease in chlorotic intensity between ~1900 (when the A0 ancestor existed) and ~1930 (when the A1 ancestor existed), thereafter chlorotic intensity increased to the point where, in MSV-A-ZW-MatA_1994 it had again reached similar levels in the R- and M-maize genotypes to those observed for A0. Although in the S-maize genotype MSV-A-ZW-MatA_1994 induces lower degrees of chlorotic intensity than the A0, it induces substantially more chlorotic intensity than A1.

This pattern of symptom evolution through time is consistent with an evolutionary trend that has been primarily driven by increased cell colonization with concomitant selection favouring the minimization of collateral damage to chloroplasts that is incurred by increasing numbers of infected leaf chlorenchyma cells (Monjane *et al*., 2020).

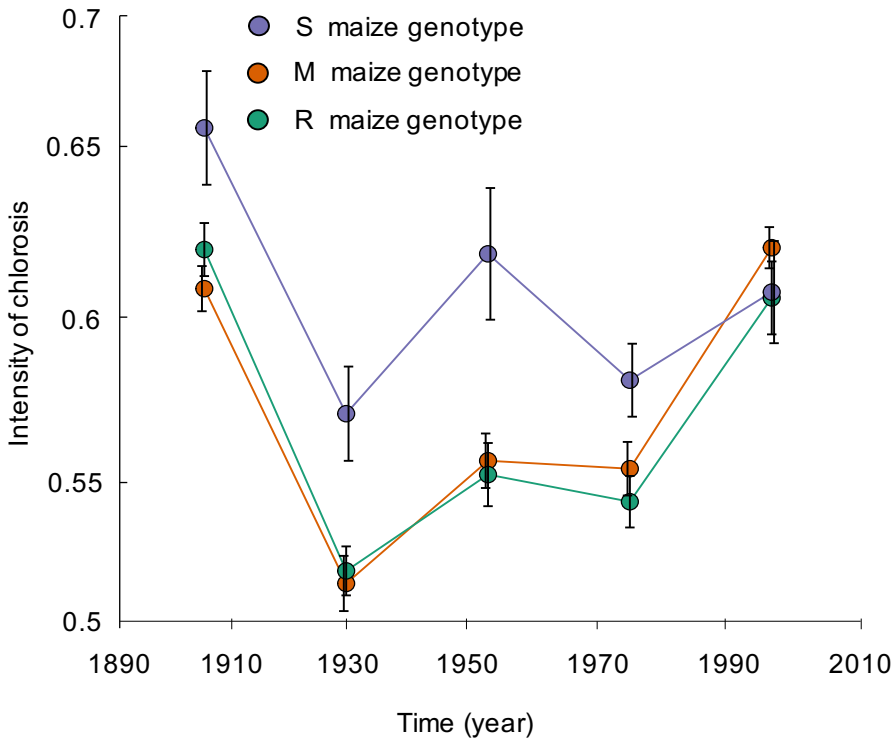


Figure 5.2: Changes in the observed intensities of chlorosis induced over time by viruses in the lineage of MSV-A variants beginning with the most recent common ancestor of all sampled MSV-A variants, A0 in ~1900 through to the MSV-A-ZW-MatA_1994 (the first MSV-A_1 variant ever discovered) in 1994. Included here are mean symptom measurements for individual synthesised direct ancestors of MSV-A-ZW-MatA_1994 that likely existed in ~1975 (A6), ~1955 (A2) and ~1930 (A1). Symptom measurements were made in sensitive (S), moderately resistant (M) and resistant (R) maize genotypes. Error bars represent 95% confidence intervals of means.

Increased chlorotic areas are beneficial for the virus because MSV is only acquired by leafhoppers from chlorotic lesions (Peterschmitt *et al*., 1992) such that larger chlorotic lesion sizes on leaves should correspond with increased probabilities of onward virus transmission. Besides selection favouring the preservation of photosynthesis within infected cells, selection for optimization of chlorotic intensity may have been driven by leafhopper feeding preferences since leafhoppers and other sap-feeding insects are generally attracted to specific shades of the colour yellow (Mauck, De Moraes, & Mescher, 2010). The rising and falling patterns of chlorotic intensity change along the lineage of viruses between A0 and MSV-A-ZW-MatA_1994 might therefore reflect an evolutionary trade-off that balances pressures to reduce chloroplast damage to ensure photosynthesis against pressures to increase chloroplast damage to ensure optimal attractiveness of lesions to leafhoppers.

5.5.2 Changes over time in leaf deformation and leaf stunting from A0 through to MSV-A-ZW-MatA_1994

Besides chlorotic intensity other direct measures of the harm that MSV-infections impose on infected plants are leaf deformation and leaf stunting. Leaf deformation, measured as the deviation of the shape of infected plant leaf segments from a perfect rectangular reflects the degree to which the virus impacts the normal cellular replication of infected cells. Deformation symptoms seem to have changed over time in different ways in different maize genotypes (Figure 5.3). Whereas the greatest degrees of leaf deformation were generally seen in the S-maize genotype, the lowest were generally seen in the M-maize genotype. Whereas in the M- and R-maize genotypes degrees of leaf deformation appear to have remained reasonably constant from the A0 through A1, A2, A6 and MSV-A-ZW-MatA_1994, in the S-maize

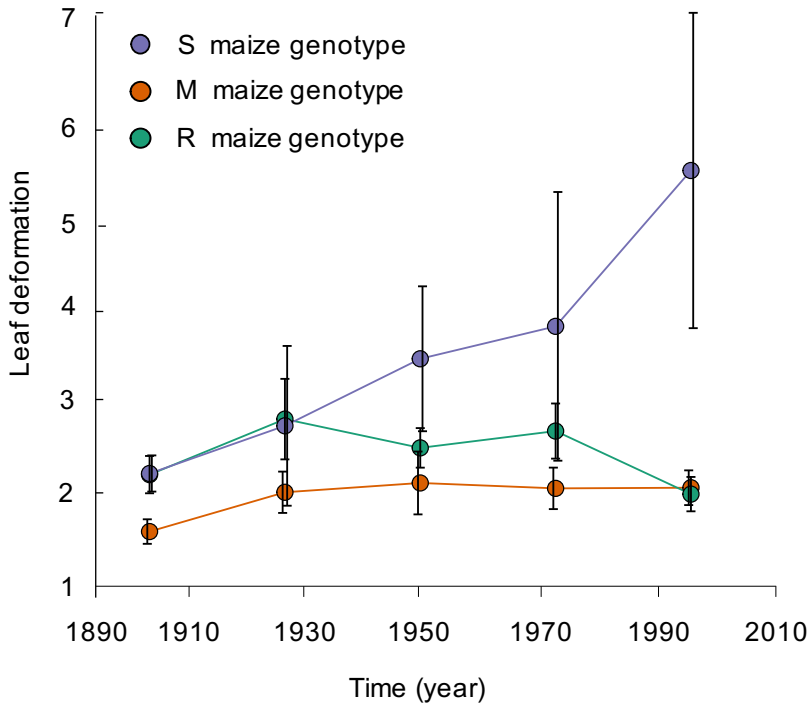


Figure 5.3: Changes in the observed leaf deformation induced over time by viruses in the lineage of MSV-A variants beginning with the most recent common ancestor of all sampled MSV-A variants, A0 in ~1900 through to the MSV-A-ZW-MatA_1994 (the first MSV-A_1 variant ever discovered) in 1994. Included here are mean symptom measurements for individual synthesised direct ancestors of MSV-A-ZW-MatA_1994 that likely existed in ~1975 (A6), ~1955 (A2) and ~1930 (A1). Symptom measurements were made in sensitive (S), moderately resistant (M) and resistant (R) maize genotypes. Error bars represent 95% confidence intervals of means.

genotype they appear to have consistently increased over the ~94 years separating A0 from MSV-A-ZW-MatA.

Whereas leaf stunting relative to uninfected maize plants decreased slightly between A0 and MSV-A-ZW-MatA_1994 in the S- and M-maize genotype the opposite trend was apparent (Figure 5.4). In the M-maize genotype there appears to have been a large increase in stunting between ~1900 (corresponding with the A0 ancestor) and ~1955 (corresponding with the A2 ancestor) whereafter stunting seems to have remained relatively constant.

The consistency of stunting between A2 and MSV-A-ZW-MatA_1994 for the M-maize genotype is somewhat surprising in that it occurred against a backdrop of both increased chlorotic areas and increased chlorotic intensities (Figures 5.1 & 5.2): a combination of symptom changes that one might predict would result in diminished photosynthesis and, therefore, increased stunting. Similarly, in the S-maize and R-maize the overall-decreases in stunting occur against a backdrop of increased chlorotic areas and varying degrees of chlorotic intensity. These patterns suggest that whatever combinations of chlorotic areas and chlorotic intensities occur in the ancestral viruses, these photosynthesis-impacting symptom types are balanced to minimize the degrees of stunting that they induce.

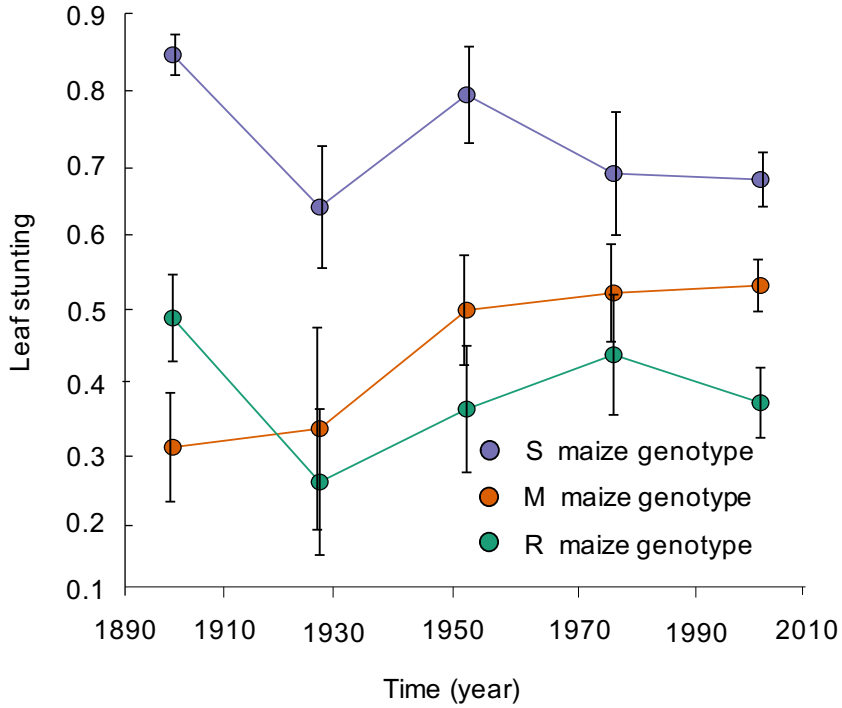


Figure 5.4: Changes in the observed leaf stunting induced over time by viruses in the lineage of MSV-A variants beginning with the most recent common ancestor of all sampled MSV-A variants, A0 in ~1900 through to the MSV-A-ZW-MatA_1994 (the first MSV-A_1 variant ever discovered) in 1994. Included here are mean symptom measurements for individual synthesised direct ancestors of MSV-A-ZW-MatA_1994 that likely existed in ~1975 (A6), ~1955 (A2) and ~1930 (A1). Symptom measurements were made in sensitive (S), moderately resistant (M) and resistant (R) maize genotypes. Error bars represent 95% confidence intervals of means.

5.6 Conclusion

The MSV-A_1 subtype to which MSV-A-ZW-MatA_1994 belongs is today the most prevalent, widely distributed and economically important group of MSV variants in Africa. I have shown that the evolution of this lineage since ~1900 has involved changes over time in the intensities of certain symptom types in particular maize genotypes. The symptom type that has most consistently changed has been chlorotic leaf areas which have displayed a strong tendency to increase through time across all maize genotypes: an observation that is consistent with previous computer modelling of MSV symptom evolution (Monjane *et al*., 2020). Also consistent with previous computer modelling was the observation that between A0 and MSV-A-ZW-MatA_1994 degrees of chlorotic intensity within chlorotic lesions decreased.

However, unlike for the modelling study, I have shown here that (1) chlorotic intensity symptoms have not consistently fallen through time and (2) stunting and deformation symptoms have also risen and fallen over time (the modelling study indicated that these had remained approximately constant). Therefore while increased chlorotic areas seems to be an adaptive change that is associated with increased MSV fitness, it is unclear what the other symptom types collectively reveal about MSV fitness. Clearly in some instances the observed extremes of these symptom types are not unduly detrimental to virus fitness.

