The Phylogeography, Epidemiology and Determinants of

Maize streak virus dispersal across Africa and the adjacent Indian Ocean Islands

Eugene. T. Madzokere

A thesis submitted in partial fulfilment of the requirements for the degree of

MAGISTER SCIENTIAE (M.Sc.)

South African National Bioinformatics Institute (SANBI)

University of the Western Cape

Supervisors: Dr. Gordon W. Harkins and Professor Darren P. Martin

December 2015

KEYWORDS

Maize streak virus

Maize streak disease

Phylogeography

Epidemiology

Source-sink dynamics

ABBREVIATIONS

AFTZ	African Free Trade Zone
AICc	corrected Akaike Information Criterion
AN	Anjouan
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BF	Bayes Factor
BF	Burkina Faso
BJ	Benin
BSSVS	Bayesian Stochastic Search Variable Selection
CABI	Commonwealth Agricultural Beaurex International
CBIO	Computational Biology Unit
CF	Central African Republic
CIMMYT	International Maize and Wheat Improvement Center
CIRAD	French Agricultural Research Center for International Development
CM	Cameroon
CMG	*Cassava mosaic geminiviruses*
COMESA	Common Market for Eastern and Southern Africa
CP	Coat Protein
CRI	Crop Research Institute
CSIR	Centre for Scientific and Industrial Research
CTMC	Continuous time Markov chain
DNA	Deoxyribonucleic Acid
EAC	East African Community
EPPO	European and Mediterranean Plant Protection Organization
ERS	Export Retention Scheme

ESS	Effective Sample Size
ET	Ethiopia
FAO	Food and Agricultural Organization
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
FICA	Farm Inputs Care Centre Limited
FUBAR	Fast Unbiased Bayesian Approximation
FG	Full Genome
GARD	Genetic Algorithm for Recombination Detection
GDP	Gross Domestic Product
GH	Ghana
GLM	Generalized Linear Model
GMRF	Gaussian Markov random field
GTR+G+I	General Time Reversible + Gamma + Invariant sites
HKY	Hasegawa Kishino Yano
HME	Harmonic mean estimate
HPD	Highest Posterior Density
HyPhy	Hypothesis testing using Phylogenies
IITA	International Institute of Tropical Agriculture
IMPALE	Improved Alignment Editor
IOC	Indian Ocean Commission
IR	Intergenic Region
IRAT	Institute of Agronomic and Tropical Research
JIC	John Innes Centre
KARI	Kenyan Agricultural Research Institute
KE	Kenya

KH	Kishino-Hasegawa
KML	key markup language
LIR	Long Intergenic Region
LS	Lesotho
MBYMIV	*Mung bean yellow mosaic India virus*
MCB	Molecular and Cell Biology
MCC	Maximum Clade Credibility
MCMC	Metropolis-Hastings Markov chain Monte Carlo
MD	Madagascar
MEGA	Molecular Evolutionary Genetics Analysis
MH	Moheli
ML	Maximum Likelihood
MP	Movement Protein
MP	Maximum Parsimony
MPCP	Movement Protein-Coat Protein
MRCA	Most Recent Common Ancestor
MSD	Maize streak disease
MSV	*Maize streak virus*
MSV-A	*Maize streak virus* subtype strain A
MZ	Mozambique
NASECO	Nalweyo Seed Company Limited
NRF	National Research Foundation
NCBI	National Center for Biotechnology Information
OECD	Organization for Economic Cooperation and Development

PARRIS	Partitioning approach for Robust Inference of Selection
PRF	Poliomyelitis Research Foundation
PS	Path Sampling
RCR	Rolling Circle Replication
RDP	Recombination Detection Program
RDR	Recombination Dependent Replication
RTC	Regional Trade Communities
SADC	Southern African Development Community
SANBI	South African National Bioinformatics Institute
SARI	Savanna Agricultural Research Institute
SCRI	Scottish Crop Research Institute
SDT	Sequence Demarcation Tool
SIR	Short Intergenic Region
SPREAD	Spatial Reconstruction of Evolutionary Dynamics
SS	Stepping stone
ssDNA	single-stranded Deoxyribonucleic Acid
TD	Chad
tMRCA	Time to the Most Recent Common Ancestor
TYLCV	*Tomato yellow leaf curl virus*
UCT	University of Cape Town
UG	Uganda
US/ USA	United States of America
USAID	United States Agency for International Development
USP	Uganda Seed Project
UWC	University of the Western Cape

WAEMU	West African Economic and Monetary Union
WNV	West Nile virus
ZM	Zambia
ZW	Zimbabwe

ABSTRACT

The Phylogeography, Epidemiology and determinants of *Maize streak virus* dispersal across Africa and the adjacent Indian Ocean Islands

E. T. Madzokere

Master of Science Thesis, South African National Bioinformatics Institute,

University of the Western Cape

Maize streak disease (MSD), caused by variants of the *Maize streak virus* (MSV) A strain, is the world's third and Africa's most important maize foliar disease. Outbreaks of the disease occur frequently and in an erratic fashion across Africa and Islands in the Indian Ocean causing devastating yield losses such that the emergence, resurgence and rapid diffusion of MSV-A variants in this region presents a serious threat to maize production, farmer livelihoods and food security. To compliment current MSD management systems, a total of 689 MSV-A full genomes sampled over a 32 year period (1979-2011) from 20 countries across Africa and the adjacent Indian Ocean Islands, 286 of which were novel, were used to estimate: (i) the levels of genetic diversity using MEGA and the Sequence Demarcation Tool v1.2 (SDT); (ii) the times of occurrence and distribution of recombination using the recombination detection program (RDP v.4) and the genetic algorithm for recombination detection (GARD); (iii) selection pressure on codon positions using PARRIS and FUBAR methods implemented on the DATAMONKEY web server; (iv) reconstruct the history of spatio-temporal diffusion for MSV-A using the discrete phylogeographic models implemented in BEAST v1.8.1; (v) characterize source-sink dynamics and identify predictor variables driving MSV-A dispersal using the generalized linear models, again implemented in BEAST v1.8.1.

Isolates used displayed low levels of genetic diversity (0.017 mean pairwise distance and ≥ 98% nucleotide sequence identities), and a well-structured geographical distribution where all of the 233 novel isolates clustered together with the -A_1 strains. A total of 34 MSV inter-strain recombination events and 33 MSV-A intra-strain recombination events, 15 of which have not been reported in previous analyses (Owor *et al*., 2007, Varsani *et al*., 2008 and Monjane *et al*., 2011), were detected. The majority of intra-strain MSV-A recombination events detected were inferred to have

occurred within the last six decades, the oldest and most conserved of these being events 19, 26 and 28 whereas the most recent events were 8, 16, 17, 21, 23, and 29. Intra-strain recombination events 20, 25 and 33, were widely distributed amongst East African MSV-A samples, whereas events 16, 21 and 23, occurred more frequently within West African MSV-A samples. Events 1, 4, 8, 10, 14, 17, 19, 22, 24, 25, 26, 28, and 29 were more widely distributed across East, West and Southern Africa and the adjacent Indian Ocean Islands. Whereas codon positions 12 and 19 within motif I in the coat protein transcript, and four out of the seven codon positions (147, 166, 195, 203, 242, 262, 267) in the Rep transcript (codons 195 and 203 in the Rb motif and codons 262 and 267 in site B of motif IV), evolved under strong positive selection pressure, those in the movement protein (MP) and RepA protein encoding genes evolved neutrally and under negative selection pressure respectively.

Phylogeographic analyses revealed that MSV-A first emerged in Zimbabwe around 1938 (95% HPD 1904 - 1956), and its dispersal across Africa and the adjacent Indian Ocean Islands was achieved through approximately 34 migration events, 19 of which were statistically supported using Bayes factor (BF) tests. The higher than previously reported mean nucleotide substitution rate [9.922×10^{-4} (95% HPD 8.54×10^{-4} to 1.1317×10^{-3}) substitutions per site per year)] for the full genome recombination-free MSV-A dataset H estimated was possibly a result of high nucleotide substitution rates being conserved among geminiviruses such as MSV as previously suggested. Persistence of MSV-A was highest in source locations that include Zimbabwe, followed by South Africa, Uganda, and Kenya. These locations were characterized by high average annual precipitation; moderately high average annual temperatures; high seasonal changes; high maize yield; high prevalence of undernourishment; low trade imports and exports; high GDP per capita; low vector control pesticide usage; high percentage forest land area; low percentage arable land; high population densities, and were in close proximity to sink locations. Dispersal of MSV-A was frequent between locations that received high average annual rainfall, had high percentage forest land area, occupied high latitudes and experienced similar climatic seasons, had high GDP per capita and had balanced maize import to export ratios, and were in close geographical proximity.

December 2015

DECLARATION

I declare that, *The Phylogeography, Epidemiology, and Determinants of Maize streak virus dispersal across Africa and the adjacent Indian Ocean Islands*, is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Eugene. T. Madzokere December 2015

Signed:.....

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisors Dr. Gordon William Harkins and Professor Darren Martin for their patience, exceptional guidance and support during the study period. I would also like to thank Dr. Arvind Varsani for sequencing and providing data on both novel and publicly available samples used in the study. The financial assistance of the National Research Foundation (NRF), the Poliomyelitis Research Foundation (PRF), and the Thuthuka Board towards this research is hereby acknowledged. Opinions expressed and arrived at, are those of the author and are not necessarily to be attributed to the NRF, PRF and/or the Thuthuka Board.

My deepest gratitude goes to the University of the Western Cape's (UWC) Computer Science and Physics Departments and the University of Cape Town's (UCT) Computer Science Department, the South African Center for High Performance Computing (CHPC), and the South African National Bioinformatics Institute (SANBI), for their provision of computational resources and platforms on which to analyze MSV datasets.

Thank you Leendert Cloete, Batsirai Mabvakure, Emil Tanov, Brejnev Muhire, Peter van Heusden, Dr. Dionne Shepherd, Dane Kennedy, Inus Scheepers, and Mariam Salie, for all the assistance you gave me throughout the project. Last, but not least, I am also very grateful to my family and friends for their love, prayers and unwavering support.

December 2015

CONTENTS

List of Figures

List of Tables

APPENDICES

List of Figures

List of Tables

PREFACE

Maize streak disease (MSD), caused by *Maize streak virus* (MSV; family *Geminiviridae*, genus *Mastrevirus*: Bock, 1974; McClean, 1947; Mullineaux *et al*., 1984), is the world's third and Africa's most important maize foliar disease (Bousque-Perez, 2000; Pratt and Gordon, 2006). Unfortunately, *Cicadulina* leafhopper vector species implicated in the transmission of MSV are widely distributed across Africa and the adjacent Indian Ocean Islands, and since the first report of MSD in the 1870s (Fuller, 1901; Monjane *et al*., 2011), more than 20 countries in this region have reported symptoms resembling those of MSD, the presence of MSV, and the occurrence of numerous disease outbreaks which are difficult to predict and are influenced by several interacting factors (Esenam *et al*., 1966; Fajemisin *et al*., 1976; Teklamariam *et al*., 1986; Efron *et al*., 1989; Cardwell *et al*., 1997; Caulfield, 1994; Martin and Shepherd, 2009; Oppong *et al*., 2015; Owor, 2008). At their worst, MSD outbreaks can cause 100% yield losses, with losses ranging from 6%-10% being estimated to cost farmers between US$120-US$400 million per epidemic year (Martin and Shepherd, 2009). Thus, introduction, resurgence and rapid diffusion of MSV between countries across this region, where maize is widely grown and consumed as a staple, threatens maize production, food security and more than 100 million livelihoods.

As such, to complement current MSD control strategies, elucidate MSV-A migration routes and possibly mitigate future spread of MSV-A variants, I analyzed the genetic diversity of 689 MSV-A isolates sampled from 20 locations over a 32 year period (1979-2011) across Africa and the adjacent Indian Ocean Islands where MSD is endemic, and performed phylogeographic analyses on one mostly recombination-free dataset (H) using discrete (Lemey *et al*., 2009) phylogeography diffusion models implemented in the evolutionary analyses program BEAST (Drummond and Rambaut, 2007). In addition to mapping the geographical distribution of MSV-A, intra-strain recombination events, dates of their probable occurrence, the selection pressure acting at codon positions in the MSV-A genome, and the historical movement pathways of MSV-A variants, I also characterized the source-sink dynamics of MSV-A dispersal using the generalized linear model (GLM) in BEAST and identified predictor variables potentially acting

as key determinants in the spatio-temporal diffusion of MSV-A across Africa and the adjacent Indian Ocean Islands.

1.0 INTRODUCTION

1.1 Maize streak disease (MSD)

Maize streak disease (MSD) is the world's third and Africa's most important maize foliar disease (Bousque-Perez, 2000; Pratt and Gordon, 2006). It occurs when maize is infected by the single-stranded DNA (ssDNA) *Maize streak virus* (MSV; family *Geminiviridae*, genus *Mastrevirus*: Appendix 1 - Figure 1.1: Howell, 1984; Lazarowitz, 1988; Mullineaux *et al*., 1984; Zhang *et al*., 2001), which is obligately transmitted by nine distinct species of cicadellid leafhopper within the genus *Cicadulina* (Dabrowksi, 1987; Webb, 1987).

In maize, symptoms of MSD include morphological teratology which is characterized by leaf margin splitting, broad-pale, cream colored to light green or yellow parallel leaf vein streaking, tip twisting, necrosis of emerging leaves, reduced leaf size, tassel sterility and shoot stunting (Bock *et al*., 1974; Damsteegt, 1981; Fajemisin, 1984; Storey, 1925; Oppong *et al*., 2015). MSD symptoms resembling those recorded by Fuller (1901) in the Natal outbreak and direct confirmation of MSV through molecular sequencing has been reported in more than 20 African and adjacent Indian Ocean Island countries (Fajemisin *et al*., 1976; Goodman, 1981; Kim *et al*., 1981; Kim *et al*., 1989; Lazarowitz, 1987, 1988; Malithano *et al*., 1997; Rossel and Thottapilly, 1985; Owor, 2008; Cardwell *et al*., 1997; Oppong *et al*., 2015). The most recent MSD outbreak was reported in Ghana in 2010 (Oppong *et al*., 2015), and although MSD epidemic outbreaks can cause 100% maize yield losses for individual farmers, on average, countrywide MSD incidence is estimated to be well below 40% in most epidemic years and below 5% in non-epidemic years (Martin and Shepherd, 2009). For example, during the 2005 Ugandan MSD epidemic, an ~30% countrywide MSD incidence was recorded (Owor, 2008), whereas a proportion of only ~2% maize plants experienced MSD infections in Cameroon during 1993, a non-epidemic year (Cardwell *et al*., 1997).

It has also been calculated that on average, maize yield losses caused by MSD ranging from 6% to 10% cost African farmers between US$120 and US$480 million per epidemic year (Martin and Shepherd, 2009). As a result, the emergence, resurgence, and rapid dispersal of MSV strains that cause severe MSD, continues to seriously threaten the food security and livelihoods of farmers across Africa and the adjacent Indian Ocean Islands.