Overall, however, the observed trend that chlorotic intensity, leaf deformation and leaf stunting have either slightly decreased or stayed constant between A0 and MSV-A-ZW-MatA_1994 indicates that the harms to infected maize plants that these symptoms types reflect are more likely to have been detrimental to viral fitness than beneficial. Put in another way, with a few exceptions such as leaf deformation in the S-Maize genotype, natural selection seems to have

disfavoured the survival of MSV variants that cause higher degrees of chlorotic intensity, leaf deformation and leaf stunting than those caused by A0 while at the same time favouring the survival of MSV variants that induced larger chlorotic areas than A0. This balance between harm to the host and benefits for the virus that are revealed by the changes in MSV symptoms over time are broadly consistent with the generalized transmission-virulence trade off hypothesis by Alizon *et al*. (2009) and Cressler *et al*. (2016) for explaining the evolution of pathogen virulence

5.7 Authors' contributions and acknowledgements

Main author's contribution

I constructed relevant Rhizobium-infectious clone for the MRCA A0, phenotyped it, analysed and further edited the figures. Lastly, I wrote ~80% of the manuscript.

Co-authors' contributions

1. Darren Martin inspired this project and supervised it. He wrote about 20% of the manuscript and provided critical text and figure editing.
2. Arvind Varsani inspired this project, supervised it, and helped with sequencing the data.
3. Symptom data for the A1, A2 and A6 ancestral genotypes was obtained from Aderito Monjane.

Acknowledgements

I am thankful to everyone in Joann Passmore and Lara Donaldson laboratories and to the laboratory staff in the Molecular and Cell Biology department of UCT.

Chapter 6: Concluding Remarks

6.1 Summary of findings

Food insecurity is a persistent global concern that is greatly exacerbated by a seemingly unending stream of emerging crop pathogens (Gómez *et al*., 2009; Rybicki, 2015). Most at risk are vulnerable, disadvantaged, and impoverished populations such as farmers in the sub-Saharan Africa. Cereals such as maize are among the most essential staples of impoverished people and their cultivation throughout the world remains seriously constrained by both pathogens and climate change (Redinbaugh & Zambrano, 2014; Tigchelaar *et al*., 2018).

Among the most important factors driving food insecurity in sub-Saharan Africa are maize yield losses to *Maize streak virus* (Charles, 2014; Shepherd *et al*., 2010). It is, therefore, useful to assume that efforts directed at increasing yields incorporate a multidisciplinary approach that leverages a combination of molecular biology, plant pathology, epidemiology and computational biology approaches. Such integrative strategies require constant ”molecular surveillance” of viral genomes enabling both early detection of emerging virus threats, and the tracking of viral spread. When performed well, such surveillance can also illuminate the processes underlying viral emergence and the adaption of emerging viruses to their hosts.

Following the recent recovery of what were initially believed to be grass-adapted mastrevirus strains from maize plants displaying severe maize streak disease symptoms (Kraberger *et al*., 2017; Pande *et al*., 2012), it was considered important to assess under controlled laboratory conditions the virulence in maize of these newly discovered mastrevirus variants: MSRV and MSV-C. Such an analysis formed the first part of my project (Chapter 2) whereby I used

infectious clones of *Maize streak Reunion virus* (MSRV) and MSV-C to determine that neither of these viruses are likely capable of causing severe maize streak disease on their own. It is, however, interesting that MSV-B and MSRV in mixed infection appeared to produce more intense chlorotic symptoms in maize than when either were alone. It is, therefore, possible that synergism has a role in the evolution and emergence of geminiviruses.

Even though it was not likely that the particular MSV-C and MSRV strains that I analysed are on their own serious threats to maize production it remains possible that other mastreviruses might emerge as a threat to maize either alone or as part of a disease complex. Following such a hypothetical emergence event the virus will likely spread via the same routes taken by other African mastreviruses that are transmitted by leafhoppers.

The second part of my project therefore focused on revealing some of these pathways using analysis of MSV-A sequences sampled from across the known geographical range of MSV. In the first of these MSV movement studies, described in Chapter 3, I reconstructed the movements of MSV-A in and out of Madagascar using newly determined genome sequences of viruses sampled from the island. This revealed that the viruses causing MSD on Madagascar had likely only been introduced to the island from East Africa in the last 40 years. Crucially, this analysis revealed that there have been frequent introductions of MSV-A variants to the island since the 1990s which indicates that sufficient steps are not being taken to avoid the transfer of leafhoppers and/or MSV-A infected plant materials to the island from East Africa.

In the second MSV-A movement study, described in Chapter 4, I analysed newly determined MSV genomes from Ethiopia and Rwanda to reconstruct the movements of MSV-A in and out of these East African countries. This analysis revealed that neither of these two countries seems

to be as epidemiologically important with respect to trans-African MSV-A movements as are Kenya and Uganda. Specifically, MSV-A variants found in both countries seem to have originated in Uganda or Kenya with no evidence being found of onward transmissions of viruses from Rwanda or Ethiopia to any countries outside of East Africa. In this regard, Rwanda and Ethiopia are likely to be sinks rather than sources of epidemiologically important MSV variants.

In my last research chapter, I took a retrospective look at the evolution of MSV-A over the century following its emergence as a severe maize pathogen. Specifically, I tracked symptom evolution over the past 90 years of the lineage of viruses that yielded the most economically important MSV-A subtype: MSV-A_1, the subtype that dominates MSV-A populations in all parts of Africa besides southern Africa. Here I found that while successively evolved MSV-A_1 variants consistently induced increasingly more extensive chlorotic streaking on the leaves of the maize plants that they infected, these variants tended to induce degrees of stunting, leaf deformation and chloroplast destruction within infected photosynthesizing cells that tended to either stay constant or decrease. These observations support the hypothesis that during the course of their adaptation to infecting maize, the lineage of MSV variants that yielded the successful MSV-A_1 lineage minimized the amount of collateral harm it did to the plants it infected while maximizing its transmission probability through increasing the number of photosynthesizing host leaf cells that it infects

6.2 Challenges encountered during my PhD project

The main challenges experienced while completing the work presented in Chapter 2 are wet laboratory-related. Specifically, it was very difficult to construct *Rhizobium* infectious clones

for MSV-C and MSV-B as their dimerization could not be readily accomplished by the common, relatively convenient methods applicable to other well-studied strains. The procedure used to dimerize these, although more challenging, provided me with an opportunity to hone my cloning skills.

Computational resources and time challenges posed significant challenges for Chapters 3 and 4. These are generic challenges of the bioinformatics field, the extent of which depends on the nature of the analyses. My analyses would generously benefit from higher computing power, which for me, most Ph.D. students and scientists in most developing countries, is generally either difficult to access or non-existent. For my kind of analysis, the greater the number of taxa, the greater the required computing power.

6.3 Prospects

Despite the current advances in unravelling the molecular basis for viral diversity, emergence, virulence evolution, and host adaptive processes, we still lack a comprehensive understanding of the host adaptive process in MSV infected wild grass. Therefore, future multidisciplinary research including the fields of virology, bioinformatics, molecular biology and environmental ecology to understand the true niche that MSV fills in the biome would benefit tremendously from the development of a practical protocol enabling the *Rhizobium* mediated inoculation of wild grasses with cloned MSV genomes. Such a protocol would provide us access to experimental investigations into the host range evolution of MSV and related mastreviruses: the key component of virus evolution that ultimately underlies all virus disease emergence events.

It is estimated that less than 1% of virus species are presently known (Webster *et al*., 2015; Yozwiak *et al*., 2012). With the increasing rate of viral discovery we can now say with an increasing degree of confidence that the time is close where we will be able to determine in a completely unbiased manner the underlying causes of, and constraints on, the evolution of virulence in viruses that infect multiple different hosts. Illuminating this phenomenon is key to understanding past instances of viral host-switching and predicting future emergence events.

References

Abouzid, A. M., Frischmuth, T., & Jeske, H. (1988). A putative replicative form of the abutilon mosaic virus (gemini group) in a chromatin-like structure. *MGG Molecular & General Genetics*, *212*(2), 252–258. https://doi.org/10.1007/BF00334693

Abraham, A. (2019). Emerged Plant Virus Disease in Ethiopian Agriculture : Causes and Control Options. *Ethiopian Journal of Agricultural Science*, *29*(1), 39–55.

Acosta-Leal, R., Duffy, S., Xiong, Z., Hammond, R. W., & Elena, S. F. (2011). Advances in Plant Virus Evolution: Translating Evolutionary Insights into Better Disease Management. *Phytopathology*, *101*(10), 1136–1148. https://doi.org/10.1094/Phyto-01-11-0017

Adar, K. G. (2011). East African Community. In: First International Democracy Report: The democratization of international organizations. International democracy Watch (IDW), Centre for Studies on African Federalism, pp. 1 - 39.

Africa, S., & Maize, R. (1968). Towards self sufficiency. *Nature*, *218*(5143), 713. https://doi.org/10.1038/218713a0

Alegbejo, M. D., & Olalekan O. Banwo. (2005). Journal of plant protection research. *Journal of Plant Protection Research*, *45*(2), 99–105.

Alizon, S., Hurford, A., Mideo, N., & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of Evolution ary Biology*, *22*(2), 245–259. https://doi.org/10.1111/j.1420-9101.2008.01658.x

Alusala, N. (2010). African Security Review Informal cross-border trade and arms smuggling along the Uganda-Rwanda border Informal cross-border trade and arms smuggling along the Uganda-Rwanda border. *African Security Review*, 19:3, 15-26, DOI:10.1080/10246029.2010.519875

Ammar, E. D., Tsai, C. W., Whitfield, A. E., Redinbaugh, M. G., & Hogenhout, S. A. (2009).

Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts. *Annual Review of Entomology*, *54*, 447–468. https://doi.org/10.1146/annurev.ento.54.110807.090454

Anderson, R. M., & May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, *85*(2), 411–426. https://doi.org/10.1017/S0031182000055360

Antwerpen, Van T., Mcfarlane, S. A., Potier, B. A. M., Way, M. J., Shepherd, D. N., Martin, D, P., & Webster, T. M. (2011). Report on Maize Streak Virus in the South African Sugar Industry. *Proc S Afr Sug Technol Ass,* (Vol. 84). DOI:10.1.1.410.2465

Argüello-Astorga, G., & Ruiz-Medrano, R. (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol*, *146*.

Asanzi, C. M., Bosque-Perez, N. A., & Nault, L. R. (1995). Movement of Cicadulina storeyi (Homoptera: Cicadellidae) in maize fields and its behaviour in relation to maize growth stage. *Insect Science and Its Application*, *16*(1), 39–44. https://doi.org/10.1017/s1742758400018300

Asiimwe, T., Stewart, L. R., Willie, K., Massawe, D. P., Kamatenesi, J., & Redinbaugh, M. G. (2020). Maize lethal necrosis viruses and other maize viruses in Rwanda. *Plant Pathology*, *69*(3), 585–597. https://doi.org/10.1111/ppa.13134

Autrey, L. (1983). Maize mosaic virus and other maize virus diseases in the islands of the western Indian Ocean. In D. T. Gordon, J. K. Knoke, L. R. Nault, and R. M. Ritter, Eds. Proceedings International Maize Virus Disease Colloquium and Workshop, 167–181. Retrieved from http://ag.udel.edu/delpha/7114.pdf

Barka, H. Ben. (2012). Border posts, checkpoints, and Intra-African trade: challenges and solutions. *OPEV Newsletter*, (January), 1–18.

Bediako, E., Asare, A., Puije, G. C., van der, K. J., Frimpong, K. A., Amenorpe, G., Kubi, A.

A., Lamptey, J. N., Oppong, A., Mochiah, B., Adama, I., & Tetteh, F. N. (2017). Spatio-Temporal Variations in the Incidence and Severity of Maize Streak Disease in the Volta Region of Ghana. *Journal of Plant Pathology & Microbiology*, *08*(03). https://doi.org/10.4172/2157-7471.1000401.

Berger, J. (1962). Maize production and the manuring of maize. Centre d'Etude de l'Azote.

Betts, H. C., Puttick, M. N., Clark, J. W., Williams, T. A., Donoghue, P. C. J., & Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology and Evolution*, *2*(10), 1556–1562. https://doi.org/10.1038/s41559-018-0644-x

Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Molecular Biology and Evolution*, *33*(8), 2167–2169. https://doi.org/10.1093/molbev/msw082

Bigirwa, G., Gibson, R., Page, W., J. H., Ransom, S., & Pixley, J. (1995). A new maize disorder in Uganda caused by *Cicadulina niger*. Maize Research for Stress Environments. Proceedings of the Eastern and Southern Africa. *Sidalc.Net*. Retrieved from http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=CIMMYT.xis&method=post&formato=2&cantidad=1&expresion=mfn=021765

Bjarnason, M. (1986). Progress in breeding for resistance to the maize streak virus disease. A Proceedings of the First Eastern, Central and Southern Africa Regional Maize Workshop. Lusaka, Zambia. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf%5Cnhttp://www.cabdirect.org/abstracts/19876762709.html

Bock, K. R. (1974). Maize streak virus. C.M.I./A.A.B. *Descriptions of Plant Viruses*, *No. 133,4*.

Borrow, P. (1997). Mechanisms of viral clearance and persistence. *Journal of Viral Hepatitis*,

4(s2), 16–24. https://doi.org/10.1111/j.1365-2893.1997.tb00176.x

Bosque-Pérez, N.A., Olojede, S. O., & Buddenhagen, I. W. (1998). Effect of maize streak virus disease on the growth and yield of maize as influenced by varietal resistance levels and plant stage at time of challenge. *Euphytica 1998 101:3*, *101*(3), 307–317. https://doi.org/10.1023/A:1018345921770

Bosque-Pérez, N. A. (2000). Eight decades of maize streak virus research. *Virus Research*, *71*(1–2), 107–121. https://doi.org/10.1016/S0168-1702(00)00192-1

Bosque-Pérez, Nilsa A, Buddenhagen, I. W., & Bosque-Perez·, N. A. (1999). Biology of *Cicadulina* leafhoppers and epidemiology of maize streak virus disease in West Africa. *South African Journal of Plant and Soil*, *16*(1), 50–55. https://doi.org/10.1080/02571862.1999.10634845

Boulton, M. I. (2002). Functions and interactions of mastrevirus gene products. *Physiological and Molecular Plant Pathology*, *60*(5), 243–255. https://doi.org/10.1006/pmpp.2002.0403

Boulton, M. I., Buchholz, W. G., Marks, M. S., Markham, P. G., & Davies, J. W. (1989). Specificity of Agrobacterium-mediated delivery of maize streak virus DNA to members of the Gramineae. *Plant Molecular Biology*, *12*(1), 31–40. https://doi.org/10.1007/BF00017445

Boulton, M. I., Steinkellner, H., Donson, J., Markham, P. G., King, D. I., & Davies, J. W. (1989). Mutational Analysis of the Virion-sense Genes of Maize Streak Virus. *Journal of general Virology,* (Vol. 70). Retrieved from www.microbiologyresearch.org

Brewbaker, J. L. (1978). Diseases of Maize in the Lowland Tropics and the Collapse of the Classic Maya Civilization. In *Ciat (Colombia) 000031*.

Briddon, R. W., Lunness, P., Chamberlin, L. C. L., & Markham, P. G. (1994). Analysis of the genetic variability of maize streak virus. *Virus Genes*, *9*(1), 93–100. https://doi.org/10.

1007/BF01703439

Broadbent, L. Martini, C. (1959). The Spread of Plant Viruses. *Advances in Virus Research*.

Buck, K. . (1999). Geminiviruses (*Geminiviridae*). *Academic Press*.

C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, L. A. B. (2005). The single stranded DNA viruses. Virus Taxonomy*:* Eighth Report of the International Committee on Taxonomy of Viruses. https://doi.org/10.1016/b978-0-12-249951-7.50011-0

CABI; EPPO. (1997). *Maize streak monogeminivirus. [Distribution map]*. Commonwealth Mycological Institute. Retrieved from https://www.cabi.org/isc/abstract/20066500739

CABI. (2019). *Maize streak virus*. Retrieved from https://www.cabi.org/isc/datasheet/32620#REF-DDB-183911

Cassola, F. (2003). *Coleoptera: Cicindelidae, tiger beetles (Studies of tiger beetles CXI).* (ed. S. M. G. & J. P. Benstead, Ed.). In The natural history of Madagascar (ed. S. M. Goodman & J. P. Benstead), pp. 669–677. University of Chicago Press. In The natural history of Madagascar, pp. 669–677. University of Chicago Press.

Castellano, M. M., Sanz-Burgos, A. P., & Gutiérrez, C. (1999). Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology*, 290(3), 639–652. https://doi.org/10.1006/JMBI.1999.2916

Charles, K. (2014). Maize streak virus: A review of pathogen occurrence, biology and management options for smallholder farmers. *African Journal of Agricultural Research*, *9*(36), 2736–2742. https://doi.org/10.5897/AJAR2014.8897

Chen, Q., He, J., Phoolcharoen, W., & Mason, H. S. (2011). Geminiviral vectors based on bean yellow dwarf virus for production of vaccine antigens and monoclonal antibodies in plants. *Human Vaccines*, *7*(3), 331–338. https://doi.org/10.4161/hv.7.3.14262

Chen, S., Huang, Q., Wu, L., & Qian, Y. (2015). Identification and characterization of a maize-

associated mastrevirus in China by deep sequencing small RNA populations. *Virology Journal*, 1–9. https://doi.org/10.1186/s12985-015-0384-3

Claverie, S., Bernardo, P., Kraberger, S., Hartnady, P., Lefeuvre, P., Lett, J.-M., Roumagnac, P. (2018). From Spatial Metagenomics to Molecular Characterization of Plant Viruses: A Geminivirus Case Study. *Advances in Virus Research*, 101, 55–83. https://doi.org/10.1016/BS.AIVIR.2018.02.003

Claverie, S., Ouattara, A., Hoareau, M., Filloux, D., Varsani, A., Roumagnac, P., Lefeuvre, P. (2019). Exploring the diversity of Poaceae-infecting mastreviruses on Reunion Island using a viral metagenomics-based approach. *Scientific Reports*, *9(1)*, 12716. https://doi.org/10.1038/s41598-019-49134-9

Coffin, J. M. (2013). Virions at the Gates: Receptors and the Host–Virus Arms Race. *PLoS Biology*, *11*(5), e1001574. https://doi.org/10.1371/journal.pbio.1001574

Cohen, F. S. (2016). How Viruses Invade Cells. *Biophysical Journal*, *110*(5), 1028–1032. https://doi.org/10.1016/j.bpj.2016.02.006

Collin, S., Mari´, M., Ferna´ndez, M., Ferna´ndez-Lobato, F., Gooding, P. S., Mullineaux, P. M., & Fenoll, C. (1996). The Two Nonstructural Proteins from Wheat Dwarf Virus Involved in Viral Gene Expression and Replication Are Retinoblastoma-Binding Proteins. *Virology,* (Vol. 219).

Comoros. (2014). Mining Laws and regulations Handbook (Volume 1).

Congress, W. G. for the L. of. (1994). Comoros: A Country Study.

Creamer, R. (2020). Beet curly top virus transmission, epidemiology, and management. *Applied Plant Virology*, INC. https://doi.org/10.1016/b978-0-12-818654-1.00037-2

Cressler, C. E., McLeod, D. V., Rozins, C., Van Den Hoogen, J., & Day, T. (2016). The adaptive evolution of virulence: A review of theoretical predictions and empirical tests. Parasitology. Cambridge University Press. https://doi.org/10.1017/S003118201 500092X

Cuevas, J. M., Moya, A., & Elena, S. F. (2003). Evolution of RNA virus in spatially structured heterogeneous environments. *Journal of Evolutionary Biology*, *16*(3), 456–466. https://doi.org/10.1046/J.1420-9101.2003.00547.X

Dabrowski, Z. T. (1987). Two new species of Cicadulina China (Hemiptera: Euscelidae) from West Africa. *Bulletin of Entomological Research*, *77(1)*, 53–56. https://doi.org/10.1017/S0007485300011524

Dabrowski, Z. T., & Cwikla, P. S. (1991). Some New Observations on Cicadulina Leafhoppers Taxonomic Structures. *International Journal of Tropical Insect Science*, *12(1-2–3)*, 237–247. https://doi.org/10.1017/S1742758400020750

Dáder, B., Then, C., Berthelot, E., Ducousso, M., Ng, J. C. K., & Drucker, M. (2017). Insect transmission of plant viruses: Multilayered interactions optimize viral propagation. *Insect Science*, *24(6)*, 929–946. https://doi.org/10.1111/1744-7917.12470

Damsteegt, V. D. (1983). Maize Streak Virus: I. Host Range and Vulnerability of Maize Germ Plasm. *Plant Disease*, *67*(7), 734. https://doi.org/10.1094/pd-67-734

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9(8*), 772. https://doi.org/10.1038/nmeth.2109

Davies, J. W., Boulton, M. I., & Liu, H. (1997). Maize streak virus coat protein binds single- and double-stranded DNA in vitro. *Journal of General Virology*, *78*(6), 1265–1270. https://doi.org/10.1099/0022-1317-78-6-1265

Dawit, A., Chilot, Y., Adam, B., & Agajie, T. (2014). Situation and Outlook of Maize in Ethiopia of Maize in Ethiopia. Ethiopian Institure of Agricultural Research.

Dellicour, S., Rose, R., Faria, N. R., Lemey, P., & Pybus, O. G. (2016). Seraphim: Studying environmental rasters and phylogenetically informed movements. *Bioinformatics*, *32*(20), 3204–3206. https://doi.org/10.1093/bioinformatics/btw384

Deribe, K., Meribo, K., Gebre, T., Hailu, A., Ali, A., Aseffa, A., & Davey, G. (2012). The burden of neglected tropical diseases in Ethiopia, and opportunities for integrated control and elimination. *Parasites & Vectors*, *5*(1), 240. https://doi.org/10.1186/1756-3305-5-240

Donnelly, T. W., & Parr, M. J. (2003). *Odonata, dragonflies and damselflies.* (ed. S. M.). In The natural history of Madagascar, pp. 645–654. University of Chicago.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 214. https://doi.org/10.1186/1471-2148-7-214

Duffy, S, & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *Journal of Virology*, *82*.

Duffy, S., & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. *Journal of General Virology*, *90*(6), 1539–1547. https://doi.org/10.1099/vir.0.009266-0

Durzyńska, J., & Goździcka-Józefiak, A. (2015). Viruses and cells intertwined since the dawn of evolution Emerging viruses. *Virology Journal*, *12*(1), 1–10. https://doi.org/10.1186/s12985-015-0400-7

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Engering, A., Hogerwerf, L., & Slingenbergh, J. (2013). Pathogen-host-environment interplay and disease emergence. *Emerging Microbes & Infections*, *2*(2), e5. https://doi.org/10.1038/emi.2013.5

Fajinmi, A. A., Dokunmu, A. O., Akheituamen, D. O., & Onanuga, K. A. (2012). Incidence and infection rate of *Maize streak virus* disease by *Cicadulina triangular* on maize plants and its distribution from the lowest diseased leaf under tropical conditions. *Archives Of*

Phytopathology And Plant Protection, *45*(13), 1591–1598. https://doi.org/10.1080/03235408.2012.694251

Fargette, D., Konaté, G., Fauquet, C., Muller, E., Peterschmitt, M., & Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annual Review of Phytopathology*, *44*, 235–260. https://doi.org/10.1146/annurev.phyto.44.120705.104644

Fauquet, C. M., Bisaro, D. M., Briddon, R. W., Brown, J. K., Harrison, B. D., Rybicki, E. P., Stanley (Study Group Chair), J. (2003). Virology division news : Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of Virology*, *148*(2), 405–421. https://doi.org/10.1007/s00705-002-0957-5

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Archives of Virology*, *153*(4), 783–821. https://doi.org/10.1007/s00705-008-0037-6

Fauquet, C. M., & Stanley, J. (2005). Revising the way we conceive and name viruses below the species level: A review of geminivirus taxonomy calls for new standardized isolate descriptors. *Archives of Virology*, *150*(10), 2151–2179. https://doi.org/10.1007/s00705-005-0583-0

Fermin, G. (2018). Host Range, Host–Virus Interactions, and Virus Transmission. *Viruses*, 101–134. https://doi.org/10.1016/B978-0-12-811257-1.00005-X

Firth, C., Charleston, M. A., Duffy, S., Shapiro, B., & Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *Journal of Virology*, *83*(24), 12813–12821. https://doi.org/10.1128/JVI.01719-09

Fondong, V. N. (2013). Geminivirus protein structure and function. *Molecular Plant Pathology*, *14*(6), 635–649. https://doi.org/10.1111/mpp.12032

Fondong, V., Pita, J., Rey, M., A, de K., & Beachy, R. (2000). Evidence of synergism between

African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology*, *81*.