1.2 *Maize streak virus* (MSV)

Eleven major strains of MSV (MSV-A to MSV-K) are currently known to exist (Appendix 3 - Figure 1.2) (Martin *et al*., 2001; Schnippenkoetter *et al*., 2001; Varsani *et al*., 2008; Willment *et al*., 2001), but only the five variants or subtypes of the MSV-A strain (MSV-A_1, -A_2, -A_3, -A_4, -A_6) infect maize causing severe MSD symptoms. Strains from MSV-B to MSV-K infect mostly wild grasses and/or cereals (McClean, 1947; Storey and McClean, 1930; Martin and Shepherd, 2009; Shepherd *et al*., 2010; Monjane *et al*., 2011; Oppong *et al*., 2015). However, MSV-B, MSV-C, MSV-D, and MSV-E strains produce mild MSD symptoms in maize (Damsteegt, 1983; ICTVdb Management, 2006; Martin *et al*., 1999, 2001; Konate and Traore, 1992). Thresholds of >78% and >94% genome-wide pair-wise sequence identity are adopted as demarcations of *Mastrevirus* species and strains respectively (Muhire *et al*., 2014).

The closest relatives of MSV are the seven African streak virus species (Bock *et al*., 1974; Shepherd *et al*., 2010; Pande *et al*., 2012). These mostly infect wild grasses and cereals and include *Axonopus compressus streak virus* (ACSV; Oluwafemi *et al*., 2014), *Eragostris streak virus* (ESV), *Sugarcane streak virus* (SSV), *Sugarcane streak Reunion virus* (SSRV), *Urochloa streak virus* (USV), *Sugarcane streak Egypt virus* (SSEV) and *Panicum streak virus* (PanSV) (Appendix 3 - Figure 1.2; Shepherd *et al*., 2010). While MSVs are widely distributed across Africa and the adjacent Indian Ocean Islands (Appendix 4 - Table 1.2: Shepherd *et al*., 2010), their existence in the South and South-eastern Asian region within India, Indonesia and Yemen has also been reported (Appendix 5 - Table 1.3: Bock *et al*., 1974; Brunt *et al*., 1990; EPPO, 2014; CABI and

EPPO, 1997), although no publicly available full genome sequences exist in NCBI to confirm their existence there. However, the geographical distribution of the five epidemiologically relevant MSV-A subtype strains across Africa and the adjacent Indian Ocean Islands and their severities on different maize genotypes varies greatly. MSV-A_1 strains cause the most severe MSD symptoms in maize and have the widest geographical distribution throughout mainland Africa (Shepherd *et al*., 2010). In contrast, MSV-A_2 strains are restricted to West Africa; -A_3 strains to East Africa; -A_4 strains to Southern Africa; and -A_6 strains to Islands in the Indian Ocean (Shepherd *et al*., 2010). Identifying factors that underlie this spatially restricted and well structured MSV-A strain distribution pattern in this region may expand our current knowledge of MSD epidemiology and improve our ability to predict and manage MSD outbreaks (Martin *et al*., 2001; Monjane *et al*., 2011; Owor *et al*., 2007; Varsani *et al*., 2008; Willment *et al*., 2001).

1.3 Viral Diversity

Noticeably, several processes have contributed to the diversification of geminiviruses such as MSV. These include mechanistic processes such as mutation, recombination and re-assortment (Padidam *et al*., 1999) and population processes such as genetic drift and diversifying selection (Lefeuvre *et al*., 2011). Amongst the mechanistic processes, recombination is ubiquitous within partially conserved and peripherally distributed hotspots in the MSV genome and may occur at interspecies, inter-strain and intra-strain levels where it can potentially provide a selective advantage in the evolution and emergence of new geminiviruses (Padidam *et al*., 1999; Lefeuvre *et al*., 2009; Varsani *et al*., 2008; Monjane *et al*., 2011). For example, the MSV-A strain is reported to be a product of an ancient recombination event that occurred between MSV-B and MSV-G/F variants sometime around 1870 (Monjane *et al*., 2011).

1.4 Genome Organization

Maize streak virus (MSV) has an approximately 2.7 kilobase sized, circular, single stranded DNA (ssDNA) genome (Appendix 1 - Figure 1.1: Howell, 1984; Lazarowitz, 1988; Mullineaux *et al*., 1984; Zhang *et al*., 2001), that replicates via rolling circle and recombination-dependent mechanisms (Preiss and Jeske, 2003), and encodes four proteins: (i) a Movement protein (MP) through gene V1; (ii) a Coat protein through gene V2; and two Replication associated proteins, (iii) Rep, encoded by post-transcriptionally spliced C1 and C2 genes, and (iv) RepA, which is only encoded by the C1 gene (Pratt and Gordon, 2006). Bidirectional transcription from the long intergenic region (LIR) leads to virion sense expression of the MP and the CP and the complementary-sense expression of the replication-associated proteins, Rep and RepA respectively. These four proteins each play significant roles during host infection. Movement protein (MP) and coat protein (CP) encoding genes are required for systemic infection of host plants by MSV (Boulton *et al*., 1991a, and 1991b; Lazarowitz *et al*., 1989; Woolston *et al*., 1989). The MP facilitates cell-to-cell movement of virus within the host, whereas the CP is required both for entry of viral DNA into the nucleus and the inter-cellular movement of viral DNA, whilst the replication-associated proteins, Rep and RepA, enable usurping of the host replication machinery and rapid production of high copy numbers of viral progeny (Preiss and Jeske, 2003; Zhang *et al*., 2001).

1.5 The Host: Maize

Maize (*Zea mays* L; family *Poaceae* and tribe *Maydeae*) is not an indigenous African plant, instead its origins have been traced back to the Mesoamerican region, now Mexico and Central America (Matsuoka *et al*., 2002; Piperno and Flannery, 2001), and teosinte (*Z. mexicana*) is believed to be the ancestor of the crop plant (Warburton *et al*., 2011). Maize was first introduced into Nigeria in West Africa, by the Portuguese in the 16th century and thereafter into southern Africa by the Dutch East India Company in the middle of the 17th century (Jeffreys, 1963; McCann, 2001). The first disease reports of symptoms resembling those of contemporary MSV infection were in the Natal region in

South Africa (Fuller 1901), indicating that within approximately 200 years of the introduction of this crop plant into Southern Africa, a viral pathogen had emerged with the capacity to infect maize and subsequently spread to more than 20 African and Indian Ocean Island countries. To date, maize is cultivated in 46 of the 54 African countries, and because of its high economic (Andreu *et al*., 2006; Pingali, 2001; Sleper and Poehlman, 2006), and nutritional value (Prasanna *et al*., 2001; Rosegrant, 2008), it is an important staple food across this continent and Islands in the Indian Ocean (Morris *et al*., 1999; WABS, 2008). In addition to large numbers of large-scale commercial farmers, maize is also cultivated by more than 100 million subsistence farmers in this region both for consumption and economic empowerment (FAO, 2007; Martin and Shepherd, 2009) and its production in this region exceeds that of cereals such as wheat, rice, millet and sorghum (Sleper and Poehlman, 2006). Following the emergence of MSV-A in southern or east-central Africa (Fuller, 1901; Harkins *et al*., 2009 and Monjane *et al*., 2011), subsequent MSD outbreaks have periodically occurred at local, regional and/or countrywide levels in a seemingly erratic fashion, often resulting in devastating crop losses for farmers across Africa and Islands in the Indian Ocean (Martin and Shepherd, 2009; Shepherd *et al*., 2010; Monjane *et al*., 2011). Currently, the top maize producers in this region are South Africa, followed by Nigeria, Ethiopia, Tanzania, Kenya, Malawi, and Zambia (Table 1.1: FAOSTAT, 2012).

1.6 MSD Epidemiology

A number of factors that may broadly be classified as ecological, climatic, genetic, sociopolitical, economic, and physical, have been observed to affect and influence MSD epidemiology (Martin and Shepherd, 2009; Shepherd *et al*., 2010). Extremely complex interactions between these factors, which appear to converge every three to ten years, are believed to produce conditions that promote MSV dispersal and an increase in MSD incidence on the continent and the adjacent Indian Ocean Islands (Efron *et al*., 1989). MSD outbreaks have largely been difficult to predict firstly because of the erratic manner in which they occur, and secondly, because the aforementioned factors include interactions among multiple MSV and African streak virus strains that increase the

diversity, distribution and possibly the range of plant hosts accessible to epidemiologically relevant MSV-A strain variants; the distribution, host range and interactions of the nine viruliferous vector species of cicadellid leafhoppers in the Genus *Cicadulina* (Rose, 1974; Dabrowski *et al*., 1987; Fennah, 1959; Nielson, 1986; Ruppel, 1965; Soto, 1978), and finally the ability of such viruses to persist in over 80 grass species (ICTVdb Management, 2006; Damsteegt, 1983; Konate and Traore, 1992).

Furthermore, it is also clear that climatic (temperature, rainfall, relative humidity) and geographical factors that influence the composition of grass and leafhopper populations also add to the complexity of MSD epidemiology (Dabrowski, 1987; Reynaud *et al*., 2009). This is reflected by reports suggesting that MSD outbreaks occur more frequently in locations that (i) lie anywhere from sea level up to an altitude of 2000 meters (Magenya *et al*., 2008), (ii) have high average annual temperatures and precipitation (Asanzi *et al*., 1994; Okoth and Dabrowski, 1987; Rose 1972) (iii) where drought conditions are followed by irregular rains at the beginning of the growing season (Bjarnason, 1986; Welz *et al*., 1998; Efron *et al*., 1989), (iv) are MSV diversification hotspots (such as eastern and southern Africa: Monjane *et al*., 2011), and (v) during the second season where there are two maize growing seasons a year (Martin and Shepherd, 2009). Economic factors that include exorbitant pesticide prices, poorly implemented MSD agronomic management practices, and systematic flooding of the African maize seed-market with low-to-negligible MSV - resistant varieties by seed companies and traders also make it difficult to understand MSD epidemiology and predict the occurrence of outbreaks (Martin and Shepherd, 2009). An unfortunate consequence of the interaction and changes in the factors influencing MSD epidemiology and the frequency of disease outbreak occurrence across Africa and Islands in the Indian Ocean is that, it complicates the proper scheduling of the maize planting and harvesting seasonal calendar by farmers seeking to avoid huge yield losses to MSD.

1.7 MSD Management

In an attempt to mitigate the growing threat posed by MSD on food security in this region, farmers have resorted to implementing an integrated pest management system (Damsteegt, 1983; Shepherd *et al*., 2010). Such a system typically includes cultivation of treated and certified maize seed; use of MSV tolerant or resistant (usually conventionally bred or transgenic) varieties; early rouging; chemical control of the MSV leafhopper vectors through application of pesticides (Rose, 1973; Magenya *et al*., 2008; Karavina, 2014); and rotation with broadleaved crops such as groundnuts, beans, cowpeas, cotton and pumpkin which appear immune to MSV (Damsteegt, 1983 and Shepherd *et al*., 2010). The primary goal of such systems is to guide farmers in appropriately scheduling their planting and harvesting calendars from season to season so as to avoid and/or minimize yield losses due to MSD. However, for this goal to be realized, farmers, maize seed and agrochemical companies, governments and research institutes within countries where MSD occurs periodically, must seriously commit to implementing, monitoring and updating such systems. Unfortunately, most African countries, with the exception of South Africa and Nigeria, do not have robust legislative frameworks and resources to structure the implementation, monitoring and updating of MSD surveillance systems.

As a process, disease surveillance involves the systematic collection, analysis, interpretation and distribution of large volumes of data (usually raw health or agricultural data) originating from a variety of sources, for the purpose of planning, implementing, and evaluating public health, agronomic and/or other interventions, with the ultimate goal of reducing the morbidity and mortality of susceptible host populations (Scallan *et al*., 2011a; Trifonov *et al*., 2009). A successful disease surveillance system uses data captured to: (i) evaluate the effectiveness of control and preventative measures; (ii) monitor changes in infectious agents such as trends in disease development or occurrence of new, highly viruliferous viral strains over time, including, (iii) the occurrence and frequency of disease outbreaks (Torok and Anderson, 2008). Estimates of baseline levels of disease obtained from such surveillance systems may permit identification of high-risk populations, areas and/or climatic seasons to target interventions and as a consequence,

such estimates act as important references for future outbreaks and may guide policy development and maize production (Janes *et al*., 2000).

One obvious consequence of not having surveillance systems is avoidable maize yield losses and without a sustained level of commitment to MSD surveillance, each country in this region risks multiple introductions and/or re-introductions of epidemiologically relevant variants of the MSV-A strain. Trans-continental movements of such variants are most likely to occur and increase between countries that: (i) either have weak or no MSD management and/or epidemic outbreak surveillance systems in implementation; (ii) have weak border control and trade policies on trans-boundary movement of plant material such as maize; (iii) share physical borders or ports with a country that has attributes in parts (i) and (ii) and trade frequently (Monjane *et al*., 2011). However, where present, data from a structured MSD surveillance system can complement phylogeographic methods in elucidating the extent to which factors stated above influence the rate of MSV-A dispersal between source and sink locations.

1.8 The MSV Vector

Cicadulina species are widely distributed throughout tropical and subtropical Africa (CABI, 1986; Dabrowksi, 1987; Mylonas, *et al*., 2014; Oluwafemi *et al*., 2007; Reynaud *et al*., 2009; Rose, 1978). Amongst these species, *C. mbila* and *C. storeyi* are the most viruliferous, particularly the females and large-winged individuals capable of long distance flight and long-range MSV dispersal (CABI, 1986; Dabrowksi, 1987; Mylonas, *et al*., 2014). *C. mbila* occurs in more than 20 countries across Africa and the adjacent Indian Ocean Islands and its development, distribution and persistence is mostly affected by fluctuations in temperature and precipitation (Rose, 1973a), and also by changes in land-use patterns (CABI, 1986; Reynaud *et al*., 2009; Webb, 1987; Mylonas *et al*., 2014).

1.9 MSV Infectious Disease Dynamics

Elucidation of the factors that shape the spread and evolution of diseases such as MSD will increase our understanding of the dynamics of viral infections and inform prevention and control measures (Pybus and Rambaut, 2009). Traditionally, viral infectious disease dynamics have been investigated using epidemiological methods (Pybus and Rambaut, 2009), however, with recent concurrent advances in whole genome sequencing, mathematical modeling and computer processing power, the use of evolutionary approaches unified with traditional epidemiological methods have increasingly been used to investigate viral infectious disease dynamics (Grenfell *et al*., 2004; Pybus and Rambaut, 2009; Pybus *et al*., 2012).

There are several advantages to unifying evolutionary approaches with traditional epidemiological methods and employing both to investigate viral infectious disease dynamics. First of all, because evolutionary approaches allow the reconstruction of the demographic history of an entire epidemic whether there is very little or no surveillance data, they complement traditional epidemiological methods; secondly, evolutionary approaches require fewer samples of viral pathogens to achieve the latter and estimate population parameters such as the dispersal and nucleotide substitution rates; and thirdly, through use of these approaches, a plausible history of the migration routes used by the virus to attain its geographical distribution can be reconstructed. Finally, it is also possible to infer epidemiological linkages among infections in time and space using evolutionary approaches (Pybus and Rambaut, 2009). Moreover, because evolutionary approaches are an integral component of phylogeographic analyses methods, it is now possible to investigate the spatial and temporal patterns present in viral phylogenies (Lemey *et al*., 2009; Lemey *et al*., 2010; Lemey *et al*., 2014; Monjane *et al*., 2011).