Fontes, E. (1994). Geminivirus Replication Origins Have a Modular Organization. *The Plant Cell Online*, *6*(3), 405–416. https://doi.org/10.1105/tpc.6.3.405

Forterre, D. R. and P. (2008). Redefining viruses: lessons from Mimivirus. Retrieved from www.nature.com/reviews/micro

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research*, *117*(1), 5–16. https://doi.org/10.1016/j.virusres.2006.01.010

Gallitelli, D. (2000). The ecology of Cucumber mosaic virus and sustainable agriculture. *Virus Research*, *71*(1–2), 9–21. https://doi.org/10.1016/S0168-1702(00)00184-2

Garcia-Arenal, F., Fraile, A., & Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int Microbiol*, *6*, 225–232. https://doi.org/10.1007/s10123-003-0142-z

Gilbertson, R. L., Melgarejo, T. A., Rojas, M. R., William M Wintermantel, & Stanley, J. (2019). Beet Curly Top Virus (*Geminiviridae*). *Reference Module in Life Sciences*, *(July)*, 301–307. https://doi.org/10.1016/b978-0-12-809633-8.21238-7

Goh, C. S., Bogan, A. A., Joachimiak, M., Walther, D., & Cohen, F. E. (2000). Co-evolution of proteins with their interaction partners 1 1Edited by B. Honig. *Journal of Molecular Biology*, *299*(2), 283–293. https://doi.org/10.1006/jmbi.2000.3732

Gómez, P., Rodríguez-Hernández, A. M., Moury, B., & Aranda, M. A. (2009). Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European Journal of Plant Pathology, 125*(1), 1–22. https://doi.org/10.1007/S10658-009-9468-5

Guadie, D., Tesfaye, K., Knierim, D., Winter, S., & Abraham, A. (2019). Molecular analysis of maize (Zea mays L.)-infecting mastreviruses in Ethiopia reveals marked diversity of virus genomes and a novel species. *Virus Genes*, 1–7. https://doi.org/10.1007/s11262-

019-01655-1

Guerrero, J., Regedanz, E., Lu, L., Ruan, J., Bisaro, D. M., & Sunter, G. (2020). Manipulation of the Plant Host by the Geminivirus AC2/C2 Protein, a Central Player in the Infection Cycle. *Frontiers in Plant Science*. Frontiers Media S.A. https://doi.org/10.3389/fpls.2020.00591

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gutierrez, C. (1999). Geminivirus DNA replication. *Cellular and Molecular Life Sciences*, *56*(3–4), 313–329. https://doi.org/10.1007/s000180050433

Gutierrez, C. (2000). DNA replication and cell cycle in plants: learning from geminiviruses. *The EMBO Journal*, *19*(5), 792–799. https://doi.org/10.1093/emboj/19.5.792

Gutierrez, Crisanto, Ramirez-Parra, E., Mar Castellano, M., Sanz-Burgos, A. P., Luque, A., & Missich, R. (2004). Geminivirus DNA replication and cell cycle interactions. *Veterinary Microbiology*, *98*(2), 111–119. https://doi.org/10.1016/J.VETMIC.2003.10.012

Gutiérrez, S., Michalakis, Y., Munster, M. Van, & Blanc, S. (2013). Plant feeding by insect vectors can affect life cycle, population genetics and evolution of plant viruses. *Functional Ecology*, *27*(3), 610–622. https://doi.org/10.1111/1365-2435.12070

Hanley-Bowdoin, L., Settlage, S. B., & Robertson, D. (2004). Reprogramming plant gene expression: A prerequisite to geminivirus DNA replication. *Molecular Plant Pathology*, *5*(2), 149–156. https://doi.org/10.1111/j.1364-3703.2004.00214.x

Harkins, G., Delport, W., Duffy, S., Wood, N., & Monjane, A. (2009). Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J*, *6*.

Harkins, G. W., Martin, D. P., Duffy, S., Monjane, A. L., Shepherd, D. N., Windram, O. P.,

Varsani, A. (2009). Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *Journal of General Virology*, *90*(12), 3066–3074. https://doi.org/10.1099/vir.0.015537-0

Hefferon, K. L., Moon, Y. S., & Fan, Y. (2006). Multi-tasking of nonstructural gene products is required for bean yellow dwarf geminivirus transcriptional regulation. *FEBS Journal*, *273*(19), 4482–4494. https://doi.org/10.1111/j.1742-4658.2006.05454.x

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., & Gronenborn, B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Research*, *23*(6), 910–916. https://doi.org/10.1093/nar/23.6.910

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hodge, S., & Powell, G. (2008). Do Plant Viruses Facilitate Their Aphid Vectors by Inducing Symptoms that Alter Behavior and Performance? *Environmental Entomology*, *37*(6), 1573–1581. https://doi.org/10.1603/0046-225X-37.6.1573

Hofer, J. M., Dekker, E. L., Reynolds, H. V, Woolston, C. J., Cox, B. S., & Mullineaux, P. M. (1992). Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *The Plant Cell*, *4*(2), 213–223. https://doi.org/10.1105/tpc.4.2.213

Hohn, T., & Sharma, P. (2014). Plant Viru*s* – Host Interaction. https://doi.org/10.1016/B978-0-12-411584-2.01001-5

Holland, J. J., Kennedy, S. I. T., Semler, B. L., Jones, C. L., Roux, L., & Grabau, E. A. (1980). Defective Interfering RNA Viruses and the Host-Cell Response. In Comprehensive Virology (pp. 137–192). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4613-3129-2_3

Horváth, G. V., Pettkó-Szandtner, A., Nikovics, K., Bilgin, M., Boulton, M., Davies, J. W.,

Dudits, D. (1998). Prediction of functional regions of the maize streak virus replication-associated proteins by protein-protein interaction analysis. *Plant Molecular Biology*, *38*(5), 699–712. https://doi.org/10.1023/A:1006076316887

Huffnagel, H. P. (1961). Agriculture in Ethiopia. *Agriculture in Ethiopia.*

Hull, R., & Hull, R. (2014a). Chapter 1 – Introduction. In *Plant Virology* (pp. 3–14). https://doi.org/10.1016/B978-0-12-384871-0.00001-7

Hull, R., & Hull, R. (2014b). Chapter 7 – Replication of Plant Viruses. In *Plant Virology* (pp. 341–421). https://doi.org/10.1016/B978-0-12-384871-0.00007-8

Hull, R., & Hull, R. (2014c). Origins and Evolution of Plant Viruses. *Plant Virology*, 423–476. https://doi.org/10.1016/B978-0-12-384871-0.00008-X

ICTV. (2019). Genus: Mastrevirus. Retrieved from https://ictv.global/taxonomy/

IOM. (2015). Ict Strategy for Integrated Border Management in Rwanda. Government of Rwanda

Ivanowski, D. (1892). Über die Mosaikkrankheit der Tabakspflanze. *St. Petersb. Acad. Imp. Sci. Bull.*, *35,* 67–70.

Iyer, L. M., Balaji, S., Koonin, E. V., & Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Research*, *117*(1), 156–184. https://doi.org/10.1016/j.virusres.2006.01.009

Javaux, E. J., & Lepot, K. (2018). The Paleoproterozoic fossil record: Implications for the evolution of the biosphere during Earth's middle-age. *Earth-Science Reviews*, *176*, 68–86. https://doi.org/10.1016/J.EARSCIREV.2017.10.001

Jeske, H., Lütgemeier, M., & Preiss, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J*, *20*.

Johansson, J. (2008). Evolutionary responses to environmental changes: how does competition affect adaptation? *Evolution*, *62*.

Jones, R. A. C. (2009). Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Research*, *141*(2), 113–130. https://doi.org/10.1016/J.VIRUSRES.2008.07.028

Jordan, I. K., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. *Genome Research*, *12*(6), 962–968. https://doi.org/10.1101/gr.87702

Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*, *389*(6646), 33–39. https://doi.org/10.1038/37918

Khoosal A., & Martin, D. P. (2015). IMPALE:Improved Alignment Editor, (http://www.cbio.uct.ac.za/~arjun/).

King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (2012). Virus Taxonomy Classification and Nomenclature of Viruses Ninth Report of the International Committee on Taxonomy of Viruses. Retrieved from www.macmillansolutions.com

Knoema. (2020a). Ethiopia - Maize production quantity. https://knoema.com/atlas/Ethiopia/topics/Agriculture/Crops-Production-Quantity-tonnes/Maize-production

Knoema. (2020b). Rwanda - Maize production quantity. https://knoema.com/atlas/Rwanda/topics/Agriculture/Crops-Production-Quantity-tonnes/Maize-production

Kong, L. J., Orozco, B. M., Roe, J. L., Nagar, S., Ou, S., Feiler, H. S., Hanley-Bowdoin, L. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *The EMBO Journal*, *19*(13), 3485–3495. https://doi.org/10.1093/emboj/19.13.3485

Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, *479–480*, 2–25. https://doi.org/10.1016/

j.virol.2015.02.039

Koonin, E. V., Senkevich, T. G., & Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biology Direct*, *1*, 1–27. https://doi.org/10.1186/1745-6150-1-29

Koonin, E. V, & Dolja, V. V. (2006). Evolution of complexity in the viral world: The dawn of a new vision. *Virus Research*, *117*, 1–4. https://doi.org/10.1016/j.virusres.2006.01.018

Kraberger, S., Saumtally, S., Pande, D., Khoodoo, M. H. R., Dhayan, S., Dookun-Saumtally, A., & Varsani, A. (2017). Molecular diversity, geographic distribution and host range of monocot-infecting mastreviruses in Africa and surrounding islands. *Virus Research*, *238*, 171–178. https://doi.org/10.1016/J.VIRUSRES.2017.07.001

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. https://doi.org/10.1016/J.Virol.2009.03.024

Krupovič, M., & Bamford, D. H. (2010). Order to the Viral Universe. *Journal of Virology*, *84*(24), 12476–12479. https://doi.org/10.1128/jvi.01489-10

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology 2010 8:5*, *8*(5), 317–327. https://doi.org/10.1038/nrmicro2315

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., & Barraclough, T. G. (2012). Species Interactions Alter Evolutionary Responses to a Novel Environment. *PLoS Biology*, *10*(5), e1001330. https://doi.org/10.1371/journal.pbio.1001330

Lazarowitz, S. G., Pinder, A. J., Damsteegt, V. D., & Rogers, S. G. (1989). Maize streak virus genes essential for systemic spread and symptom development. *The EMBO Journal*, *8*(4), 1023–1032. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16453874

Lefeuvre, P, Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Lett, J. (2016). Begomovirus ' melting pot ' in the south-west Indian Ocean islands : molecular diversity and evolution through recombination, (2007), 3458–3468. https://doi.org/10.1099/vir.0.83252-0

Lefeuvre, Pierre, Harkins, G. W., Lett, J.-M., Briddon, R. W., Chase, M. W., Moury, B., & Martin, D. P. (2011). Evolutionary Time-Scale of the Begomoviruses: Evidence from Integrated Sequences in the Nicotiana Genome. *PLoS ONE*, *6*(5), e19193. https://doi.org/10.1371/journal.pone.0019193

Lefeuvre, Pierre, Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., & Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology*, 1–13. https://doi.org/10.1038/s41579-019-0232-3

Legg, J. P., & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology*, *56*(4), 585–599. https://doi.org/10.1007/s11103-004-1651-7

Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian Phylogeography Finds Its Roots. *PLoS Computational Biology*, *5*(9), e1000520. https://doi.org/10.1371/journal.pcbi.1000520

Lemey, P., Rambaut, A., Welch, J. J., & Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, *27*(8), 1877–1885. https://doi.org/10.1093/molbev/msq067

Locklear, M. (2017). Viruses Would Rather Jump to New Hosts Than Evolve With Them.

Lodish, H. F. (2000). Molecular cell biology. W.H. Freeman.

Lucy, A P, Boulton, M. I., Davies, J. W., & Maule, A. J. (1996). Tissue specificity of Zea mays

infection by maize streak virus. *Molecular Plant-Microbe Interactions : MPMI (USA)*. https://doi.org/https://doi.org/10.1094/MPMI-9-0022

Lucy, Andrew P., Boulton, M. I., Davies, J. W., & Maule, A. J. (1996). Tissue specificity of Zea mays infection by maize streak virus. *Molecular Plant-Microbe Interactions*. https://doi.org/10.1094/MPMI-9-0022

Magenya, O. E. V., Mueke, J., & Omwega, C. (2008). Significance and transmission of maize streak virus disease in Africa and options for management: A review. *African Journal of Biotechnology*, *7*(25), 4897–4910. https://doi.org/10.4314/ajb.v7i25.59697

Magenya, O. E. V, Mueke, J., & Omwega, C. (2009). *Association of maize streak virus disease and its vectors (Homoptera: Cicadelidae) with soil macronutrients and altitudes in Kenya*. *African Journal of Agricultural Research* (Vol. 4). Retrieved from http://www.academicjournals.org/AJAR

Malik, A., Briddon, R. W., & Mansoor, S. (2011). Infectious clones of Tomato leaf curl Palampur virus with a defective DNA B and their pseudo-recombination with Tomato leaf curl New Delhi virus. *Virology Journal*, *8*(1), 173. https://doi.org/10.1186/1743-422X-8-173

Maramorosch, K. (1955). Multiplication of Plant Viruses in Insect Vectors. *Advances in Virus Research*, *3*(C), 221–249. https://doi.org/10.1016/S0065-3527(08)60637-5

Maramorosch, K. (1980). Spread of Plant Viruses and Spiroplasmas Through Airborne Vectors. *Annals of the New York Academy of Sciences*, *353*(1), 179–185. https://doi.org/10.1111/j.1749-6632.1980.tb18921.x

Marchand, J.-L., Peterschmitt, M., Reynaud, B., & Dintinger, J. (1995). Maize streak virus, maize stripe virus and maize mosaic virus in the tropics (Africa and islands in the Indian Ocean). *Agriculture et Développement*, 55–69.

Maret, F. (2007). Distortions to Agricultural Incentives in Madagascar. World Bank's

Development Research

Martin, D. P. (2019). *DIA (Program)*.

Martin, D. P., Shepherd, D. N., & Rybicki, E. P. (2008). Maize Streak Virus. *Encyclopedia of Virology*, 263–272. https://doi.org/10.1016/B978-012374410-4.00707-X

Martin, D., & Rybicki, E. (2002). Investigation of Maize streak virus Pathogenicity determinants using chimaeric genomes. *Virology*, *300*.

Martin, D. P, Willment, J. A., Billharz, R., Velders, R., Odhiambo, B., Njuguna, J., Rybicki, E. P. (2001). Sequence diversity and virulence in Zea mays of Maize streak virus isolates. *Virology*, *288*(2), 247–255. https://doi.org/10.1006/viro.2001.1075

Martin, D. P, Willment, J. A., & Rybicki, E. P. (1999). Evaluation of Maize Streak Virus Pathogenicity in Differentially Resistant Zea mays Genotypes. *Phytopathology*, *89*(8), 695–700. https://doi.org/10.1094/Phyto.1999.89.8.695

Martin, D. P., Lefeuvre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B., Lett, J. M. (2011). Complex Recombination Patterns Arising during Geminivirus Coinfections Preserve and Demarcate Biologically Important Intra-Genome Interaction Networks. *PLoS Pathogens*, *7*(9), e1002203. https://doi.org/10.1371/journal.ppat.1002203

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), 1–5. https://doi.org/10.1093/ve/vev003

Martin, D. P., & Shepherd, D. N. (2009). The epidemiology, economic impact and control of maize streak disease. *Food Security*, *1*(3), 305–315. https://doi.org/10.1007/s12571-009-0023-1

Martin, D. P., & Rybicki, E. P. (1998). Microcomputer-Based Quantification of Maize Streak Virus Symptoms in *Zea mays*. *Phytopathology*, *88*(5), 422–427. https://doi.org/10.1094/PHYTO.1998.88.5.422

Mauck, K. E., De Moraes, C. M., & Mescher, M. C. (2010). Deceptive chemical signals induced by a plant virus attract insect vectors to inferior hosts. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3600–3605. https://doi.org/10.1073/pnas.0907191107

McCann, J. (2001). *Maize and Grace: History, Corn, and Africa's New Landscapes, 1500-1999*. Retrieved from https://about.jstor.org/terms

McCann, J. (2001). Maize and grace: History, corn, and Africa's new landscapes, 1500-1999. *Comparative Studies in Society and History*, *43*(2), 246–272. https://doi.org/10.1017/S0010417501003486

Meacham, C. E., & Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature*, *501*(7467), 328–337. https://doi.org/10.1038/nature12624

Mergeay, J., & Santamaria, L. (2012). Evolution and Biodiversity: the evolutionary basis of biodiversity and its potential for adaptation to global change. *Evolutionary Applications*, *5*(2), 103–106. https://doi.org/10.1111/j.1752-4571.2011.00232.x

Mesfin, T., & Bosque-Perez·, N. A. (1998). Feeding behaviour of Cicadu/ina storeyi China (Homoptera: Cicadellidae) on maize varieties susceptible or resistant to maize streak virus. Retrieved from https://journals.co.za/docserver/ fulltext/ento/6/2/227.pdf?expires=1597396242&id=id&accname=guest&checksum=3499B607B7F5B6E26D297747A5ED19F4

Mesfin, T., Den Hollander, J., & Markham, P. G. (1991). *Cicadulina* species and maize streak virus in Ethiopia. *Tropical Pest Management*, *37*(3), 240–244. https://doi.org/10.1080/09670879109371592

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). *Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees*. Retrieved from http://www.phylo.org/sub_sections/portal/sc2010_paper.pdf

Moffat, A. S. (1999). Plant Pathology:Geminiviruses Emerge as Serious Crop Threat. *Science*, *286*(5446), 1835–1835. https://doi.org/10.1126/science.286.5446.1835

Monaghan, M. T., Gattolliat, J. L., Sartori, M., Elouard, J. M., James, H., Derleth, P., Vogler, A. P. (2005). Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1574), 1829–1836. https://doi.org/10.1098/rspb.2005.3139

Monjane, A. L., Harkins, G. W., Martin, D. P., Lemey, P., Lefeuvre, P., Shepherd, D. N., Varsani, A. (2011). Reconstructing the History of Maize Streak Virus Strain A Dispersal To Reveal Diversification Hot Spots and Its Origin in Southern Africa. *Journal of Virology*, *85*(18), 9623–9636. https://doi.org/10.1128/JVI.00640-11

Monjane, A. L., Martin, D. P., Lakay, F., Muhire, B. M., Pande, D., Varsani, A., Rybicki, E. P. (2014). Extensive Recombination-Induced Disruption of Genetic Interactions Is Highly Deleterious but Can Be Partially Reversed by Small Numbers of Secondary Recombination Events. *Journal of Virology*, *88*(14). https://doi.org/10.1128/JVI.00709-14

Monjane, A. L., Dellicour, S., Hartnady, P., Oyeniran, K. A., Owor, B. E., Bezuidenhout, M., Linderme, D., Syed, R. A., Donaldson, L., Murray, S., Rybicki, E. P., Kvarnheden, A., Yazdkhasti, E., Lefeuvre, P., Froissart, R., Roumagnac, P., Shepherd, D. N., Harkins, G. W.,Suchard, M. A., Lemey, P., Varsani, A., & Martin, D. P. (2020). Symptom evolution following the emergence of maize streak virus. *ELife*, *9*. https://doi.org/10.7554/eLife.51984

Monjane, A. L., Martin, D. P., Lakay, F., Muhire, B. M., Pande, D., Varsani, A., Rybicki, E. P. (2014). Extensive recombination-induced disruption of genetic interactions is highly deleterious but can be partially reversed by small numbers of secondary recombination events. *Journal of Virology*, *88*(14), 7843–7851. https://doi.org/10.1128/JVI.00709-14

Monsalve, Z. I., Arguello-Astorga, G., Rivera-Bustamante, R. F., Monsalve-Fonnegra, Z. I., giiello-Htorga, G., & Rivera-Bustamante, R. F. (2014). Geminivirus Replication and Gene Expression Engineering of the central carbon metabolism in native strains of Saccharomyces cerevisiae for the use of fermentable sugars View project Gene expression regulation View project Geminivirus Replication and Gene Expression. https://doi.org/10.1201/9781482277890-12

Moreno, A. (2009). Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus Research*, *141*(2), 158–168. https://doi.org/10.1016/J.VIRUSRES.2008.10.020

Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). *Archives of Virology*, *158*(6), 1411–1424. https://doi.org/10.1007/s00705-012-1601-7

Mundi. (2019a). Comoros Economy Profile. Retrieved from https://www.indexmundi.com/comoros/economy_profile.html

Mundi. (2019b). Ethiopia - Air transport. Retrieved from https://www.indexmundi.com/facts/ethiopia/air-transport#IS.AIR.PSGR

Mundi. (2019c). Rwanda - Air transport. Retrieved from https://www.indexmundi.com/facts/rwanda/air-transport#IS.AIR.PSGR

Muñoz-Martín, A., Collin, S., Herreros, E., Mullineaux, P. M., Fernández-Lobato, M., & Fenoll, C. (2003). Regulation of MSV and WDV virion-sense promoters by WDV nonstructural proteins: a role for their retinoblastoma protein-binding motifs. https://doi.org/10.1016/S0042-6822(02)00072-7

Murdoch, C. (2016). Porous borders: prosecuting at source. Retrieved January 21, 2021, from https://www.counselmagazine.co.uk/articles/porous-borders-prosecuting-source

Murray, G. G. R., Wang, F., Harrison, E. M., Paterson, G. K., Mather, A. E., Harris, S. R., Welch, J. J. (2016). The effect of genetic structure on molecular dating and tests for temporal signal. *Methods in Ecology and Evolution*, *7*(1), 80–89. https://doi.org/10. 1111/ 2041-210X.12466

Nguyen, L.T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/ msu300

Nikovics, K., Simidjieva, J., Peres, A., Ayaydin, F., Pasternak, T., Davies, J. W., Horváth, G. V. (2001). *Cell-Cycle, Phase-Specific Activation of Maize streak virus Promoters*. */ 609 MPMI* (Vol. 14). Retrieved from https://apsjournals.apsnet.org/doi/pdf/10.1094/MP MI.2001.14.5.609

Okoth, V. A. O., & Dabrowski, Z. (1987). Population density, species composition and infectivity with maize streak virus (MSV) of Cicadulina spp. leafhoppers in some ecological zones in Nigeria. *Cabdirect.Org*, *Acta Oecol*(8), 191–200. Retrieved from https://www. cabdirect.org/cabdirect/abstract/19880547503

Oldstone, M. B. A. (2009). Anatomy of Viral Persistence. *PLoS Pathogens*, *5*(7), e1000523. https://doi.org/10.1371/journal.ppat.1000523

Oluwafemi, S., Kraberger, S., Shepherd, D. N., & Martin, D. P. (2014). A high degree of African streak virus diversity within Nigerian maize fields includes a new mastrevirus from Axonopus compressus. *Archives of Virology*, (159(10)), 2765–2770. https://doi.org/ 10.1007/s00705-014-2090-7

Owor, B. E., Martin, D. P., Shepherd, D. N., Edema, R., Monjane, A. L., Rybicki, E. P., Varsani, A. (2007). Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *Journal of General Virology*, *88*(11),

3154–3165. https://doi.org/10.1099/vir.0.83144-0

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Varsani, A. (2007). Successful application of FTA® Classic Card technology and use of bacteriophage ϕ29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *Journal of Virological Methods*, *140*(1–2), 100–105. https://doi.org/10.1016/j.jviromet.2006.11.004

Owor, B., Martin, D., Shepherd, D., Edema, R., & Rybicki, E. (2007). Genetic analysis of Maize streak virus (MSV) isolates from Uganda reveals widespread distribution of a recombinant MSV variant in Uganda. *J Gen Virol*, *88*.

Oyeniran, K. A., Hartnady, P., Claverie, S., Lefeuvre, P., Monjane, A. L., Donaldson, L., Martin, D. P. (2021). How virulent are emerging maize - infecting mastreviruses ? *Archives of Virology*, (0123456789). https://doi.org/10.1007/s00705-020-04906-x

Page, W. W., Smith, M. C., Holt, J., & Kyetere, D. (1999). Intercrops, Cicadulina spp., and maize streak virus disease. *Annals of Applied Biology*, *135*(1), 385–393. https://doi.org/10.1111/j.1744-7348.1999.tb00865.x

Palmer, K. E., & Rybicki, E. P. (1998). The molecular biology of mastreviruses. *Advances in Virus Research*, *50*, 183–234. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9521000

Pande, D., Kraberger, S., Lefeuvre, P., Shepherd, J. L. D. N., Varsani, A., & Martin, D. P. (2012). A novel maize-infecting mastrevirus from La Reunion Island. *Archives of Virology*, (157(8)), 1617–1621. https://doi.org/10.1007/s00705-012-1314-y

Pande, D., Madzokere, E., Hartnady, P., Kraberger, S., Hadfield, J., Rosario, K., Harkins, G. W. (2017). The role of Kenya in the trans-African spread of maize streak virus strain A. *Virus Research*, *232*, 69–76. https://doi.org/10.1016/J.virusres.2017.02.005

Parrish, C. R., & Kawaoka, Y. (2005). The Origins of New Pandemi Viruses: The Acquisition

of New Host Ranges by Canine Parvovirus and Influenza A Viruses. *Annual Review of Microbiology*, *59*(1), 553–586. https://doi.org/10.1146/annurev.micro.59.030804.121059

Payne, S. (2017). Chapter 8 - Virus Evolution and Genetics. In S. Payne (Ed.), *Viruses* (pp. 81–86). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-803109-4.00008-8

Pazos, F., & Valencia, A. (2008). Protein co-evolution, co-adaptation and interactions. *The EMBO Journal*, *27*(20), 2648–2655. https://doi.org/10.1038/emboj.2008.189

Peake, J. (1971). The evolution of terrestrial faunas in the western Indian Ocean. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *260*(836), 581–610. https://doi.org/10.1098/rstb.1971.0027

Peel, M. C., Finlayson, B. L., & Mcmahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification Updated world map of the K ¨ oppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss* (Vol. 4). Retrieved from www.hydrol-earth-syst-sci-discuss.net/4/439/2007/

Peele, C., Jordan, C. V., Muangsan, N., Turnage, M., Egelkrout, E., Eagle, P., Robertson, D. (2001). Silencing of a meristematic gene using geminivirus-derived vectors. *The Plant Journal*, *27*(4), 357–366. https://doi.org/10.1046/J.1365-313X.2001.01080.X

Penn, D. J., & Smith, K. R. (2007). Differential fitness costs of reproduction between the sexes. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(2), 553–558. https://doi.org/10.1073/pnas.0609301103

Peterschmitt, M., Quiot, J. B., Reynaud, B., & Baudin, P. (1992). Detection of maize streak virus antigens over time in different parts of maize plants of a sensitive and a so-called tolerant cultivar by ELISA. *Annals of Applied Biology*, *121*(3), 641–653. https://doi.org/10.1111/j.1744-7348.1992.tb03473.x

Pilartz, M., & Jeske, H. (2003). Mapping of abutilon mosaic geminivirus minichromosomes.