It is important to note that, phylogeographic analyses involve a joint estimation of both the virus phylogenetic tree that represents the evolutionary relationships between sampled pathogens and the locations of un-sampled, most recent common ancestors, thereby producing a full history of viral dispersal (Grenfell *et al*., 2004; Pybus *et al*., 2009). Once

identified, the spatio-temporal patterns embedded within viral phylogenies can be matched to historically dated and/or confirmed reports of epidemic outbreaks as well as interventions to control MSD incidence amongst the sampling locations, which ultimately allow us to elucidate the relationship between these patterns and the occurrence of MSD outbreaks. Phylogeographic analyses methods implemented in the evolutionary software Bayesian Evolutionary Analysis by Sampling Trees (BEAST: Drummond and Rambaut, 2007), allow for the reconstruction of the dispersal history of viral pathogens among discrete locations (Lemey *et al*., 2009) and in a continuous space (Lemey *et al*., 2010). They also yield a Bayes factor (BF) statistical support for the best-supported epidemiological linkages between locations (Lemey *et al*., 2009), and are able to account for spatial and temporal uncertainty in phylogenetic tree reconstruction by evaluating ancestral reconstructions over a posterior distribution of trees as opposed to doing so based on a single tree (Baele *et al*., 2012; Pagel *et al*., 2004). In addition to accounting for phylogenetic uncertainty in tree reconstruction, Bayesian analyses with BEAST are more computationally efficient and therefore widely used compared to Hill-climbing methods such as Maximum Likelihood (ML) and Maximum Parsimony (MP) in reconstructing the spatio-temporal diffusion history of viral pathogens (Baele *et al*., 2012; Drummond and Bouckaert, 2014).

Such phylogeographic methods provide a platform to test evolutionary hypotheses and determine the most likely molecular clock, demographic and diffusion models from a wide range of models given the data (Baele *et al*., 2012, 2013; Drummond and Bouckaert, 2014). Bayesian analyses with BEAST take as input, a multiple nucleotide sequence alignment of viral pathogens, their sampling dates, locations and a set of proper priors for all viral population parameters being estimated (Drummond and Rambaut, 2007; Baele *et al*., 2012, 2013), to reconstruct the most plausible and/or otherwise hidden spatio-temporal movement pathways underlying the observed geographical distributions for a given viral pathogen by calling on a probabilistic framework implemented through the phylogeographic methods (Drummond and Rambaut, 2007, 2009; HIV: Bedford *et al*., 2010). In addition, BEAST can also be used to estimate the dispersal rate, wave front velocity and directionality of epidemic spread (Biek *et al*., 2007; Lemey *et al*., 2010).

Evolutionary methods implemented using BEAST have since been used to reconstruct the spatio-temporal diffusion of both animal and plant viruses. Animal-infecting viruses investigated using these methods include the Ebola virus (EBOV; Azarian *et al*., 2015), Human immunodeficiency virus (HIV; Faria *et al*., 2012), Dengue virus (Allicock *et al*., 2012; Nunes *et al*., 2014; Morato *et al*., 2015), Rabies (Kuzmina *et al*., 2013), and West Nile virus (WNV; Zehender *et al*., 2011), plant viruses whose evolutionary dynamics have been studied using BEAST include the *Tomato yellow leaf curl virus* (TYLCV: Lefeuvre *et al*., 2010, 2011), *Cassava mosaic virus* (CMG: De Bruyn *et al*., 2012), and *Maize streak virus* (MSV: Monjane *et al*., 2011). In fact, phylogeographic analyses of 353 full genome MSV-A isolates has recently revealed that this strain emerged in southern Africa around 1863, is trans-continentally dispersing at an average rate of 32.5 km/year across Africa, and has diversified into 24 recombinant lineages that currently circulate around the continent, most of which may have emerged within the past 40 years in southern and/or east-central African diversification hotspots (Monjane *et al*., 2011).

1.9.1 Identifying potential determinants of MSV-A diffusion

The generalized linear model (GLM) has recently been added as an extension to the methods implemented in BEAST (Faria *et al*., 2013; Lemey *et al*., 2012; Lemey *et al*., 2014). This model parameterizes the logarithm of the instantaneous DNA rate matrix as the logarithm of a combination of a set of predictor variables using the Bayesian stochastic search variable selection (BSSVS) probabilistic framework in BEAST (Faria *et al*., 2013; Lemey *et al*., 2012; Lemey *et al*., 2014). Given a predictor or set of predictor variables, the GLM approach calculates both the posterior inclusion probability (PIB: a Bayes factor support for each predictor) and a conditional effect size (cES) representing the degree to which a predictor is either included or excluded in the model, and using the PIB and cES statistics for each predictor variable considered, determinants that are possibly driving the viral dispersal process can be inferred. This is important because several such interacting predictive variables may drive the viral dispersal process, and amongst those widely investigated using the GLM approach are variables that can be

classified into ecological, climatic, genetic, sociopolitical, economic, and physical factors (Lemey *et al*., 2012, 2014; Faria *et al*., 2013; Magee *et al*., 2015; Nunes *et al*., 2013).

By comparing variations observed in these predictive variables under their different classes to specific instances and/or frequencies of disease outbreak occurrence, it is possible to: (i) identify determinants of viral dispersal and/or conditions prevailing in sources (locations where there is a demographic surplus of the virus, and its rapid multiplication and persistence are more likely) and sinks (locations where there is a demographic deficit of the virus and its extinction is more likely), and (ii) quantify the contribution that each predictive variable makes to the viral dispersal process (Lemey, *et al*., 2012, 2014). This helps to elucidate the different distribution and dispersal pathways for distinct variants of epidemiologically relevant pathogens such as those of the MSV-A strain. In light of this, the GLM approach has been used to investigate the determinants of spatio-temporal diffusion for animal viral infections that include, Human influenza H3N2 (Lemey *et al*., 2012; 2014) and H5N1 (Magee *et al*., 2015), Dengue viral serotypes 1-3 in Brazil (Nunes *et al*., 2013), and Bat rabies virus in North America (Faria *et al*., 2013). In this study, the GLM approach will be used to estimate the potential determinants of plant viral dispersal using MSV-A as a model organism.

Several statistical advantages are associated with the use of the GLM approach in testing spatial hypotheses (Lemey *et al*., 2014). Firstly, there is strong evidence that Bayesian measures of model fit (e.g. harmonic mean estimate of the marginal likelihood) which can be applied to models with among-location movement rates fixed to a particular predictor, perform poorly (Baele *et al*., 2012, 2013; Baele and Lemey, 2013). Secondly, Bayesian measures of model fit provide only a relative ranking of different models and unlike the GLM approach, do not identify which of the top ranked predictors needs to be jointly considered as explanatory variables. Finally, the GLM approach is advantageous over alternative approaches because it provides a measure of support for each predictor by estimating the associated coefficients (β) (Lemey *et al*., 2014).

1.9.2 Accounting for sources of confounding in Bayesian Inference

Although there are a number of distinct advantages to the Bayesian inference approach using probabilistic models it has to be borne in mind that there are a number of factors that may confound or bias the results obtained using such methods. These factors include the presence of evidence of recombination within molecular sequence datasets (Schierup and Hein, 2000a; Schierup and Hein, 2000b), unevenness in sample sizes and sampling times (Duchene *et al*., 2015), and the inability to account for un-sampled areas as possible sources or sinks in the viral dispersal process (Stack *et al*., 2010; Pulliam, 1988; Pulliam *et al*., 1991).

The presence of evidence of recombination at the interspecies, inter-strain and/or intra-strain levels in molecular sequence datasets is a result of lateral transfer of nucleotides between viruses (Ubeda and Wilkins, 2011). Recombination in viral datasets can invalidate phylogeny reconstruction, estimates of selection pressure at codon positions, selection of evolutionary models, and inferences from model parameters (Schierup and Hein, 200a, 2000b). It is therefore important to detect and remove recombinant sequence tracts or identify recombination breakpoint positions and focus phylogeographic analyses exclusively on those genome regions that are free of recombination (Martin *et al*., 2015).

Similarly, opportunistic sampling schemes often lead to datasets with uneven sample sizes either in space or time or both. To ensure that the inference of the location of the origin of the most recent common ancestor of a viral population is not systematically biased towards locations with larger sample sizes from amongst the locations under consideration (Lemey *et al*., 2010), a tip-swap null model can be used to account for uneven sample sizes amongst the sampling locations (Frost *et al*., 2015; Stack *et al*., 2010). As the name suggests, this involves randomly shuffling taxon labels across the tips of the phylogeny followed by calculation of the root state probability for each of the sampling locations, which is the probability of each sampling location being the origin of the most recent common ancestor of the viral population (Frost *et al*., 2015; Stack *et al*., 2010). Comparison of these probabilities to those obtained without the tip-swap null

model for each of the sampling locations, gives a clearer picture of which location is the most probable origin of the MRCA and more importantly, whether the inference of this estimated population parameter is likely biased towards locations with larger sample sizes.

Unevenness in sampling times can also introduce bias that results in over-estimation or under-estimation of the nucleotide substitution rate (Duchene *et al*., 2015), thereby providing misleading inferences on how fast a virus population is evolving. Such bias is usually accounted for by estimating the temporal signal within the dataset using Path-O-Gen (Rambaut *et al*., 2013) or estimating the nucleotide substitution rate for several randomly generated smaller sample size datasets through the date-randomization test (Ramsden *et al*., 2008). Path-O-Gen estimates the correlation between sampling times and the genetic distance of the samples in the data, and this is used as a proxy of the strength of the temporal signal in the data, where the stronger the correlation is, the stronger the temporal signal, and the lower the chances of unevenness in sampling times possibly influencing estimates of the nucleotide substitution rate. However, the date randomization test, which is computationally intensive but more efficient compared to Path-O-Gen, involves randomly reassigning the sampling times of the sequences, which effectively breaks the association between substitutions and time (Duchene *et al*., 2015), and generates an expectation of substitution rate estimates in the absence of temporal signal in the data.

In this investigation, I analyzed 689 MSV-A full genomes sampled over 32 years (1979-2011) from 20 locations across Africa and the adjacent Indian Ocean Islands. This included 286 novel full genomes, 111 of which were sampled in six countries [Anjouan (n=1), Moheli (n=2), Ethiopia (n=1), Mauritius (n=1), Madagascar (n=53), and Ghana (n=53; Oppong *et al*., 2015)] from which no data had been previously available (Monjane *et al*. 2011). My approach involved first an estimation of the genetic diversity of MSV-A isolates, and subsequent detection, dating and characterization of inter-and-intra strain recombination events across the genome and sampling regions, estimation of the best-fit

nucleotide substitution and evolutionary models and mapping of the selection pressures acting at codon positions in the MSV-A genome.

This produced a recombination-free MSV-A dataset that I then used to reconstruct the spatio-temporal diffusion history employing the discrete phylogeographic model implemented in BEAST (Drummond and Rambaut, 2007; Lemey *et al*., 2009). Using this approach it was possible to estimate the MSV-A nucleotide substitution rate, to infer the location where and dates when the most recent common ancestor (MRCA) of MSV-A samples existed and to identify statistically supported epidemiological linkages between countries, and elucidate the migration pathways and directions that MSV-A has used to attain its current distribution. Lastly, using the generalized linear model (GLM) approach implemented in BEAST, I characterized the sampling countries into source and sink locations and estimated the probable contributions of 27 predictor variables in the dispersal of MSV-A across Africa and Islands in the Indian Ocean.

2.0 MATERIALS AND METHODS

2.1 Sample Collection

Maize streak virus (MSV) samples collected over a period of 32 years (1979-2011) from a total of 20 countries distributed across Africa and the adjacent Indian Ocean Islands were used in this study (Table 2.0). The MSV dataset used comprised full genome sequences from 403 MSV-A (Appendix 20 - Table 3.7) and 182 MSV (MSV-B to MSV-K) held in publicly available databases (NCBI), along with 286 unpublished MSV-A full genome sequences collected by our research collaborators based at the University of Cape Town's Computational Biology Unit (CBIO) and Molecular and Cell Biology (MCB) Department's; the John Innes Centre (JIC) in the United Kingdom; the Centre for International Agronomic Research Development (CIRAD) in France, and the Scottish Crop Research Institute (SCRI) in Scotland. A total of 58 of the 286 unpublished full genome MSV-A samples came from Anjouan (n=1), Ethiopia (n=1), Madagascar (n=53), Mauritius (n=1), and Moheli (n=2), whereas most of the new sequences were sampled from Kenya (n = 175). Sampled isolate sequences were then divided into nine separate MSV datasets (A - H) for analyses, from which different estimates and inferences were drawn (Table 2.1). All phylogeographic, evolutionary, and epidemiologic population parameter estimates and inferences concluded on in this study were therefore primarily based on analyses carried out on datasets containing the more epidemiologically relevant MSV-A strain variants.

Table 2.0 Sample size and collection date per location for dataset A

Location	Samples	Collection date
Anjouan	1	2009
Benin	1	1999
Burkina Faso	5	2008
Cameroon	10	1998-2008
Central African Republic	33	2008
Chad	2	1987
Ethiopia	1	2010
Ghana	53	2010
Kenya	199	1983-2011
Lesotho	3	2005
Madagascar	53	2009-2010
Mauritius	1	Unknown
Moheli	2	2009
Mozambique	38	2006-2007
Nigeria	37	1983-2011
Reunion	12	1986-1997
South Africa	129	1979-2010
Uganda	68	2005
Zambia	17	2008
Zimbabwe	24	1987-2010

Table 2.1 Datasets analyzed in the study

Dataset	Description	Sequences	Analyses
A	Full Genome non-recombination free MSV-A (Appendix 21 - Table 3.8)	689	Genetic Diversity
B_1	Full Genome MSV sample subtypes A to K	871	Inter-strain recombination
B_2	Full Genome inter-strain recombination free MSV-A	689	Intra-strain recombination
C	Full Genome inter-and intra-strain recombination free MSV-A	689	Nucleotide substitution model selection
D	Movement Protein (MP)	13	Positive Selection
E	Coat Protein (CP)	37	Positive Selection
F	Replication Associated Protein (Rep)	63	Positive Selection
G	Replication Associated Protein A (Rep A)	20	Positive Selection
H	Full genome MSV-A samples with country centroid latitude and longitude coordinates specified	668	-Evolutionary model selection -Discrete Phylogeography -Predictor

2.2 Genome Sequencing

Sequencing of all novel MSV-A full genome samples used in this study was accomplished by my research collaborators following the methods proposed by Owor *et al*. (2007) and Shepherd *et al*. (2008).

2.3 Multiple Sequence Alignment and Editing

All alignments for the eight datasets (A-H) analyzed in this study were done using MUSCLE (Edgar, 2004) implemented in MEGA version 6.06 (Tamura *et al*., 2013), and subsequently manually edited using the Improved Alignment Editor (IMPALE: http://web.cbio.uct.ac.za/~arjun/).

2.4 Genetic Diversity Analyses

The levels of genetic diversity in the MSV-A dataset A were estimated using both MEGA version 6.06 (Tamura *et al*., 2013) and version 1.2 of the Species Demarcation Tool (SDT: Muhire *et al*., 2014; http://www.web.cbio.uct.ac.za/SDT).