Journal of Virology, *77*(20), 10808–10818. https://doi.org/10.1128/jvi.77.20.10808-10818.2003

Pinner, M. S., Markham, P. G., Markham, R. H., & Dekker, L. (1988). Characterization of maize streak virus: description of strains; symptoms. *Plant Pathology*, *37*(1), 74–87. https://doi.org/10.1111/j.1365-3059.1988.tb02198.x

Pinner, M. S., Medin, V., Plaskitt, K. A., & Markham, P. G. (1993). Viral inclusions in monocotyledons infected by maize streak and related geminiviruses. *Plant Pathology*, *42*(1), 75–87. https://doi.org/10.1111/j.1365-3059.1993.tb01472.x

Pinto, V. B., Quadros, A. F. F., Godinho, M. T., Silva, J. C., Alfenas-Zerbini, P., & Zerbini, F. M. (2021). Intra-host evolution of the ssDNA virus tomato severe rugose virus (ToSRV). *Virus Research*, *292*, 198234. https://doi.org/10.1016/j.virusres.2020.198234

Preiss, W., & Jeske, H. (2003). Multitasking in Replication Is Common among Geminiviruses. *J Virol*, *77*.

Pulat Batirbaev, TaeYoung Kim, Rezvan Ma'ani, Ryung Shim, Jen Singer, Matt Snyder, F. Y. (2013). Maize in Rwanda : A Value Chain Analysis, 0–45.

Radcliffe, E. B., & Lagnaoui, A. (2007). *Insect pests in potato*. *Potato Biology and Biotechnology: Advances and Perspectives*. Elsevier B.V. https://doi.org/10.1016/B978-044451018-1/50067-1

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, *2*(1), vew007. https://doi.org/10.1093/ve/vew007

Read, A. F. (1994). The evolution of virulence. *Trends in Microbiology*, *2*(3), 73–76.

https://doi.org/10.1016/0966-842X(94)90537-1

Redinbaugh, M. G., & Stewart, L. R. (2018). Maize Lethal Necrosis: An Emerging, Synergistic Viral Disease. *Annual Review of Virology*, *5*(1), 301–322. https://doi.org/10.1146/annurev-virology-092917-043413

Redinbaugh, M. G., & Zambrano, J. L. (2014). Control of Virus Diseases in Maize. *Advances in Virus Research,* (Vol. 90). https://doi.org/10.1016/B978-0-12-801246-8.00008-1

Reynaud, B., Peterschmitt, & M. (1992). A study of the mode of transmission of maize streak virus by Cicadulina mbila using an enzyme-linked immunosorbent assay. *Annals of Applied Biology*, *121*(1), 85–94. https://doi.org/10.1111/j.1744-7348.1992.tb03989.x

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Huelsenbeck, J. P. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1548), 1899–1905. https://doi.org/10.1098/rstb.2010.0057

Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology 2011 9:2*, *9*(2), 99–108. https://doi.org/10.1038/nrmicro2491

Rose, D. J. W. (1978). Epidemiology of Maize Streak Disease. *Annual Review of Entomology*, *23*(1), 259–282. https://doi.org/10.1146/annurev.en.23.010178.001355

Rose, D. J. W. (1974). The epidemiology of maize streak disease in relation to population densities of Cicadulina spp. *Annals of Applied Biology*, *76*(2), 199–207. https://doi.org/10.1111/j.1744-7348.1974.tb07973.x

Rousseau, E., Tamisier, L., Fabre, F., Simon, V., Szadkowski, M., Bouchez, O., Moury, B. (2018). Impact of genetic drift, selection and accumulation level on virus adaptation to its

host plants. *Molecular Plant Pathology*, *19*(12), 2575–2589. https://doi.org/10.1111/MPP.12730

Ruschhaupt, M., Martin, D. P., Lakay, F., Bezuidenhout, M., Rybicki, E. P., Jeske, H., & Shepherd, D. N. (2013). Replication modes of Maize streak virus mutants lacking RepA or the RepA-pRBR interaction motif. *Virology*, *442*(2), 173–179. https://doi.org/10.1016/j.virol.2013.04.012

Rybicki, E. P. (2015). A Top Ten list for economically important plant viruses. *Archives of Virology*, *160*(1), 17–20. https://doi.org/10.1007/S00705-014-2295-9

Saleem, H., Nahid, N., Shakir, S., Ijaz, S., Murtaza, G., Khan, A. A., Nawaz-ul-Rehman, M. S. (2016). Diversity, Mutation and Recombination Analysis of Cotton Leaf Curl Geminiviruses. *Plos One*, *11*(3), e0151161. https://doi.org/10.1371/journal.pone.0151161

Schalk, H. J., Matzeit, V., Schiller, B., Schell, J., & Gronenborn, B. (1989). Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. *The EMBO Journal*, *8*(2), 359. Retrieved from /pmc/articles/PMC400814/?report=abstract

Schnippenkoetter, W. H., Martin, D. P., Hughes, F. L., Fyvie, M., Willment, J. A., James, D., Rybicki, E. P. (2001). The relative infectivities and genomic characterisation of three distinct mastreviruses from South Africa. *Archives of Virology*, *146*(6), 1075–1088. https://doi.org/10.1007/s007050170107

Scoville, H. (2018). Differential Reproductive Success in the Science of Evolution. Retrieved May 22, 2019, from https://www.thoughtco.com/differential-reproductive-success-1224662

Settlage, S., Miller, A., Gruissem, W., & Hanley-bowdoin, L. (2001). Dual Interaction of a Geminivirus Replication Accessory Factor with a Viral Replication Protein and a Plant Cell Cycle Regulator. *Virology*, *279*.

Shackelton, L. A., Parrish, C. R., Truyen, U., & Holmes, E. C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(2), 379–384. https://doi.org/10.1073/pnas.0406765102

Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. *Cell*, *138*(5), 822–829. https://doi.org/10.1016/j.cell.2009.08.017

Shepherd, D. N., Mangwende, T., Martin, D. P., Bezuidenhout, M., Thomson, J. A., & Rybicki, E. P. (2007). Inhibition of maize streak virus (MSV) replication by transient and transgenic expression of MSV replication-associated protein mutants. *Journal of General Virology*, *88*(1), 325–336. https://doi.org/10.1099/vir.0.82338-0

Shepherd, D. N., Martin, D. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Rybicki, E. P., & Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *Journal of Virological Methods*, *149*(1), 97–102. https://doi.org/10.1016/j.jviromet.2007.12.014

Shepherd, D. N., Martin, D. P., Van Der Walt, T, E., Dent, K., Varsani, A., & Rybicki, E. P. (2010). Maize streak virus: an old and complex â€˜emergingâ€™™ pathogen. *Molecular Plant Pathology*, *11*(1), 1–12. https://doi.org/10.1111/j.1364-3703.2009.00568.x

Shepherd, D. N, Martin, D. P., Van Der Walt, E., Dent, K., Varsani, A., & Rybicki, E. P. (2010). Maize streak virus: an old and complex “emerging” pathogen. *Molecular Plant Pathology*, *11*(1), 1–12. https://doi.org/10.1111/j.1364-3703.2009.00568.x

Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, *16*(8), 1114–1116. https://doi.org/10.1093/oxfordjournals.molbev.a026201

Soto, P. E., Buddenhagen, I. W., & Asnani, V. L. (1982). Development of streak virus-resistant

maize populations through improved challenge and selection methods. *Annals of Applied Biology*, *100*(3), 539–546. https://doi.org/10.1111/j.1744-7348.1982.tb01420.x

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs withinthe conserved nonanucleotide motif during the initiation of rolling circle DNA replication. *Virology*, *206*(1), 707–712. https://doi.org/10.1016/S0042-6822(95)80093-X

Stores, H. H. (1925). The Transmission of Streak Disease of Maize By the Leafhopper Balclutha Mbila Nalide. *Annals of Applied Biology*, *12*(4), 422–439. https://doi.org/10.1111/j.1744-7348.1925.tb04238.x

Storey, H. (1930). The Transmission of Streak Disease Between Maize, Sugar Cane and Wild Grasses. *The Annals of Applied Biology*, *17*(4).

Storey, H. H. (1933). Investigations of the Mechanism of the Transmission of Plant Viruses by Insect Vectors.--I. *Proceedings of the Royal Society B: Biological Sciences*, *113*(784), 463–485. https://doi.org/10.1098/rspb.1933.0060

Storey, H. H. (1936). Virus Diseases of East African Plants. *The East African Agricultural Journal*, *1*(6), 471–475. https://doi.org/10.1080/03670074.1936.11663710

Storey, H. H. (1938). Investigations of the mechanism of the transmission of plant viruses by insect vectors II. The part played by puncture in transmission. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, *125*(849), 455–477. https://doi.org/10.1098/rspb.1939.0038

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/ve/vey016

Tavare, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Some Mathematical Questions in Biology / DNA Sequence Analysis Edited by*

Robert M. Miura. Retrieved from https://agris.fao.org/agris-search/search.do?recordID=US201301755037

Thrall, P. H., & Burdon, J. J. (2003). Evolution of virulence in a plant host-pathogen metapopulation. *Science*, *299*(5613), 1735–1737. https://doi.org/10.1126/science.1080070

Thresh, J. (1983). The long-range dispersal of plant viruses by arthropod vectors. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1111), 497–528. https://doi.org/10.1098/rstb.1983.0071

Thresh, J. M. (2004). Control of plant virus diseases in sub-Saharan Africa: the possibility and feasibility of an integrated approach. *African Crop Science Journal*, *11*(3). https://doi.org/10.4314/acsj.v11i3.27571

Tigchelaar, M., Battisti, D. S., Naylor, R. L., & Ray, D. K. (2018). Future warming increases probability of globally synchronized maize production shocks. *Proceedings of the National Academy of Sciences*, *115*(26), 6644–6649. https://doi.org/10.1073/PNAS.1718031115

Trewick SA. (2000). Molecular evidence for dispersal rather than vicariance as the oprigin of flightless insect species on the Chatham Islands, New Zealand. *Journal of Biogeography*, *27*(5), 1189–1200.

Turi, D. G. G. (2020). Cross-border smuggling between Kenya and Ethiopia continues apace - ENACT Africa. Retrieved January 21, 2021, from https://enactafrica.org/enact-observer/cross-border-smuggling-between-kenya-and-ethiopia-continues-apace

UNdata. (2019). Madagascar - Air transport. Retrieved March 26, 2021, from https://www.indexmundi.com/facts/madagascar/air-transport#IS.AIR.PSGR

USAID. (2019). *Country Overview;Madagascar*.

USDA. (2020). Madagascar Corn Imports by Year. Retrieved January 15, 2021, from

https://www.indexmundi.com/agriculture/?country=mg&commodity=corn&graph=imports

Uzest, M., Gargani, D., Drucker, M., Hébrard, E., Garzo, E., Candresse, T., Blanc, S. (2007). A protein key to plant virus transmission at the tip of the insect vector stylet. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17959–17964. https://doi.org/10.1073/pnas.0706608104

Drake, A. V. A. (1995). *Insect Migration: Tracking Resources through Space and Time*.

Van Der Walt, Eric, Martin, D. P., Varsani, A., Polston, J. E., & Rybicki, E. P. (2008). Experimental observations of rapid maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virology Journal*, *5*(1), 104. https://doi.org/10.1186/1743-422X-5-104

Van Der Walt, Eric, Palmer, K. E., Martin, D. P., & Rybicki, E. P. (2008). Viable chimaeric viruses confirm the biological importance of sequence specific maize streak virus movement protein and coat protein interactions. *Virology Journal*, *5*, 1–11. https://doi.org/10.1186/1743-422X-5-61

Varsani, A., Martin, D. P., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Murilo Z. F., Brown, J. K. (2014). Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Archives of Virology*, *159*(7), 1873–1882. https://doi.org/10.1007/s00 705-014-1982-x

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Murilo Z. F., Martin, D. P. (2014a). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Archives of Virology*, *159*(8), 2193–2203. https://doi.org/10.1007/s00705-014-2050-2

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Murilo Z. F., Martin, D. P. (2014b). Establishment of three new genera in the family

Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Archives of Virology*, *159*(8), 2193–2203. https://doi.org/10.1007/s00705-014-2050-2

Varsani, A., Roumagnac, P., Fuchs, M., Jesú S Navas-Castillo, Moriones, E., Idris, A., Briddon, R. W., Rivera-Bustamante, R., Murilo Z. F., Martin, P. (2017). Capulavirus and Grablovirus: two new genera in the family Geminiviridae. *Archives of Virology*, *162*. https://doi.org/10.1007/s00705-017-3268-6

Varsani, A., Shepherd, D. N., Dent, K., Monjane, A. L., Rybicki, E. P., & Martin, D. P. (2009). A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. *Virology Journal*, *6*(1), 36. https://doi.org/10.1186/1743-422X-6-36

Varsani, A., Shepherd, D. N., Monjane, A. L., Owor, B. E., Erdmann, J. B., Rybicki, E. P., Peterschmitt, M., Briddon, R. W., Markham, P. G., Oluwafemi, S., Windram, O. P., Lefeuvre, P., Lett, J. M., & Martin, D. P. (2008). Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *Journal of General Virology*, *89*(9), 2063–2074. https://doi.org/10.1099/vir.0.2008/003590-0

Walt, E van der, Rybicki, E., Varsani, A., Polston, J., & Billharz, R. (2009). Rapid host adaptation by extensive recombination. *J Gen Virol*, *90*.

Wangia, C., Wangia, S., & Groote, H. De. (2002). Review of Maize Marketing in Kenya: Implementation and Impact of Liberalisation, 1989-1999. *Integrated Approaches to Higher Maize Productivity in the New Millenium. Proceedings of the 7th Eastern and Southern Africa Regional Maize Conference*, (February), 1989–1999.

Webster, C. L., Waldron, F. M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J. F., Obbard, D. J. (2015). The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. *PLOS Biology*, *13*(7), e1002210. https://doi.org/10.1371/

journal.pbio.1002210

Wetterdienst, D. (2017). Klimatafel von Toliary (Tulear) / Madagaskar. *Baseline Climate Means (1961-1990) from Stations All over the World*. Retrieved from http://www.dwd.de/DWD/klima/beratung/ak/ak_671610_kt.pdf

Willment, J. A., Martin, D. P., & Rybicki, E. P. (2001). Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data. *Journal of Virological Methods*, *93*(1–2), 75–87. https://doi.org/10.1016/S0166-0934(00)00299-8

Willment, J. A., Martin, D. P., Van der Walt, E., & Rybicki, E. P. (2002). Biological and Genomic Sequence Characterization of *Maize streak virus* Isolates from Wheat. *Phytopathology*, *92*(1), 81–86. https://doi.org/10.1094/PHYTO.2002.92.1.81

WITS. (2009a). Ethiopia(excludes Eritrea) Product Imports from Kenya| WITS Data. Retrieved from https://wits.worldbank.org/CountryProfile/en/Country/ETH/Year/2007/TradeFlow/Import/Partner/KEN/Product/all-groups

WITS. (2009b). Ethiopia(excludes Eritrea) Product Imports from Uganda | WITS Data. Retrieved from https://wits.worldbank.org/CountryProfile/en/Country/ETH/Year/2007/TradeFlow/Import/Partner/UGA/Product/all-groups

World Integrated Trade Solution. (2018). Product Exports by Madagascar to Comoros. Retrieved from https://wits.worldbank.org/CountryProfile/en/Country/MDG/Year/2018/TradeFlow/Export/Partner/COM/Product/all-groups

Worobey, M., Bjork, A., & Wertheim, J. O. (2007). Point, Counterpoint: The Evolution of Pathogenic Viruses and their Human Hosts. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 515–540. https://doi.org/10.1146/annurev.ecolsys.38.091206.095722

WTO. (2001). WTO | Trade policy review - Madasgascar 2001. Retrieved from https://www.wto.org/english/tratop_e/tpr_e/tp156_e.htm

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with

variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, *39*(3), 306–314. https://doi.org/10.1007/BF00160154

Yozwiak, N. L., Skewes-Cox, P., Stenglein, M. D., Balmaseda, A., Harris, E., & DeRisi, J. L. (2012). Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Neglected Tropical Diseases*, *6*(2), e1485. https://doi.org/10.1371/journal.pntd.0001485

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., & Varsani, A. (2017). ICTV virus taxonomy profile: Geminiviridae. *Journal of General Virology*, *98*(2), 131–133. https://doi.org/10.1099/jgv.0.000738

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W., McKenna, R. (2001). Structure of the Maize Streak Virus Geminate Particle. *Virology*, *279*(2), 471–477. https://doi.org/10.1006/viro.2000.0739

Supplementary information

Supplementary Table 1: Madagascan isolates list, accession numbers, sampling year, GPS coordinates, subtypes and their recombination lineages

Isolates	Accession	Sampling Year	Latitude	Longitude	Subtype	Recombination Linages
MSV_MG_Ant_ma31_10	KY311930	2010	-20.2108	44.4089	MSV-A_1	XI
MSV_MG_FOF_mad2_09	KY311956	2009	-23.3342	43.6894	MSV-A_1	V
MSV_MG_Tsi_mad5_09	KY311957	2009	-21.2013	48.224	MSV-A_1	IV
MSV_MG_FOF_mad11_09	KY311960	2009	-23.3342	43.6894	MSV-A_1	IV
MSV_MG_Sar_mad8_09	KY311958	2009	-23.2504	44.0184	MSV-A_1	IV
MSV_MG_Mor_mad21_09	KY311964	2009	-22.876	43.5811	MSV-A_1	IV
MSV_MG_FOF_mad10_09	KY311959	2009	-23.3342	43.6894	MSV-A_1	IV
MSV_MG_And_mad23_09	KY311965	2009	-23.445	43.9026	MSV-A_1	IV
MSV_KM_Zia_mad28_09	KY311967	2009	-12.2667	43.7023	MSV-A_1	IV
MSV_MG_Bef_mad13_09	KY311961	2009	-23.3186	43.7037	MSV-A_1	IV
MSV_MG_Tso_mad20_09	KY311963	2009	-23.327	43.6599	MSV-A_1	IV
MSV_MG_Tul_mad19_09	KY311962	2009	-23.3348	43.6786	MSV-A_1	IV
MSV_KM_Mra_mad35_09	KY311968	2009	-12.191	44.5072	MSV-A_1	XI
MSV_MG_Amb_ma40_10	KY311935	2010	-18.9836	46.5328	MSV-A_1	XI
MSV_MG_Amb_ma67_10	KY311947	2010	-17.6666	48.2112	MSV-A_1	XI
MSV_MG_Lab_ma74_10	KY311952	2010	-18.7066	48.4238	MSV-A_1	XI
MSV_KM_EED_mad26_09	KY311966	2009	-12.282	43.7344	MSV-A_1	XI
MSV_MG_Amb_ma75_10	KY311953	2010	-16.4887	46.7177	MSV-A_1	XI
MSV_MG_Ant_ma32_10	KY311931	2010	-20.2978	44.4088	MSV-A_1	XI
MSV_MG_Cam_ma10_10	KY311919	2010	-20.2117	44.4308	MSV-A_1	XI
MSV_MG_And_ma02_10	KY311914	2010	-19.5962	46.2218	MSV-A_1	XI
MSV_MG_Cam_ma8_10	KY311917	2010	-20.2117	44.4308	MSV-A_1	XI
MSV_MG_Cam_ma11_10	KY311920	2010	-20.2117	44.4308	MSV-A_1	XI
MSV_MG_Amb_ma59_10	KY311944	2010	-18.202	48.2588	MSV-A_1	XI
MSV_MG_Amb_ma60_10	KY311945	2010	-18.202	48.2588	MSV-A_1	XI
MSV_MG_Ava_ma26_10	KY311928	2010	-19.6807	44.548	MSV-A_1	XI
MSV_MG_Ava_ma24_10	KY311927	2010	-19.6807	44.548	MSV-A_1	XI
MSV_MG_And_ma04_10	KY311915	2010	-19.5962	46.2218	MSV-A_1	XI
MSV_MG_Cam_ma09_10	KY311918	2010	-20.2117	44.4308	MSV-A_1	XI
MSV_MG_Ana_ma37_10	KY311934	2010	-18.9777	46.7169	MSV-A_1	XI
MSV_MG_Soa_ma21_10	KY311924	2010	-19.6829	44.5389	MSV-A_1	XI
MSV_MG_Amb_ma53_10	KY311940	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma62_10	KY311946	2010	-18.202	48.2588	MSV-A_1	XI
MSV_MG_Bez_ma89_10	KY311955	2010	-12.654	49.2813	MSV-A_1	XI
MSV_MG_And_ma01_10	KY311913	2010	-19.5962	46.2218	MSV-A_1	XI

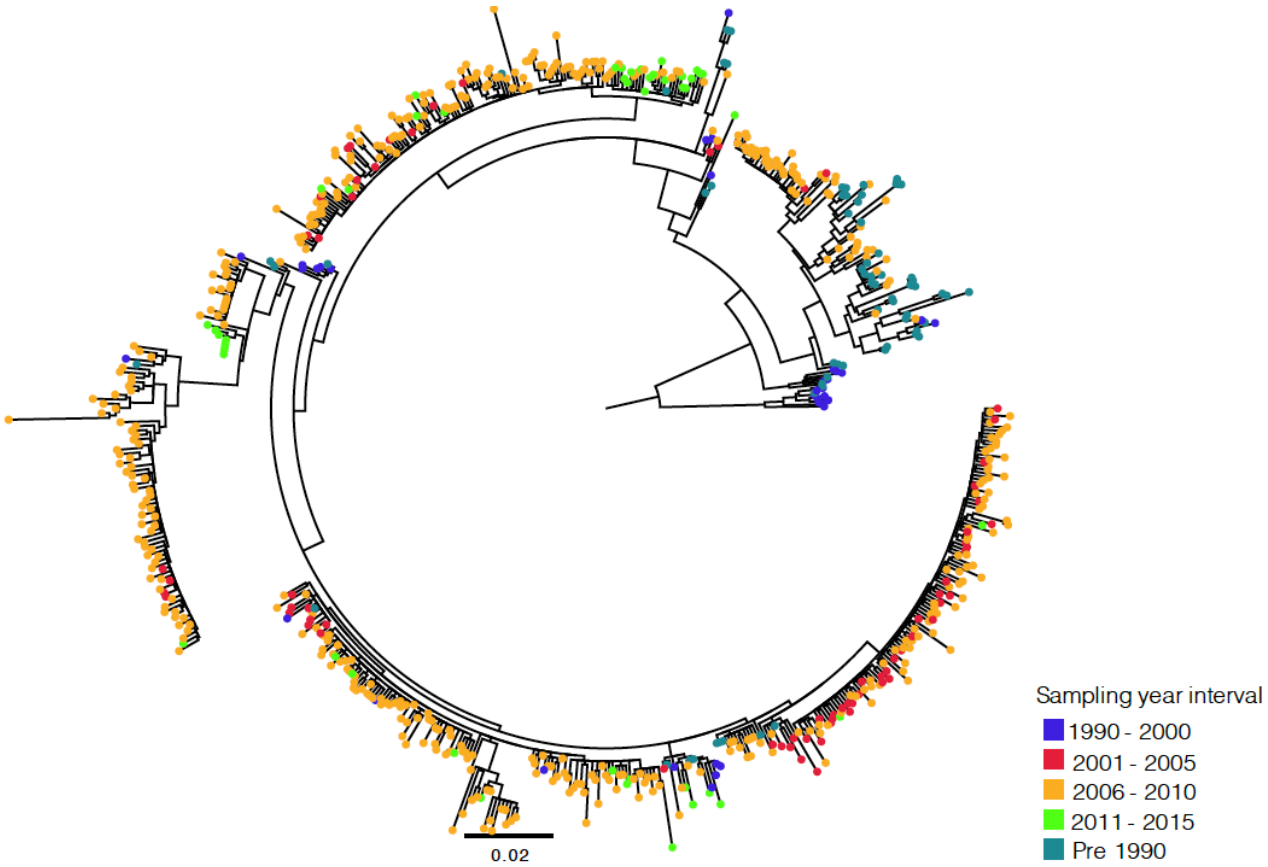
MSV_MG_Amb_ma35B_10	KY311933	2010	-18.9602	46.8182	MSV-A_1	XI
MSV_MG_Amb_ma73_10	KY311951	2010	-17.6674	48.2174	MSV-A_1	XI
MSV_MG_Amb_ma81_10	KY311954	2010	-13.6704	48.4483	MSV-A_1	XI
MSV_MG_Amb_ma45_10	KY311937	2010	-13.3573	48.8702	MSV-A_1	XI
MSV_MG_Lam_ma18_10	KY311922	2010	-19.9087	44.6003	MSV-A_1	XI
MSV_MG_Soa_ma22_10	KY311925	2010	-19.6829	44.5389	MSV-A_1	XI
MSV_MG_Soa_ma23_10	KY311926	2010	-19.6829	44.5389	MSV-A_1	XI
MSV_MG_Cam_ma12_10	KY311921	2010	-20.2117	44.4308	MSV-A_1	XI
MSV_MG_Amb_ma41_10	KY311936	2010	-18.9836	46.5328	MSV-A_1	XI
MSV_MG_Amb_ma35_10	KY311932	2010	-18.9602	46.8182	MSV-A_1	XI
MSV_MG_Ber_ma20_10	KY311923	2010	-19.7132	44.5387	MSV-A_1	XI
MSV_MG_Ank_ma29_10	KY311929	2010	-20.2151	44.422	MSV-A_1	XI
MSV_MG_Amb_ma51_10	KY311939	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma54a_10	KY311941	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma50_10	KY311938	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma54b_10	KY311942	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma54c_10	KY311943	2010	-17.6942	48.4656	MSV-A_1	XI
MSV_MG_Amb_ma69_10	KY311949	2010	-17.6666	48.2112	MSV-A_1	XI
MSV_MG_Amb_ma68_10	KY311948	2010	-17.6666	48.2112	MSV-A_1	XI
MSV_MG_Amb_ma72_10	KY311950	2010	-17.6674	48.2174	MSV-A_1	XI
MSV_MG_And_ma06_10	KY311916	2010	-19.5962	46.2218	MSV-A_1	XI