2.5 Genome-wide Recombination Analyses Using RDP4

To account for inter-and-intra strain recombination, I used the recombination detection program (RDP4: Martin *et al*., 2015) with default setting and methods implemented in the program (RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiScan, PhylPro, LARD and 3Seq), to analyze dataset B_1 first, then B_2, producing a mostly inter-and-intra strain recombination free dataset C from which I created dataset H used in the discrete phylogeographic analyses. Only potential recombination events detected by two or more of these methods and phylogenetic evidence of recombination were considered as robust evidence of recombination. I also used the Bonferroni correction to minimize type I and type II errors and the severity of correction was minimized by only searching for recombination signals in a single sequence within groups of three sequences sharing > 99.3% sequence identity in datasets B_1 and B_2 respectively.

2.6 Nucleotide Substitution Model Selection

I used jModeltest version 2.14 (Darriba *et al*., 2012) to estimate the best-fit nucleotide substitution model for dataset C, from a pool of over 88 substitution models that can be evaluated using the program.

2.7 Recombination Analyses of Coding Region Alignments Using GARD

Prior to natural selection analyses, I partitioned dataset A into four MSV-A gene encoding region alignment datasets (D, E, F, and G), and used the genetic algorithm for recombination detection (GARD: Kosakovsky Pond *et al*., 2006) implemented on the

DATAMONKEY web server (Delport *et al*., 2010: http://www.datamonkey.org) to detect for evidence of recombination in those datasets. This involved the use of an HKY85 nucleotide substitution model with four rate categories in a beta-gamma distribution to search for and identify putative breakpoint recombination delimiting regions that had distinct phylogenies for datasets D to G. Two factors motivated my use of the HKY85 model, the first of which is that the DATAMONKEY web server does not implement the GTR+G4+I model estimated earlier on for dataset C using jModeltest and secondly because the HKY85 model has recently been estimated as a good fit for a 353 MSV-A isolate dataset used in previous phylogeographic analyses by Monjane *et al*. (2011). Potential breakpoints were identified by improvement of the small-sample corrected Akaike information criterion (AICc) for phylogenetic trees constructed of individual recombinant fragments (Akaike, 1974). Based on the outcome of the GARD analyses, a level of statistical support was assigned and expressed as a breakpoint placement score (Kosakovsky Pond *et al*., 2006a, 2006b). The significance of GARD analyses breakpoints was assessed using the KH test in the "Hypothesis testing using phylogenies" (HyPhy) package (Kishino-Hasegawa, 1989; Kosakovsky Pond *et al*., 2005, 2005).

2.8 Natural Selection Analyses

Synonymous substitution rates at codon positions within MSV-A coding region datasets (D, E, F, G) were estimated using the "Partitioning approach for robust inference of selection" (PARRIS: Scheffler *et al*., 2006) and the fast-unbiased Bayes approximation (FUBAR: Murrell *et al*., 2013) maximum likelihood phylogenetic-based selection characterization methods both implemented on the DATAMONKEY web server (http://www.datamonkey.org: Delport *et al*., 2010). The PARRIS method accounts for site-to-site variation in synonymous substitution rate for each partition, which can occur as an artifact of recombination thereby reducing false positives (Scheffler *et al*., 2006), whereas FUBAR utilizes a hierarchical Bayes approach that allows a flexible prior specification with no parametric constraints on the prior shape (Murrell *et al*., 2013). Selection analyses results were considered significant at the 95% level ($p<0.05$). Thereafter, codon positions detected as evolving neutrally, or under either the influence

of positive (diversifying) or negative (purifying) selection were identified using version 1.0 of the program Selection Map (Muhire *et al*., 2014), which also takes a comma separated (csv) output file from FUBAR analyses as input and generates a coding region map plot of the selection pressure and its strength at all individual codon positions for each gene encoding region dataset (Muhire *et al*., 2014).

2.9 Bayesian Evolutionary Analyses

2.9.1Evolutionary Model Selection

Identification of the most appropriate molecular clock and demographic models for the data is crucial for making accurate divergence time and demographic change inferences from viral phylogenies (Ho and Duchene, 2014). To do this, I first calculated the harmonic mean estimate of the marginal likelihood (HME: Newton and Raftery, 2007) in TRACER (Rambaut *et al*., 2009), first for the strict and relaxed (uncorrelated) molecular clock models, and secondly for the demographic models [simple parametric (constant size: Kingman, 1982a; exponential growth: Griffith and Tavares, 1994) and complex non-parametric (Bayesian Skyride and Bayesian Skygrid: Gill *et al*., 2013). Whereas, the strict molecular clock assumes a uniform nucleotide substitution rate across the branches of the phylogeny, the relaxed molecular clocks, allow each branch to have it own rate in the phylogeny (Drummond *et al*., 2006). I used the HME approach only because the alternative and more accurate Path-Sampling (PS: Ogata, 1989; Gelman and Meng, 1998; Lartillot and Philippe, 2006) and Stepping Stone (SS) methods (Xie *et al*., 2011), proved very computationally intensive and therefore time-consuming, such that by the time of submitting this thesis, these analyses were still incomplete.

2.9.2 Phylogeographic Analyses

Discrete models of spatio-temporal diffusion are more appropriate for estimating viral movements over very long distances because they do not assume that the diffusion process follows either a restrictive homogeneous Brownian motion, and/or any of the

relaxed random walk distributions (Cauchy, Gamma, or Lognormal) (Lemey *et al*., 2009, 2010). Discrete phylogeography models also yield a Bayes factor (BF) statistical support for the best-supported epidemiological linkages between sampling locations (Lemey *et al*., 2009). I used these models on a recombination-free 668 MSV-A full genome sample dataset (H), to attain estimates of the nucleotide substitution rate, the location where and time when the most recent common ancestor of the MSV-A population may have existed. Dataset H was created by excluding 21 MSV-A samples from dataset A that either did not contain collection dates, country of sampling origin and/or caused misalignment of the dataset. One such example is the single Mauritian sample sequence, whose exclusion from dataset A, resulted in my phylogeographic analyses focusing exclusively on 19 instead of 20 MSV-A sampling locations.

A minimum of five and a maximum of twelve replicate BEAST runs were set up where each had a Markov chain length ranging between 1.0×10^8 and 9.0×10^8 steps.

To ensure ample mixing of the Markov chain and parameter sampling before MCMC chains converged on a stationery posterior distribution of trees, I ran all analyses up to a point when all effective sample sizes (ESS) of all relevant model parameters were above 200, which is recommended for analyses results intended for publication (Drummond and Rambaut, 2007). I then used version 1.8.1 of Log-Combiner, a BEAST embedded package, to combine BEAST log and tree output files when similar results from independent replicate runs of the Markov chain were obtained for dataset H. Thereafter, I analyzed the resulting BEAST log traces using TRACER version 1.6 (http://tree.bio.ed.ac.uk/software/tracer/; Rambaut *et al*., 2009), and proceeded to annotate the Maximum Clade Credibility (MCC) tree using version 1.8.1 of Tree Annotator. The MCC tree is the tree with the highest accumulated posterior support in the posterior distribution of trees produced using the phylogeographic methods implemented in BEAST (Drummond and Rambaut, 2007). I then used FigTree version 1.4.2 to view the annotated MCC tree and the evolutionary relationship amongst publicly available and novel MSV-A samples (http://tree.bio.ed.ac.uk/software/figtree/; Rambaut *et al*., 2009).

Using version 1.06 of the "Spatial phylogeographic reconstruction of evolutionary dynamics: SPREAD" software (http://www.phylogeography.org/SPREAD), and the BEAST log, MCC tree and sampling location geographical coordinate files as input, I calculated Bayes factors (BF) for statistically supported inferred MSV-A movement pathways and epidemiological linkages. I considered evidence of MSV-A viral movements between different locations yielding a BF of <5 as not well supported; a BF of >5 to reflect substantial support, and I took BFs of >10 and >100 to indicate strong and decisive statistical support respectively (Kass and Raftery, 1995; Suchard *et al*., 2001). Using SPREAD, I proceeded to create a karyotype markup language (KML) file from the MCC tree file for dataset H and then projected this file through time onto Google Earth (https://earth.google.com/) to visualize the phylogeographic spread of MSV-A among discrete geographical locations by displaying all the transition rates with a non-zero expectancy resulting in statistically significant BFs (only those larger than five). The directionality of Bayes factor (BF) supported transmission movements for established epidemiological linkages was inferred using the Bayesian stochastic search variable selection (BSSVS) approach under an asymmetric diffusion model (Lemey *et al*., 2009).

2.9.3 Accounting for unevenness in sample sizes and sampling times

Unevenness in sample sizes and in sampling times can bias first, the inference of the origin of the most recent common ancestor (MRCA), and secondly estimates of the rate at which nucleotide substitutions are occurring in reconstructed viral phylogenies. To account for bias arising from unevenness in sample sizes in space, a tip-swap null model was used as described in Frost *et al*. (2015) and Stack *et al*. (2010), and to account for unevenness in sampling times, I estimated the correlation between sampling times and the genetic distance amongst samples in dataset H using version 1.4 of Path-O-Gen (Rambaut *et al*., 2009), and used this as a proxy of the strength of the temporal signal in the data. I interpreted a strong positive correlation between sampling times and the genetic distance amongst samples, as reflective of strong temporal signal within dataset H and therefore a lower chance of nucleotide substitution rate estimates obtained using

BEAST in the discrete phylogeographic analyses being biased by unevenness in sampling times (Rambaut, *et al*., 2009; Duchene *et al*., 2014).

2.9.4 Source-sink dynamics and determinants of MSV-A dispersal

Using the generalized linear model (GLM) implemented in BEAST, I investigated a total of 27 (Appendix 6 and 7) different predictor variables possibly contributing to MSV-A dispersal across Africa and the adjacent Indian Ocean Islands. These were respectively classified into ecological, genetic, climatic, economic, sociopolitical, and physical factors for easier collective interpretation. The GLM quantifies the contribution or effect size of potential predictor variables by estimating the GLM coefficient and the frequency at which each predictor is included in the model based on an inclusion probability, which represents the support for the predictor (Lemey *et al*., 2014). To estimate the latter, I first calculated Markov jumps for the locations under consideration and included this information into the GLM BEAST input file, for which I set an MCMC chain length of 7.0×10^8. While replicate runs of this input file were run, no output log files were combined using Log-Combiner. I inspected the GLM log file statistics using TRACER (Rambaut *et al*., 2009) from which I calculated the BF support for the different individual predictive variables (Kass and Raftery, 1995), and extrapolated the net Markov jumps (expected number of transitions) and Markov rewards (waiting times in a particular location) computed using the continuous-time Markov chain (CTMC) with the GLM in BEAST (Faria *et al*., 2011; Lemey *et al*., 2014). This allowed for inference of the most important determinants to the spread of MSV-A, as well as locations contributing the most to the persistence and dispersal of MSV-A across Africa and the adjacent Indian Ocean Islands. A flow chart schematic representation of the methods and analyses used in the study are shown in Figure 2.0 below.

The 20 MSV-A Sampling location country codes:

AN, BJ, BF, CM, CF, ET, GH, KE, LS, MD, MH, MU, MZ, NG, RE, TD, UG, ZA, ZM and ZW

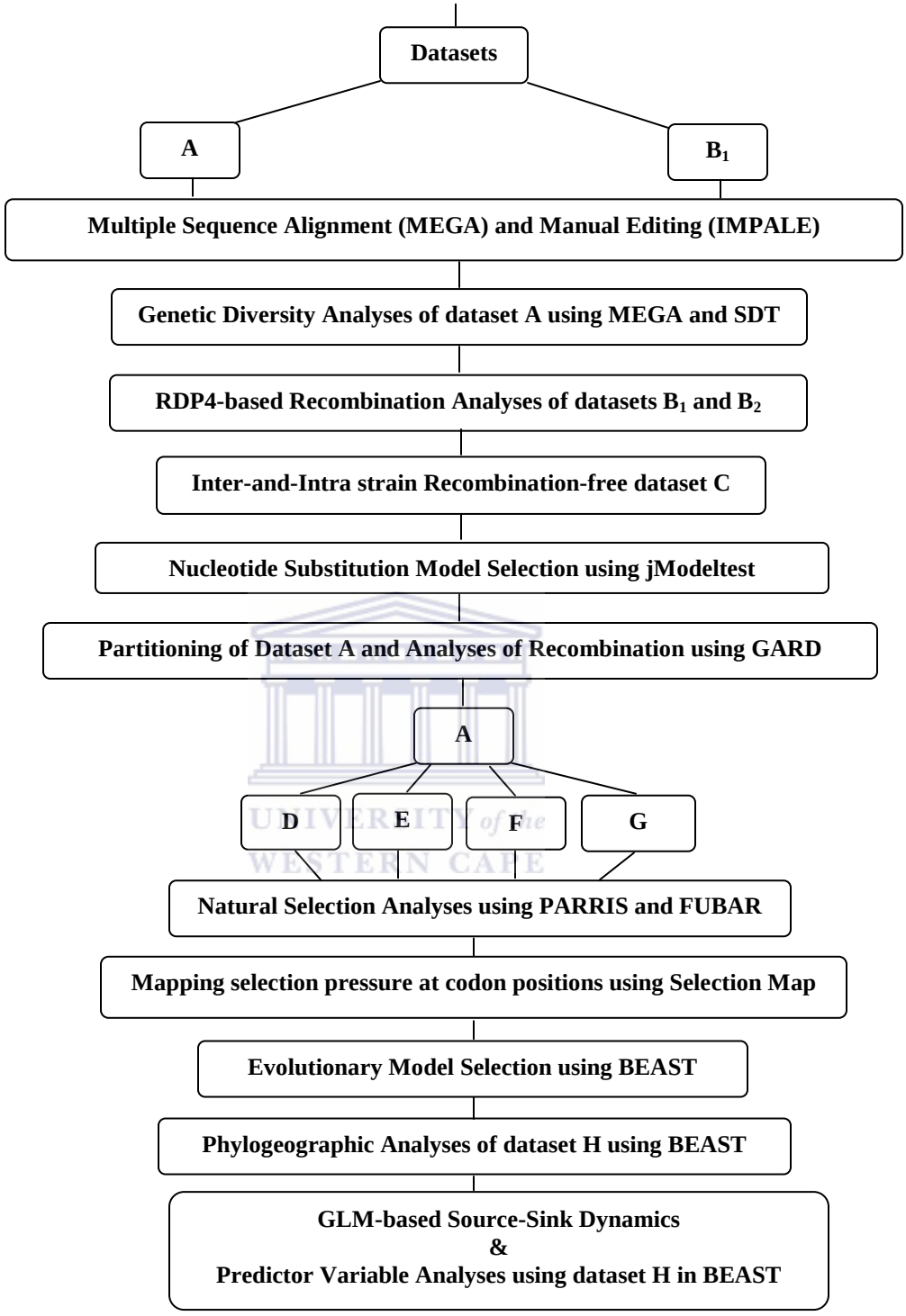


Figure 2.0 A flow chart schematic representation of the datasets and data analyses methods used in the study. AN = Anjouan; BF = Burkina Faso; BJ = Benin; CF = Central African Republic; CM = Cameroon; ET = Ethiopia; GH = Ghana; KE = Kenya; LS = Lesotho; MD = Madagascar; MH = Moheli; MU = Mauritius; MZ = Mozambique; NG = Nigeria; TD = Chad; UG = Uganda; ZA = South Africa; ZM = Zambia; and ZW = Zimbabwe.