Supplementary Table 2: MSV-A isolates from Ethiopia, subtypes and their recombination lineages

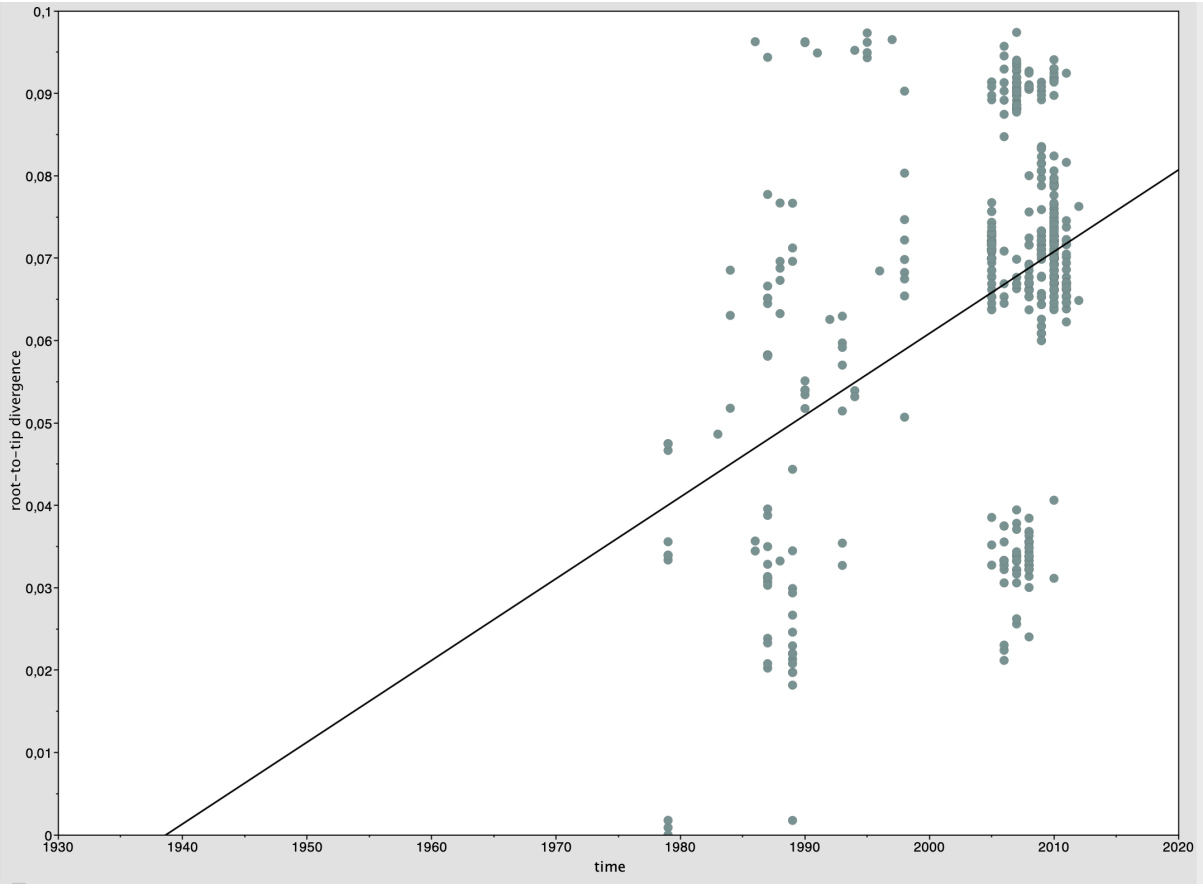
Isolate	Subtype	Recombination linages
ET9_2019	MSV-A_1	V
ET5_2019	MSV-A_1	V
ET11_SG	MSV-A_1	V
ET12_2019	MSV-A_1	V
ET13_2019	MSV-A_1	V
ET17_SG	MSV-A_1	V
ET19_2019	MSV-A_1	V
ET22_SG	MSV-A_1	V
ET34_SG	MSV-A_1	V
ET42_SG	MSV-A_1	V
ET46_2019	MSV-A_1	V
ET48_2019	MSV-A_1	V
ET51_2019	MSV-A_1	V
ET53_2019	MSV-A_1	V
ET54_2019	MSV-A_1	V
ET58_2019	MSV-A_1	V
ET59_SG	MSV-A_1	V
ET60_SG	MSV-A_1	V
ET62_2019	MSV-A_1	V
ET64_2019	MSV-A_1	V
ET65_SG	MSV-A_1	V
ET66_2019	MSV-A_1	V
ET67_2019	MSV-A_1	V
ET68_2019	MSV-A_1	V
ET70_2019	MSV-A_1	V
ET71_SG	MSV-A_1	V
ET72_2019	MSV-A_1	V
ET74_SG	MSV-A_1	V
ET75_SG	MSV-A_1	V
ET76_2019	MSV-A_1	V
ET77_2019	MSV-A_1	V
ET78_2019	MSV-A_1	V
ET79_SG	MSV-A_1	V
ET80_2019	MSV-A_1	V
ET81_SG	MSV-A_1	V
ET82_2019	MSV-A_1	V
ET84_2019	MSV-A_1	V

ET85_2019	MSV-A_1	V
ET86_SG	MSV-A_1	V
ET87_SG	MSV-A_1	V
ET88_2019	MSV-A_1	V
ET90_2019	MSV-A_1	V
ET91_2019	MSV-A_1	V
ET93_2019	MSV-A_1	V
ET95_2019	MSV-A_1	V
ET96_2019	MSV-A_1	V
ET97_2019	MSV-A_1	V
ET98_2019	MSV-A_1	V
ET99_2019	MSV-A_1	V
ET102_2019	MSV-A_1	V
ET103_2019	MSV-A_1	V
ET104_SG	MSV-A_1	V
ET18_2019	MSV-A_1	V
ET20_2019	MSV-A_1	V
ET21_2019	MSV-A_1	V
ET23_2019	MSV-A_1	V
ET24_SG	MSV-A_1	V
ET25_2019	MSV-A_1	V
ET26_SG	MSV-A_1	V
ET27_SG	MSV-A_1	V
ET3_2019	MSV-A_1	V
ET30_2019	MSV-A_1	V
ET33_SG	MSV-A_1	V
ET35_2019	MSV-A_1	V
ET36_2019	MSV-A_1	V
ET37_SG	MSV-A_1	V
ET38_SG	MSV-A_1	V
ET39_SG	MSV-A_1	V
ET4_SG	MSV-A_1	V
ET40_2019	MSV-A_1	V
ET44_SG	MSV-A_1	V
ET45_SG	MSV-A_1	V
ET47_2019	MSV-A_1	V
ET50_2019	MSV-A_1	V
ET52_2019	MSV-A_1	V
ET55_2019	MSV-A_1	V
ET56_2019	MSV-A_1	V
ET57_2019	MSV-A_1	V
ET63_2019	MSV-A_1	V

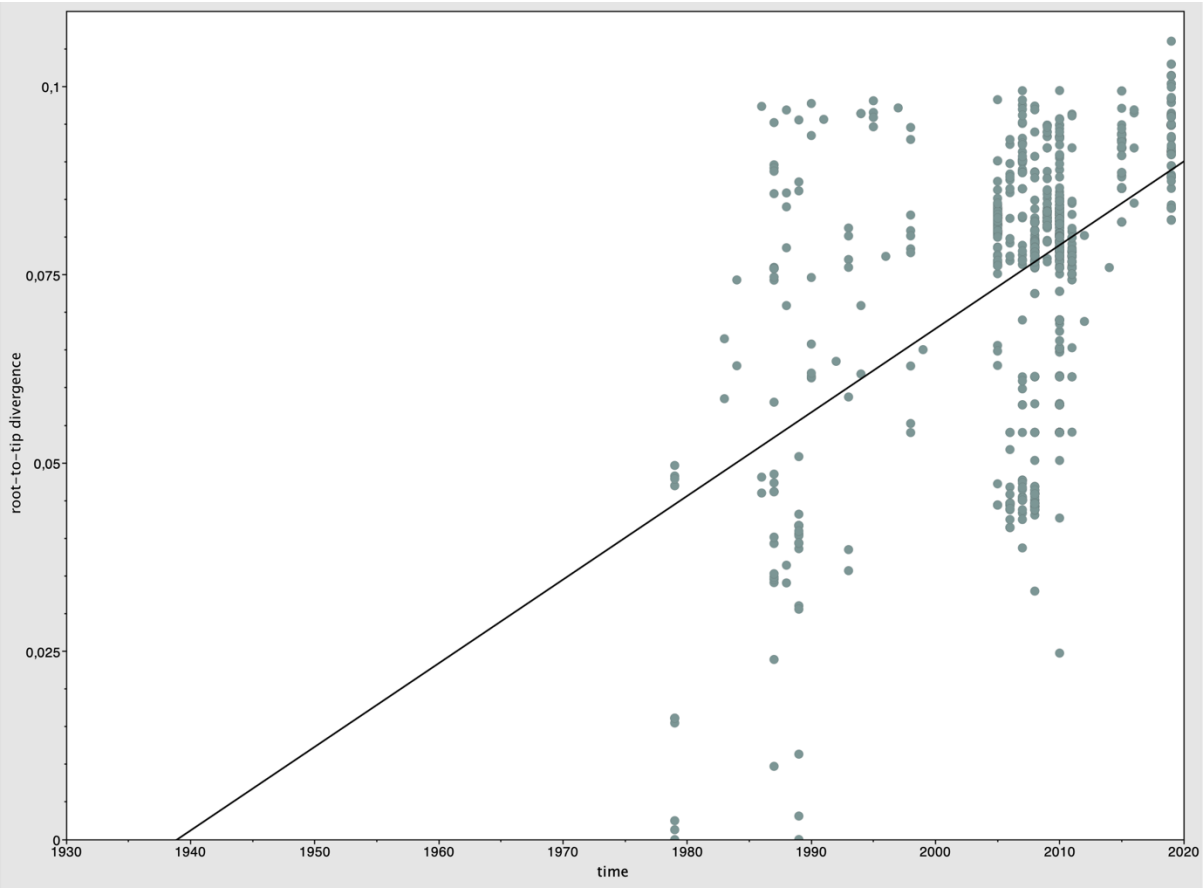
ET7_2019	MSV-A_1	V
ET89_2019	MSV-A_1	V
ET69_2019	MSV-A_1	V
ET8_2019	MSV-A_1	V
ET6_2019	MSV-A_1	V



Supplementary Figure 1: MSV-A maximum likelihood phylogeny visualized with sampling year interval. The phylogeny (model GTR+I+G_4) was rooted using the outgroup isolates from Reunion and Mauritius. The bar indicates the number of nucleotide substitutions per site.



Supplementary Figure 2: Regression plots of genetic distances versus sampling time for the MSV-A dataset used for Madagascan inference. The correlation coefficient is 0.333.



Supplementary Figure 3: Regression plots of genetic distances versus sampling time for the MSV-A dataset for Ethiopian and Rwandan inferences. The correlation coefficient is 0.505.

Appendix

Author's publications associated with the thesis

Oyeniran, K. A., Hartnady, P., Claverie, S., Lefeuvre, P., Monjane, A. L., Donaldson, L., Jean-Michel Lett., Arvind Varsani., & Martin, D. P. (2021). How virulent are emerging maize - infecting mastreviruses ? Archives of Virology, 0123456789.

Monjane, Adérito L, Dellicour, S., Hartnady, P., **Oyeniran, K. A**., Owor, B. E., Bezuidenhout, M., Daphne Linderme., Rizwan A. Syed., Lara Donaldson., Shane Murray., Edward P. Rybicki., Anders Kvarnheden., Elhman Yazdkhasti., Pierre Lefeuvre., Rémy Froissart., Philippe Roumagnac., Dionne N. Shepherd., Gordon W. Harkins., Marc A. Suchard., Philippe Lemey., Arvind Varsani., & Martin, D. P. (2020). Symptom evolution following the emergence of maize streak virus. *ELife*, *9*.

Fontenele, R.S.; Salywon, A.M.; Majure, L.C.; Cobb, I.N.; Bhaskara, A.; Avalos-Calleros, J.A.; Argüello-Astorga, G.R.; Schmidlin, K.; Khalifeh, A.; Smith, K.; Schreck, J.; Lund, M.C.; Köhler, M.; Wojciechowski, M.F; Hodgson, W.C.; Puente-Martinez, R.; Van Doorslaer, K.; Kumari, S.; **Oyeniran, K.A**.; Vernière, C.; Filloux, D.;Roumagnac, P.; Lefeuvre, P.; Ribeiro, G.S; Kraberger, S.P.; Martin, D.P.; and Varsani, A. New World Cactaceae Plants Harbor Diverse Geminiviruses. Viruses 2021, 13, 694. https://doi.org/10.3390/v13040694