3.0 RESULTS AND DISCUSSION

3.1 Genetic Diversity Analyses

An overall mean pairwise genetic distance of 0.017 and ≥ 98% nucleotide sequence identity were estimated for the 689 MSV-A dataset A using MEGA version 6.06 (Tamura *et al*., 2013) and SDT (Muhire *et al*., 2014) respectively. These estimates show that isolates used in the study were very closely related and, in accordance with the mastrevirus strain demarcation threshold suggested by Muhire *et al*. (2014) (indicating that all isolates sharing greater than 94% nucleotide sequence identity should be considered members of the same strain), all of the analyzed MSV isolates belong to the strain, MSV-A.

3.2 Genome-wide recombination and distribution of breakpoints

Using methods implemented in RDP4 (Martin *et al*., 2015), I first identified 33 well-supported inter-strain recombination events (Appendix 9 - Table 3.1a) and removed the respective recombination-derived sequence fragments from dataset B_1 which resulted in creation of an intermediate "inter-strain recombinant free" MSV-A dataset (B_2) containing 689 isolates. Further RDP4 based recombination analyses of dataset B_2 led to the detection of 34 well-supported intra-strain recombination events (Appendix 10 - Table 3.1b), and the creation of the mostly inter-and-intra-strain recombination-free dataset C, which I used to make dataset H that was analysed in BEAST by the discrete phylogeographic, and subsequent predictor and source-sink dynamics analyses. While 19 of the intra-strain recombination events (1, 2, 3, 4, 6, 10, 14, 15, 16, 17, 19, 20, 22, 23, 24, 25, 29, 31, 33) detected in this study have been described in previous MSV-A related studies (Owor *et al*., 2007; Varsani *et al*., 2008; Monjane *et al*., 2011), all of the inter-strain plus 15 of the of the 34 intra-strain recombination events (5, 7, 8, 9, 11, 12, 13, 18, 21, 26, 27, 28, 30, 32, and 34) were detected here for the first time. Characterization of the distribution of viruses carrying intra-strain recombination events (Figure 3.1) revealed that, while those carrying

events 20, 25 and 33, are widely distributed in East Africa, those carrying events 16, 21 and 23, occur more frequently in West and Central Africa. However, the viruses carrying most of the events (1, 4, 8, 10, 14, 17, 19, 22, 24, 25, 26, 28, and 29) are more spread out across East, West, Central and Southern Africa and the adjacent Islands in the Indian Ocean.

Using the highest posterior density intervals (HPD - with a lower and upper bound representing the interval containing 95% of the sampled values) on the height of the nine nodes marked by black dots in the MSV-A reconstructed phylogeny (Figure 3.1), and evidence of recombination based on RDP4 analyses (Table 3.1b), I estimated dates on which intra-strain recombination events 1, 4, 8, 10, 14, 17, 19, 20, 21, 22, 26, 28 and 29, may have occurred and summarized these in Table 3.1c below.

Table 3.1c Dates on which intra-strain MSV-A recombination events most likely occurred

Events	MSV-A subtype strain	Estimated date of occurrence
'1, '4 and '10	A_4	(95% HPD 1972 - 1983)
'8	A_1	(95% HPD 1987 - 2004)
'14, '20 and '22	A_1	(95% HPD 1974 - 1985)
'17 and '29	A_1	(95% HPD 1991 - 2001)
'19 and '28	A_1, A_2, A_3 and A_6	(95% HPD 1942 - 1976)
'16	A_1	(95% HPD 2002 - 2006)
'21	A_1	(95% HPD 1989 - 1997)
'23	A_1	(95% HPD 1989 - 1997)
'26	A_4	(95% HPD 1959 - 1971)

As shown in Table 3.1c, most of the intra-strain MSV-A recombination events detected here with RDP4 have occurred within the last six decades. The oldest and most conserved (with respect to persistence time in the genome and frequency of occurrence within sampled isolates) recombinant sequence fragments belonged to events 19, 26 and 28 whereas, the most recent to events 8, 16, 17, 21, 23, and 29. Only four of the 15 unique intra-strain recombination events (8, 21, 26, and 28) appear to be well conserved in the MSV-A genome. It is also clear that West and Central African MSV-A samples detected with evidence of event 21, also share evidence of events 16 and 23. Although unlikely, highly conserved recombinant fragments detected in the intra-strain recombination analyses have perhaps played a

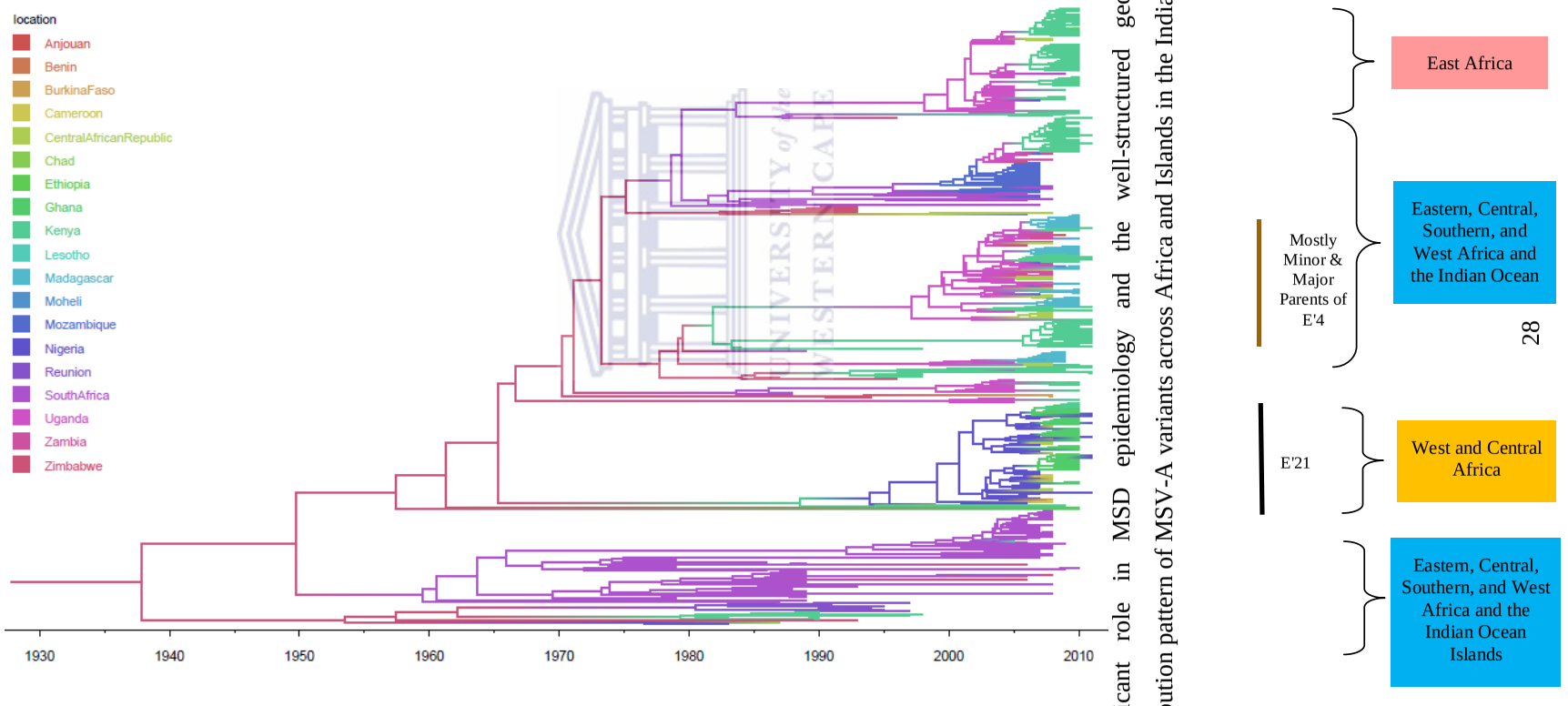


Figure 3.1 Maximum clade credibility tree showing the spatio-t
Africa and the adjacent Indian Ocean Island region. The nine noc
specific intra-strain recombination events occurred are marked by
MRCA for isolates from Moheli and Madagascar, plus those from c
migration event 18 (see section 3.6.3.3 and Figure 3.7).

significant role in MSD epidemiology and the well-structured geographical distribution pattern of MSV-A variants across Africa and Islands in the Indian Ocean.

3.3 The best-fit Nucleotide Substitution Model

Using version 2.14 of jModeltest (Darriba *et al*., 2012), I identified the best-fit DNA nucleotide substitution model for the MSV-A dataset H as the Generalized Time Reversible parameter model with four gamma rate categories and a proportion of invariant sites (GTR+G4+I), which is consistent with previous studies (Monjane *et al*., 2011).

3.4 Recombination breakpoints within coding sequence alignments

Recombination analyses of the coding sequence alignments was performed using the GARD method implemented on the DATAMONKEY web server (Delport *et al*., 2010). This analysis revealed no evidence of recombination in datasets G (RepA) and D (MP). However, one breakpoint was detected in dataset E (Rep) at position 372 and another two breakpoints were also detected in dataset F (CP) at positions 174 and 434. The detection of evidence of recombination in the Rep and CP coding regions reflected topological incongruence according to the KH test report ($P = 0.01$, $P = 0.05$, and $P = 0.1$) but not significant enough to invalidate phylogeographic inferences because while GARD used just one method to converge on this result, RDP4 based analyses invoked a total of nine different well-supported methods to estimate the presence, characterization and distribution of recombination events, breakpoint positions, possible recombinants, minor and major parental sequences. Furthermore, given that datasets D, E, F, and G were created from dataset A which had recombinant sequences because it had not been subjected to the more robust RDP4 based recombination analyses, it is not surprising that evidence of recombination was detected in datasets E and F. Dataset A was only used to create datasets D to G so as to conform with the DATAMONKEY web server data submission requirements for detecting recombination using GARD and subsequent estimation of the force of selection acting on codon positions in molecular sequences. Amongst these requirements is a recombination-free input multiple sequence alignment dataset containing ≤ 300 samples for each coding region dataset.

3.5 Natural Selection Analyses and Mapping of Selection Pressure

3.5.1 Evaluating Evidence of Positive Selection Pressure

One way of determining the role that population process such as natural selection have in shaping the MSV-A population across time and space is by estimating the strength of selection pressure acting upon different codon positions within coding regions. This is because selection pressure acting at codon positions or nucleotide sites within specific codons, may promote increased translational efficiency and accuracy, codon usage bias in species with larger effective population sizes (Ingvarsson, 2008) and/or modify translational kinetics to produce correct protein folding (Yang and Nielsen, 2008). To estimate synonymous nucleotide substitution rates (substitutions not resulting in amino acid change) within each of the four MSV-A coding region datasets (D, E, F and G) at the least functionally constrained third codon positions (Bofkin and Goldman, 2007), I used the maximum likelihood phylogenetic-based selection characterization methods of PARRIS (Scheffler *et al*., 2006) and FUBAR (Murrell *et al*., 2013), that are available for implementation on the DATAMONKEY web server (Delport *et al*., 2010). These analyses did not reveal any evidence of positive selection for datasets D (MP) and G (Rep A), however statistically significant evidence for strong positive selection (at the 95% level; $p<0.05$) was detected in dataset F (Rep) by both PARRIS ($P = 0.000263857$; Table 3.2) and FUBAR methods at seven codon positions (147, 166, 195, 203,244, 260, and 267: at the 90% level; $p<0.9$).

Table 3.2 PARRIS-based Evidence of Strong Positive Selection Pressure

Dataset	P-value	Substitutions per site	LRT	Codons
D (MP)	0.999908	0.102525	0.000183075	101
E (CP)	0.137675	0.184714	3.96572	244
F (Rep)	0.000263857	0.407206	16.4802	272
G (RepA)	0.99881	0.184714	0.155185	153

Of the two methods used to estimate synonymous nucleotide substitution rates, only FUBAR detected significant evidence (at the 90% level; p>=0.9) of positive selection within dataset E (CP) at two codon positions (12 and 19). Maps (Appendix 13 to 16 - Figures 3.3a-d) showing the strength of the selection pressure acting at codon positions within datasets D, E, F and G were respectively generated using version 1.0 of the program Selection Map (Muhire *et al*., 2014; http://etal.cbio.uct.ac.za/~brejnev/ComputationalTools.html), for positions detected as evolving neutrally, or under the influence of either negative or positive selection.

3.5.2 Determinants of Positive Selection in MSV-A Coding Regions

3.5.2.1 The MSV-A Replication associated transcript (Rep)

The Rep transcript is a splice product of the C1 and C2 open reading frames (ORFs) (Pratt and Gordon, 2006; Preiss and Jeske, 2003). It initiates viral replication through a rolling circle mechanism that entails cleavage of the positive (virion sense) strand within the conserved nonanucleotide sequence (Appendix 1-Figure 1.1) and binding covalently to the 5' and 3' ends following one round of replication leading to generation of ssDNA MSV-A genomes (Pratt and Gordon, 2006; Preiss and Jeske, 2003).

In geminiviruses such as MSV-A, the Rep transcript consists of five known motifs that have been comprehensively described (Willment *et al*., 1999; Nash *et al*., 2011). These motifs are arranged in the MSV-A Rep transcript in the following sequence: I-II-III-Rb-IV (Willment *et al*., 1999). Rb is the geminivirus retinoblastoma protein binding motif (Willment *et al*., 1999). Motifs I, II, III, and IV are involved in rolling circle replication (RCR). Motif I is five amino acids residues long (FLTYP); motif II has six residues (HLHALL); motif III is four amino acids residues long (YI/TLK); the Rb motif (PSSPDLLCNESINDW) is 15 amino acid residues long and lastly, motif IV site A (SLYIVGPTRTGKSTWARSLGV) has 21 residues while site B of motif IV (IYNIVDDIPEKE) is 12 residues long (Illya and Koonin, 1992; Laufs *et al*., 1995a; Willment, 1999; Nash *et al*., 2011). Motifs I, II, III, Rb, and IV are known to occupy codon positions 18 through to 272 in the MSV-A Rep transcript. Specifically, motif I, whose function is unknown, spans from codons 18 to 22, and is more commonly

referred to as RC1. Motif II, which is involved in binding Mn^{2+} or Mg^{2+} metal cations and may consequently influence protein conformation and/or catalysis (Laufs *et al*., 1995a), starts from position 60 up to 65, and is often called RC2. Motif III, is required for phosphodiester bond cleavage for initiation of RCR (Orozco *et al*., 1998), linkage at the virion sense origin of replication (Laufs *et al*., 1995b), and in Rep-Rep, RepA-RepA and RepA-Rep interactions (Settlage *et al*., 1996; Hovarth *et al*., 1998). It lies in-between codons 100 and 103 respectively. A reference of these three motifs is publicly available and linked to the UniprotKB identifier, P03568. The Rb motif spans from codon positions 193 up to 207 in Rep transcript, and incidentally houses codon positions 195 and 203 (two serine (S) amino acid residues marked in red above), detected here as evolving under strong positive selection using PARRIS and FUBAR, indicating their role in inhibition of the plant cell cycle and transition from the G1 to S phases, in the plant retinoblastoma regulated pathway (RBR) (Xie *et al*., 1995; Willment *et al*., 1999; Gutierrez, 2000). Motif IV site A spans from codon positions 224 up to 244 and site B spans from positions 262 up to 273. Only one of the seven codon positions within site B of motif IV (position 267: coding for Aspartic acid (D) marked in red above) detected here using PARRIS and FUBAR as evolving under the influence of strong positive selection lie in this region of the Rep transcript. It is possible that these two codons may actively participate in the MSV-A rolling circle replication process.

Motif IV is a known NTP binding motif, having the characteristic P-Ioops found in proteins with kinase and DNA helicase activity (Hanson *et al.,* 1995; Gorbalenya and Koonin, 1989), and it is essential for continuance of the replication cycle *in vivo* (Desbiez *et al.,* 1995; Hanson *et al., 1995;* Heyraud-Nitschke *et al*., 1995; Thommes *et al.,* 1993). It also probably induces virion sense gene transcription (Hofer *et al*., 1992) and/or host genes during an infection (Palmer and Rybicki, 1998). Therefore, nucleotide substitutions in either of these codon positions as a result of recombination are likely to significantly influence the rate at which the virus replicates *in vivo*, and also the quantity of viral titer available for acquisition and dispersal by *Cicadulina* leafhoppers in the physical environment. Therefore strong positive selection in the highlighted Rep motifs probably confers a selective advantage to MSV-A variants that influences their persistence in different hosts and current geographical distribution across Africa and adjacent Indian Ocean Islands.

3.5.2.2 The MSV-A Coat Protein transcript

The MSV-A coat protein (CP) is a multi-functional component of the ssDNA geminivirus genome (Zhang *et al*., 2001). Its functional roles include viral encapsidation (Mullinueaux *et al*., 1988; Townsend *et al*., 1985); intra- and inter-plant virus transmission (Boulton *et al*., 1989; Lazarowitz *et al*., 1989; Liu *et al*., 1999; Woolston *et al*., 1989); determination of vector specificity (Briddon *et al*., 1990); protection of viral ssDNA during transmission by leafhopper vectors (Azzam *et al*., 1994) and/or mechanical inoculation (Frischmuth and Stanley, 1998). While it has been reported that DNA nucleotide sequences required for vector transmission of the virus are often located in the central part of the CP in MSV and other geminiviruses (Liu *et al*., 2001; Unseld *et al*., 2001), codon positions 12 and 19 which are conserved for amino acids Serine (S) and Threonine (T) (marked in red within motif I below), and which were detected here as evolving under the influence of strong positive selection using FUBAR, are in fact not centrally positioned within the CP transcript of MSV-A. Codons 12 and 19 are in fact components of the 24 amino acid long CP motif I (MSTSKRKRGDDSNWSKRVTKKKPS), a bipartite nuclear localization sequence that interacts with the movement protein and binds both single and double stranded DNA (Liu *et al*., 1997). The following publicly available UniprotKB identifiers: P06448, P03569, and P14986, are for motif I in the CP of MSV-N, MSV-K, and MSV-S respectively. Because both codons lie in motif I, they significantly influence the yield of viral ssDNA inside infected host tissues, and therefore nucleotide substitutions within this region of the transcript (motif I) through processes such as recombination, which will either have an additive or reductive effect on the severity of MSD, vector-specificity, MP-CP interactions, viral transmission rates, and vector and host range, depending on the geographical location where recombination occurs. Still, the possibility of the CP playing a significant role in the control of vector transmission of MSV-A exists, since the CP s the initial point of contact between the virus and the vector. Unlike in geminiviruses such as *Mung bean yellow mosaic India virus* (MBYMIV), where the CP transcript participates in rolling circle replication (RCR: Saunders *et al*., 1991) by down-regulating the replication initiation activity (nicking and closing function) of the Rep transcript

(Malik *et al*., 2005), my analyses did not identify motifs controlling such activity in the MSV-A CP.

3.6 Bayesian Evolutionary Analyses

3.6. 1 Evolutionary Model Selection

Using the posterior-sampling harmonic mean estimate (HME: Newton and Raftery, 2007) method for calculating the marginal likelihood, I identified the relaxed uncorrelated lognormal molecular clock, constant population size model (Tables 3.3a and 3.3b) as the best-fit model for analyzing the evolutionary dynamics of MSV-A dataset H. Evolutionary model selection based on the path sampling (PS: Ogata, 1989; Gelman and Meng, 1998; Lartillot and Philippe, 2006) and stepping stone methods (SS: Xie *et al*., 2011), that generate better estimates of the marginal likelihood, was computationally intensive to a point where results for those analyses could not be presented here in time for submission of this thesis, but instead will be ready in time for publication.

Table 3.3a HME log Bayes factor based Molecular clock model selection

Model	Marginal LnL	Standard Error (S.E)	Relaxed uncorrelated lognormal	Strict
***Relaxed uncorrelated lognormal**	-37294.033	+/- 0.053	-	928.191
Strict	-38222.224	+/- 0.084	-928.191	-

*best model - Relaxed Uncorrelated Molecular Clock (RC)

Table 3.3b HME log Bayes factor based Demographic model selection

Model	Marginal LnL	Standard Error (S.E)	BSkygrid	Expgrowth	GMRF-Skyride	ConPopSize
BSkygrid	-37560.547	+/- 0.053	-	-226.28	215.769	-266.514
Expgrowth	-37334.267	+/- 0.195	226.28	-	442.049	-40.234
GMRF-Skyride	-37776.317	+/- 3.429	-215.769	-442.049	-	-482.284

***ConPopSize**	-37294.033	+/- 0.06	266.514	40.234	482.284	-

***** best model - Constant Population Size (ConPopSize)

3.6.2 Phylogeographic Analyses

3.6.2.1 *Maize streak virus* nucleotide substitution rate estimation

In addition to identifying positively and negatively selected codon positions within the MSV-A full genome, estimating the rate at which nucleotides in coding and non-coding genomic regions are substituted can inform on how quickly individuals within populations are changing, or are likely to change, in response to shifts in selection pressure. To estimate the MSV-A nucleotide substitution rate for dataset H, I first accounted for unevenness in sampling times by estimating the correlation between the sampling times and the genetic distance amongst samples using Path-O-Gen version 1.4 (Rambaut *et al*., 2009), and found evidence of a very strong positive correlation (Appendix 23), which I used as a proxy for high temporal structure in the dataset, meaning Bayesian estimation of the nucleotide substitution rate was unlikely to be influenced by any unevenness in sampling times. In view of this, the estimated mean nucleotide substitution rate for dataset H obtained was 9.922×10^{-4} (95% HPD 8.54×10^{-4} to 1.1317×10^{-3}) substitutions per site per year, which is much higher than the estimates obtained from short term (<60 days: Shepherd *et al*., 2005, 2006; Walt *et al*., 2009) and long-term (between 1 - 6 years: Isnard *et al*., 1998; Harkins *et al*., 2009; van der Walt *et al*., 2008b) experiments which lie in the range of 2×10^{-4} up to 7×10^{-4} substitutions per site per year. As previously reported (Harkins *et al*., 2009; Duffy and Holmes, 2009; Lefeuvre *et al*., 2011), it appears that high nucleotide substitution rates may be conserved in geminiviruses such as MSV and the East African cassava mosaic viruses, which explains why my estimate of the time to the most recent common ancestor (tMRCA), as seen later on, is more recent and does not correspond precisely with the initial report of the first MSD outbreak in South Africa around the 1870s (Fuller, 1901).

3.6.2.2 Geographical dissemination and origin of MSV-A

Using the Bayesian reconstructed maximum clade credibility (MCC) tree (Figure 3.2a), my analysis shows that while MSV-A isolates used in this study displayed a high degree of geographical clustering, the geographical ranges of the different subtypes (MSV-A_1 to -A_4, and -A_6) remain identical to those reported previously (Martin *et al*., 2001; Peterschmitt *et al*., 1996; Monjane *et al*., 2011; Varsani *et al*., 2008), but all of the novel sequences sampled from Anjouan, Ethiopia, Madagascar, Moheli, and Kenya, were of the -A_1 subtype. Whereas the -A_1 subtype strain isolates had a continent-wide distribution, isolates belonging to the -A_2, -A_3, -A_4, and -A_6 subtype strains existed only in West Africa, East Africa, and Southern Africa, and the island of Reunion, respectively. Furthermore, the -A_1, -A_2, -A_3, and -A_6 isolates nested under the same clade in the reconstructed MCC tree (Figure 3.2a), appear to have diverged from their most recent common ancestor sometime between 1950 and 1955. The high (> 0.5) posterior state probability support for the tree branches between the split that led to the genesis of the -A_3 and the -A_6 lineages and the root node of the tree, suggests that the -A_3 and -A_6 subtypes may have diverged from their most recent common ancestor sometime between 1960 and 1965.

Using the discrete phylogeographic analyses with and without the tip-swap null model (Figure 3.2b), my analysis estimated Zimbabwe (posterior state probability = 0.5192) as the most likely location where the most recent common ancestor of the 668 MSV-A samples in dataset H may have occurred. This estimate is consistent with a previous study (Monjane *et al*., 2011) but with the full genome (FG) dataset that also identified Zimbabwe (posterior state probability = 0.298) as the most probable location of the MRCA. My analyses revealed significant statistical support for the first emergence of MSV-A in Zimbabwe (posterior probability: 0.52), followed by South Africa (posterior probability: 0.19), Uganda (posterior probability: 0.09), and Kenya (posterior probability: 0.08). All other locations had posterior probabilities less than 0.02.

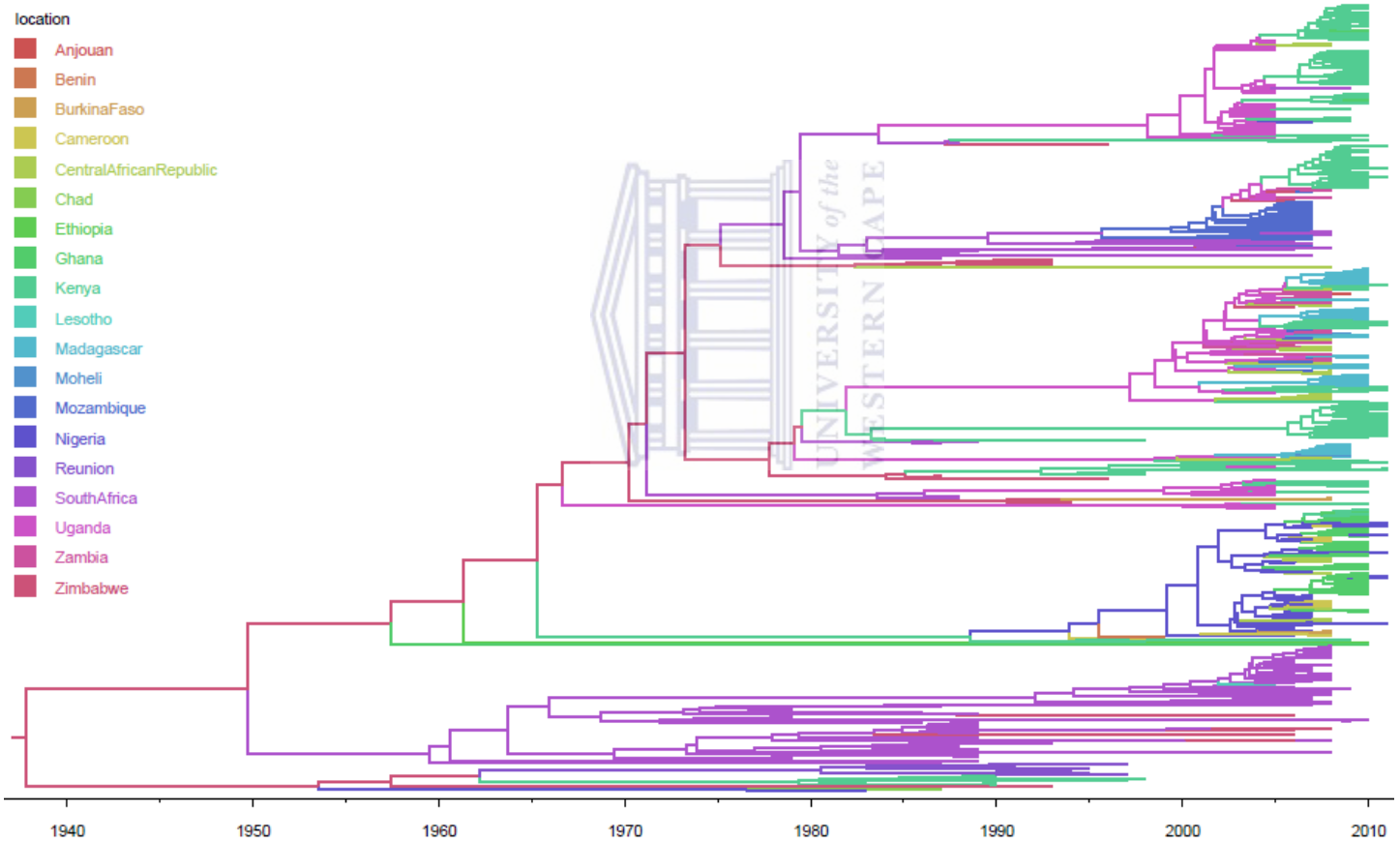


Figure 3.2a A Bayesian maximum clade credibili
A_2 = West Africa; A_3 = East Africa; A_4 = southern

Furthermore, while this result remains consistent with the most recent estimate placing the origin of the MSV-A MRCA within Southern Africa (Monjane *et al*., 2011), it also indicates that the inference of the location of the MRCA was mostly free from sampling bias which may arise due to uneven sampling sizes amongst different locations considered, and that MSV-A population parameter estimates obtained were in no way systematically influenced by such bias. In my tip-swap analyses (Figure 3.2b: Grey colored bar graph), that location was Kenya. Therefore, because Zimbabwe (and not Kenya) was selected as the location of the MRCA with the non-tip-swap analyses (Figure 3.2b: Multi-colored bar graph), this demonstrates that my analyses of dataset H was free from, and definitely not systematically influenced in any way, by sampling bias. Should a sample size bias have existed in dataset H, Kenya would have been designated the location of the MRCA in the non-tip-swap analyses.

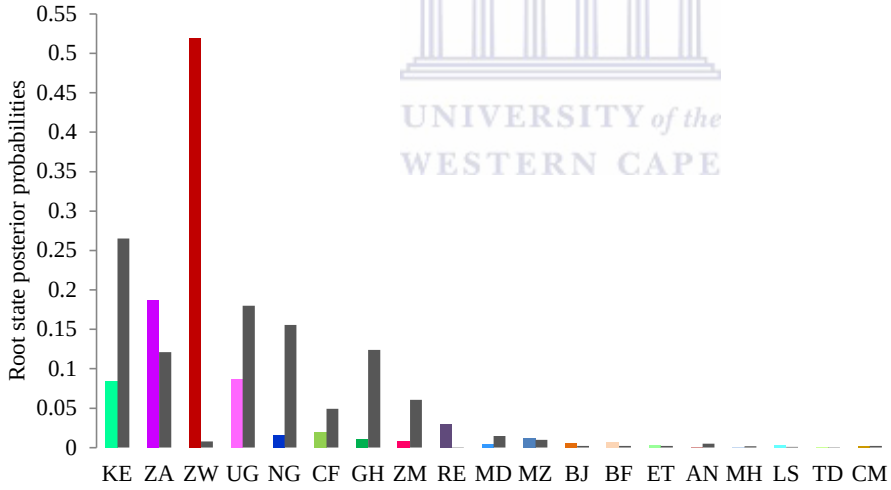


Figure 3.2b The most probable location of the most recent common ancestor of the MSV-A strain as determined using the discrete phylogeographic model with and without the tip-swap null model for dataset H. Grey color coded bar graph = location state probabilities under a tip-swap null model and Multi-color coded bar graph = location state probabilities without tip randomization. Sampling locations (states) are represented on the horizontal axis through a two-letter country code (e.g. ZW, Zimbabwe; ZA, South Africa).

3.6.2.3 The time to the most recent common ancestor (tMRCA)

The estimate of the time to the most recent common ancestor (tMRCA) obtained under the combined relaxed uncorrelated lognormal molecular clock and constant population size model was 73 years (95% HPD 54.9 -107.5). Translated into calendar years, this is equivalent to a date of 1938 (95% HPD 1904 - 1956), and is consistent with previous estimates (Monjane *et al*., 2011), which yielded a tMRCA of 1933 (95% HPD, 1867 to 1950) with a partial genome combined movement protein plus coat protein (MPCP) dataset but inconsistent with a tMRCA of 1863 (95% HPD, 1809 to 1935) inferred from the full genome (FG) dataset (Monjane *et al*., 2011).

3.6.3 Identifying the major MSV-A migration pathways

Since the emergence of the MSV-A subtype strain within southern Africa during the mid-1800s, several potentially complex patterns of MSV-A movement throughout the continent have been inferred (Monjane *et al*., 2011). Using the discrete phylogeographic model, analyses of dataset H identified a total of 34 MSV-A migration pathways, including 19 that were well supported (i.e. with a BF of > 5) (Figure 3.4a and 3.4b; Table 3.4 - Appendix 15). Incidentally, 15 of these 34 movements were concordant (Table 3.9 - Appendix 22) to those reported in previous analyses (Monjane *et al*., 2011) using both the full genome (FG) and combined movement and coat protein (MPCP) datasets. In the latter, Monjane *et al*. (2011) identified a collective total of 32 well supported movement pathways using both the discrete and continuous phylogeographic models, where eight of those were concordant between the FG and MPCP datasets. In this study, the discrete model also inferred a total of 19 Bayes factor supported epidemiological linkages for movements between locations paired with Uganda, Nigeria, Zimbabwe, South Africa, Ghana, Kenya, Mozambique and Madagascar, Zambia and the Central African Republic (Appendix 15), suggesting that these location pairings were influential in both the MSV-A dispersal process as well as the frequency of occurrence of MSD outbreaks across Africa and the adjacent Indian Ocean Islands.

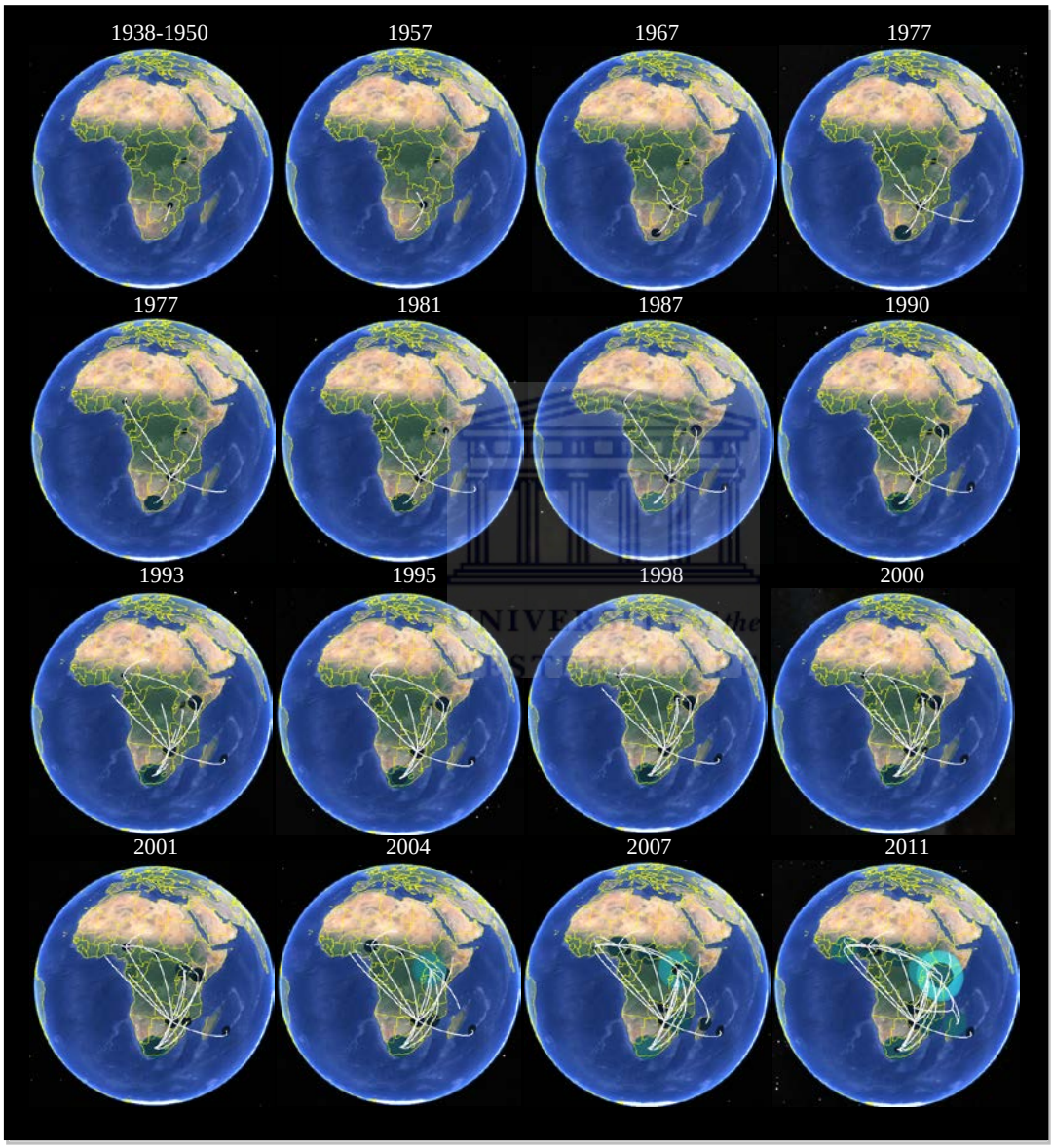


Figure 3.4a *Maize streak virus* A spread across southern Africa and into west, east, and central Africa as well as the adjacent Indian Ocean Islands through 34 migration events, 19 of which are Bayes factor (BF) supported between 1938 and 2011 as inferred using the discrete phylogeography model.

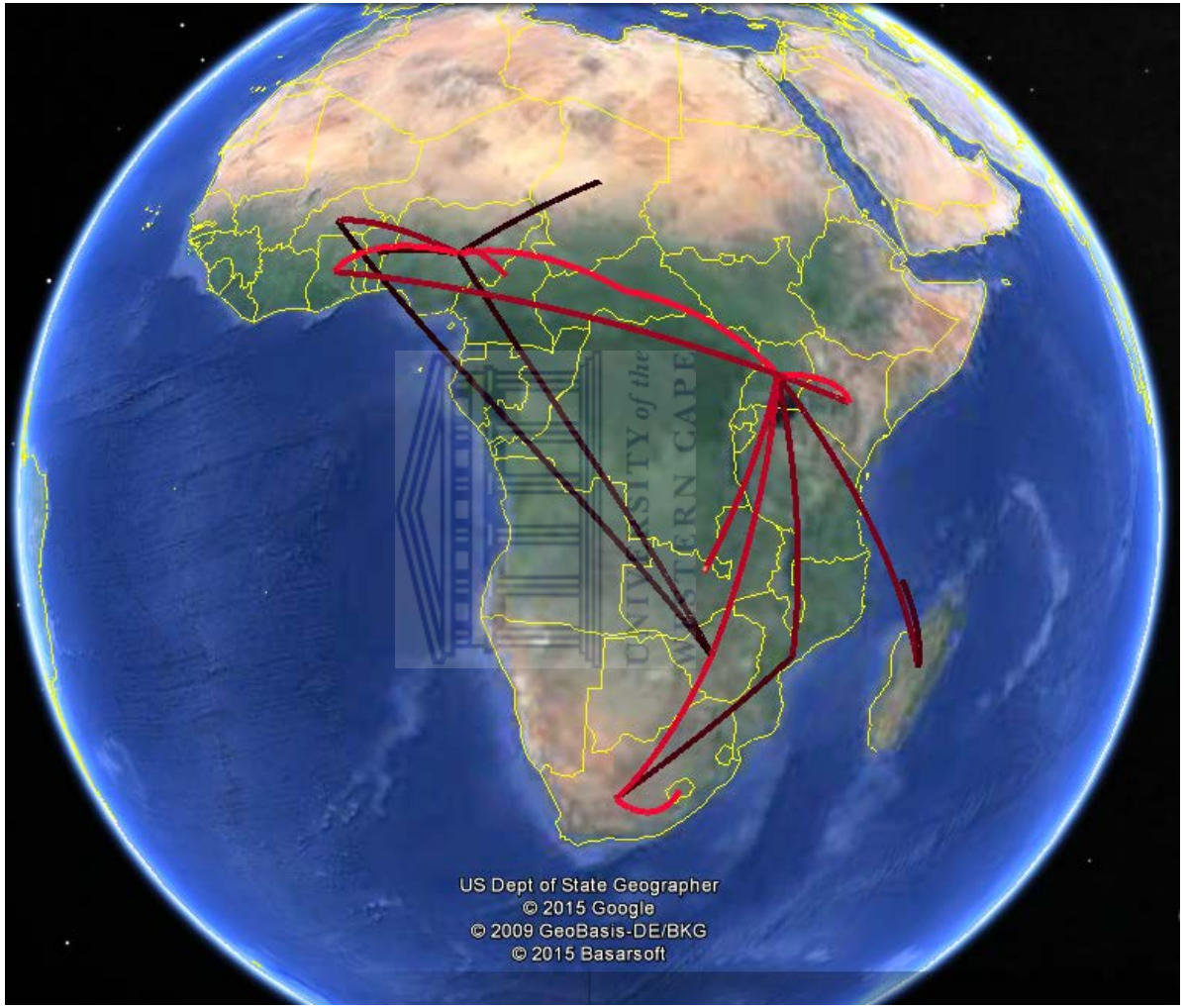


Figure 3.4b The 19 Bayes factor (BF) supported migration pathways through which MSV-A has attained its current geographical distribution across Africa and the adjacent Indian Ocean Islands between 1938 and 2011 as inferred using the discrete phylogeography model. Forty-two percent of the inferred movements were intra-regional whereas 58% were inter-regional. Light red line = significant BF support, Dark red line = strong BF support, and Black line = decisive BF support.

3.6.3.1 Movements from emergence up to 1990

In addition to inferring the emergence of MSV-A in Zimbabwe around 1938 (95% HPD 1904 - 1956), the discrete phylogeography model used here also inferred that 42% (8 events) of the 19 Bayes factor statistically supported MSV-A migration events, were intra-regional whereas 58% of them were inter-regional (11 events). Two sources of evidence corroborate the inferred emergence of MSV-A in Zimbabwe. Firstly, twelve years prior to this event, van der Merwe (1926) had reported the existence of the maize *jassid Balcutha mbila* Naude leafhopper species in the country. Storey (1932) later renamed the vector as *Cicadulina mbila*, which as is known, is widely distributed across Africa and the Indian Ocean islands and is also highly capable of transmitting MSV (CABI, 1986; Webb, 1987; Mylonas *et al*., 2014). Secondly, in 1936, Hopkins, a senior plant pathologist, documented a suspected streak disease of maize in an annual report in December of that year (Hokpins, 1936). This was perhaps the first credible report of MSD in Zimbabwe. But, prior to the Hopkins report and the inferred emergence of MSV-A in Zimbabwe around 1938, significant countrywide maize harvest failures and famine were reported in 1928, 1933, and then again in 1942, 1947 and 1960 (Iliffe, 1987), suggesting that MSV-A and leafhoppers existed in Zimbabwe; were expanding their geographical dispersal range, and significantly impeded maize production. Then as the model shows, MSV-A appears to have localized within Zimbabwe, for at least eleven years, only sequentially dispersing through five events (Events 1-5: Figure 3.5) that occurred in-between 1938 and 1990.

The first three of these five events were from Zimbabwe into first, South Africa in Southern Africa (first intra-southern Africa movement: 95% HPD 1949 - 1968; BF = 14938), secondly into Nigeria in West Africa (first southern-to-west Africa movement: 95% HPD 1962 - 1982; BF = 7.5), and thirdly into Uganda in East Africa (first southern-to-east Africa movement: 95% HPD 1957 - 1974; BF = 4974). Dispersal of MSV-A_1 into South Africa from Zimbabwe appears to have been imminent considering that the two countries shared an international border, and in-between 1948 to 1953, both enjoyed strong political and economic trade relations, the highlight of which was a jointly-operated customs agreement under which most

export and import duties on products such as maize were waived (Phimister, 1991). The proximity of the locations and the repeated cross-border trade movements definitely increased the chance of trans-boundary movement of MSV-infested vector species and/or plant material. It is however unclear as to why the timing of this inferred movement is not concordant with what is accepted as the first credible MSD outbreak reported in the KwaZulu-Natal Province of South Africa around the 1870s by Fuller (Fuller, 1901). The fourth and fifth events in this period included the first west-to-central African MSV-A_2 movement from Nigeria into Chad (HPD 1962 - 1987; BF = 6.5) and the only intra-east African movement from Kenya into Uganda (95% HPD 1979 - 1984; BF = 74723).

There are several factors that probably contributed to the spread of MSV-A_1 into and around Nigeria, and thereafter, its spill-over into west and central Africa as seen later on. For example, following its seeding with the -A_1 variant from Zimbabwe, Nigeria endured several severe MSD outbreaks between 1960 and 1990 (Esenam, 1966; Fajemisin *et al*., 1976; Efron *et al*., 1989). Its area under maize production increased from one to 5.4 million hectares in-between 1985 and 1990 (Fakorede *et al*., 1997; Fakorede, 2002), and Dabrowski (1987a) identified *Cicadulina* species of *C. China ghaurii* and *C. China hartmansi* (*Hemiptera*: *Cicadellidae*) in the country and in neighboring Cameroon in 1983 and 1986. Now, although more than 36 agricultural research institutes (including the International Institute of Tropical Agriculture (IITA), the Kenyan Agricultural Research Institute (KARI), the International Maize and Wheat Improvement Center (CIMMYT)) had responded to the threat of MSD by collaborating in the development and distribution of MSV resistant maize genotypes across west and central Africa (IITA, 1986; Buddenhagen and Bosque-Perez, 1999; Hughes and Odu, 2003), by the early 1990s, only 155 of these had been released, and at least 50% of Nigerian farmers were yet to adopt them (Manyong *et al*., 2000). The possibility of low yields was the biggest deterrent to adoption of newly improved varieties (Fakorede, 2002; Manyong *et al*., 2000; . Analyses of socio-political data reflects that the spill-over of MSV-A_2 from Nigeria into Chad occurred at a time when stronger, mutually beneficial bilateral trade ties and an unregulated informal border trade sector existed between the two countries, where Chad imported foodstuffs, maize included and manufactured goods from Nigeria and exported

livestock, dried fish, and chemicals to Nigeria (Omede, 2006). These conditions possibly encouraged the spread of MSV susceptible maize genotypes across west and central Africa, and also the dispersal of MSV-A and/or leafhopper infested maize and/or other plant materials across the region and the continent.

Meanwhile, in east Africa, the inferred introduction of MSV-A_1 into Uganda from Kenya, coincided with: (i) Uganda importing and cultivating Kenyan hybrids susceptible to MSV (Balirwa, 1992); (ii) the first report of virus-like particles associated with MSD in Uganda by Sylvester *et al*. (1973); (iii) a reduction in maize production within Uganda caused by famers struggling to manage new and poor quality maize seed varieties (Guthrie 1978; Rubaihayo *et al*., 1985). Across the border in Kenya, severe MSD epidemics had occurred in-between 1978 and 1994 (Mwangi, 1998; Hilbeck and Andow, 2004; Republic of Kenya, 2004), and Howell (1984) had identified and described the physical structure and organization of a Kenyan MSV isolate, confirming that the MSV-A progenitor from Zimbabwe had already reached Kenya by 1980 and was dispersing further across the continent just as the African Economic Community (AEC) grew larger (Adar, 2011).

Figure 3.5 Five of the 19 Bayes factor (BF) supported migration pathways through which MSV-A emigrated from Zimbabwe into South Africa (1), Nigeria (2), Uganda (3); and from Nigeria into Chad (4) and then from Uganda into Kenya (5) between 1938 and 1990. Migration event numbers are enclosed in white circles and in brackets. White arrows indicate direction of movement. Light red line = significant BF support, Dark red line = strong BF support, and Black line = decisive BF support.

3.6.3.2 Movements from the early 1990s up to the early 2000s

Since 1938, MSV-A from Zimbabwe had emigrated from southern to west, west to central and within eastern Africa. But in-between the early 1990s and the early 2000s, four MSV-A movements (Figure 3.6), all of which have previously been reported (Monjane *et al*., 2011), are inferred here using the discrete phylogeography model to have occurred. However, in this period, the dispersal trajectory included movements from southern to west Africa (n=1), those localized within southern and west Africa (n=2), and those from east headed toward central Africa (n=1). The occurrence of four events within this decade (1990-2000) indicates a substantial increase in the dispersal rate compared to that inferred between 1938 and 1990.

The first of these four movements was the second southern-to-west Africa introduction of MSV-A_1 into Burkina Faso from Zimbabwe (95% HPD 1991 - 1994; BF = 5.2). Immediately thereafter, Konate and Traore (1992, 1994) identified maize as an MSV reservoir and showed that the virus was widely distributed across the Sudan-Sahel region, Burkina Faso included. More importantly, by 1994, Burkina Faso and most member states of regional trading communities (RTCs) such as the West African Economic and Monetary Union (WAEMU) were actively trading in maize (Barka, 2012). In addition to South Africa, Nigeria, and Uganda, this movement was actually the fourth seeding of a country on mainland Africa with MSV-A from Zimbabwe, reflecting Zimbabwe's contribution to the continental dispersal of the virus. At least four factors appear to have influenced this multi-directional continent-wide dispersal of the -A_1 strain from Zimbabwe between 1938 and 1995. These include (i) the occurrence of leafhoppers plus MSD in Zimbabwe (van der Merwe, 1926; Storey, 1932; Hopkins, 1936; Fennah, 1960; Ghauri, 1961, 1964, and 1971; Caulfield, 1994); (ii) a high maize production rate in Zimbabwe well over one million metric tonnes per annum especially between 1985 and 1990 (Maphosa, 1994); (iii) bulk exports of non-MSV resistant/tolerant maize varieties (particularly CG4142 released in 1993 and C6222 in 1994) facilitated by the country's Economic Structural Adjustment Programme's (ESAP) export retention scheme, which sort to maximize returns from cash crops (such as maize) and ensure a rapid and sustained economic growth through policy reforms (Maphosa, 1994;

USAID/ZIMBABWE, 1993; Makoni, 2000; Potts, 2010; Unganai, 1994); and (iv) the mid-1990s release by Pannar seed company, the Seed Coop in Zimbabwe and the Institute of Agronomic and Tropical Research (IRAT) on the Island of Reunion of 180 MSV resistant maize genotypes across southern Africa and islands in the Indian Ocean (Rodier *et al.*,1995).

Thereafter, the discrete model infers that MSV-A_1 continued dispersing, first through the second intra-southern Africa movement from South Africa into Mozambique (95% HPD 1991 - 1998; BF = 29). Forty-years earlier, de Carvalho (1948) had reported the existence of *Cicadulina* species in Mozambique. After this, Nunes *et al*. (1985) and Denic *et al*. (2001) went on to report high MSD incidences throughout Mozambique's agricultural production regions and the presence of MSV in the country was also acknowledged (Thottappilly *et al*., 1993). MSV-A_1 is then inferred to have dispersed from Nigeria into Benin (95% HPD 1992 - 1998; BF = 14.6) in what appears to be the first intra-west African bound movement. But as early as 1974, Conte (1974), had reported the occurrence of MSD in Dahomey, a kingdom located in southern Benin. Nine years later in 1983, Zagre (1983), went on to report the existence of MSV and the transmission efficiency of the *Cicadulina triangula* leafhopper species in Benin. Meanwhile in east Africa, shortly after confirmation of the presence of MSD in Uganda in the 1970s (Sylvester, 1973; Guthrie; 1978), MSV-A_1 is inferred to have emigrated from Uganda into the Central African Republic (95% HPD 1992 - 2004; BF = 74723), in the first and only east-to-central Africa movement.

Maize exports from Uganda had just risen in response to the emergence of five new seed firms on the Ugandan market, namely East Africa Seeds, Kenya Seeds, Farm Inputs Care Centre Limited (FICA), Harvest Farm Seeds, and Nalweyo Seed Company Limited (NASECO) (Larson and Mbowa, 2004). These firms sought to outcompete Uganda Seed Project (USP), a government owned company that monopolized maize production in Uganda before the 1990s (Larson and Mbowa, 2004). Therefore USP, was perhaps responsible for the widespread cultivation in Uganda, and bulk export across Africa and islands in the Indian Ocean of MSV susceptible maize varieties that include White star and Western Queen released in 1960 and also Kawanda Composite A (KWCA) released in 1971 (Balirwa, 1992; Buddenhagen and Bosque-Perez, 1999). It is not surprising that: (i) maize yield losses

in Uganda, as high as 80% caused by MSD were reported (Balirwa, 1992), and that (ii) the widespread distribution in Uganda of the recombinant variant designated MSV-A_1 UgIII by Owor *et al*. (2007), which accounted for more than 60% of the MSV infections observed from the 155 locations sampled throughout the country between May and June of 2005 was reported. Since then however, several MSV resistant maize varieties, such as *Longe1*, *Longe4*, and *Longe5*, are available in Uganda, and reports of severe MSD outbreaks have dwindled (Bua *et al*., 2010).

Figure 3.6 Four of the 19 Bayes factor (BF) supported migration pathways through which MSV-A emigrated from Zimbabwe into Burkina Faso (6); from South Africa into Mozambique (7); and from Nigeria into Benin (8) and then from Uganda into the Central African Republic (9) between the early 1990s and the early 2000s. Migration event numbers are enclosed in white circles and in brackets. White arrows indicate direction of movement. Light red line = significant BF support, Dark red line = strong BF support, and Black line = decisive BF support.

3.6.3.3 Movements from the late 1990s into the 21st century

Moving from the late 1990s into the 21st century, ten migration events (Figure3.7: twice the number of events inferred between 1938 and 1990 and six more that those inferred between the early 1990s and early 2000s) were inferred by the discrete phylogeographic model to have occurred. The first of these ten movements was the third intra-southern African MSV- A_1emigration from South Africa into Lesotho (95% HPD 1998 - 2004; BF = 74723). While this study could not find any MSD outbreak report within Lesotho since it gained independence in 1965, Lesotho experienced several erratic, unpredictable rainfall patterns and cyclic droughts in-between 1992 and 2002 (Amstader and Eriksen, 1994; Mafura, 2015), creating maize deficits that may have increased maize imports from South Africa (Mafura, 2015; Amstader and Eriksen, 1994), and possibly dispersal of MSV-infested leafhoppers into the country. Bear in mind that, Pannar Seeds Company only released the first MSV resistant/tolerant maize varieties in South Africa in 1995 and then again in 1999 and 2001 (Martin *et al*., 1999, 2001), meaning, varieties exported prior to that were susceptible to MSV, which explains why three MSV sequences were sampled from maize in Lesotho in 2005 (Harkins *et al*., 2009; Monjane *et al*., 2011).

Next to occur, was the only east-Africa-to-Indian Ocean island MSV-A_1 movement from Uganda into Madagascar (95% HPD 1997 - 2005; BF = 187) for reasons already explained prevailing within Uganda, and because Uganda was actively establishing and growing trade partnerships with countries that relied on and consumed white maize as a staple, and whose regional climate and/or physical landscape predisposed them to recurrent droughts, low maize yields and/or famine (Balirwa, 1992). These countries include South Africa, Zimbabwe, Zambia (seen later on) and Madagascar. In fact, it was in this era that liberalized trade between member states of the East African Community (EAC), Uganda included and the Indian Ocean Commission (IOC) expanded (Wangia *et al*., 2004; Adar, 2011). However, after this, MSV-A_1 emigrated from Nigeria into Cameroon (95% HPD 1995 - 2005; BF = 24902), in the second west-to-central Africa movement, a development that coincides with 20-to-50 % (Caldwell *et al*., 1997), and at least 90% MSD incidence reports in Cameroon in 2001 and 2007 (Ngoko *et al*., 2001; Leke *et al*., 2009). Now, although relations

between Cameroon and Nigeria had long been estranged since 1993 over ownership of the Bakassi Peninsular (Omede, 2006), Nigeria was Cameroon's largest import origin in Africa (OECD, 2001). Moreover, the existence, distribution and MSV-A transmission efficiency of *C. China ghaurii* and *C. China hartmansi* leafhopper species in both countries had already been reported in 1986 (Dabrowski, 1987a). As this analyses shows, this was the third out of the six spill-over events of MSV-A_1 from Nigeria into neighboring countries in central (n=1) and west (n=5) Africa between 1960 and 2005. The movement was bound to occur given that Nigeria was the fastest developing former British colony in west Africa, it had the largest mobile and actively trading population on the continent and it also took greater strides to dictate, maintain and profit from political and economic ties with its neighbors (Nicholson, 1969; Odeme, 2006).

Thereafter, the discrete model infers that the first and only east-to-southern Africa MSV- A_1 emigration from Uganda into Zambia occurred (95% HPD 2001 - 2004; BF = 4388), followed by the second southern-to-eastern Africa MSV-A_1 emigration from Mozambique into Uganda (95% HPD 2000 - 2004; BF = 226). As stated earlier, MSV-A_1 had been introduced twice into Uganda, first from Zimbabwe between 1957 and 1974, and then again from Kenya between 1979 and 1984. Full liberalization of maize trade and marketing in Zambia in 1995 by the Zambian State Board and the Food Reserve Agency (Wangia *et al*., 2004), followed by a 30% fall in production caused by the droughts in 2000 and 2001, combined with reports of several MSD epidemics (IITA, 1986; Thottappilly *et al*., 1993; CABI and EPPO, 1997), led to the Zambian government issuing Uganda Grain Traders a contract to supply 40 000 tonnes of poor quality, MSV-susceptible maize varieties (Balirwa, 1992; Buddenhagen and Bosque-Perez, 1999; News24, 2001; FEWS, 2001; Masih *et al*., 2014). The sampling of MSV-A in 2007 and 2008 by Monjane *et al*. (2011) in the country plus collection of *Cicadulina* species and demarcation of their transmission efficiencies through the Pestnet project spearheaded by the International Center for Insect Physiology and Ecology in Zambia, support the claim that the inferred movement did indeed occur in the period indicated (ICIPE: Kaitisha, 2003). Considering Uganda's aggressive inter-and-intra regional maize trade policy (Balirwa, 1992); the known occurrence of MSV, MSD epidemics and *Cicadulina* species in Mozambique (de Carvalho, 1948; Thottappilly *et al*., 1993; Nunes *et al*., 1985; IITA,

1986; Denic *et al*., 2001; Monjane *et al*., 2011), the seeding of the -A_1 strain into Uganda through leafhoppers and/or viral infested plant materials was not surprising.

This sequence of events was immediately followed by the third west-to-central Africa bound movement from Nigeria into the Central African Republic (95% HPD 2000 - 2004; BF = 74723), and by the second and third intra-west African bound MSV-A_1 movements, first from Nigeria into Burkina Faso (95% HPD 2001 - 2005; BF = 697) and again from Nigeria into Ghana (95% HPD 2002 - 2005; BF = 74723). These three movements were the fourth, fifth, and sixth emigrations of the virus from Nigeria into central and west Africa, indicating Nigeria was clearly now a significant dispenser of MSV-A_1 across Africa between 2000 and 2005. It is estimated that Nigeria's trade relations with its regional neighbors were characterized by it offloading cheap goods, maize hybrids included, which earned Nigeria between US$1.5 and US$1.9 billion annually (OECD, 2001; Ahmed, 2012). The next and last movements inferred in this period included the first and only intra-Indian Ocean Island MSV-A_1 movement from Madagascar into Moheli (95% HPD 2004 - 2008; BF = 23.3) which has not been reported before in any other phylogeographic analyses, and the only west-to-east Africa bound MSV-A_1 movement from Ghana into Kenya (95% HPD 2005 - 2008; BF = 1351).

Prior to the introduction of MSV-A_1 into Moheli, and the most recent sampling of 53 isolates of the virus (used here) on the island in 2009 and 2010, the occurrence of MSV in Madagascar (Brunt *et al*., 1990; CABI and EPPO, 1997) and also in the islands of the western Indian Ocean (Autrey, 1983), had been reported. As the largest and most politically and economically stable of the Indian Ocean islands, Madagascar had more resources to cultivate and market its maize produce, but high levels of corruption on the island, and within the region (Comoros, 2014), plus the establishment of the African Free Trade Zone (AFTZ) and increase in liberalized maize trade between the East African Community (EAC), the Southern African Development Community (SADC) and the Indian Ocean Committee member states spurred the dispersal of the virus across mainland Africa and into islands such Moheli (Adar, 2011). Smaller islands such as Moheli and Anjouan that were and are poorly developed, do not have research institutes or centers of higher academic education, experience irregular rainfall patterns and have limited land to support mass maize

production. As such Moheli was and has largely been a net maize importer from Madagascar (Comoros, 2014). Figure 3.1 (node marked by a white star), clearly shows that the genetic distance between the two MSV-A_1 isolates sampled from Moheli and those from Madagascar is relatively small, and more significantly, these sequences cluster with and share a recent common ancestor with those from southern (Lesotho, South Africa, Zambia and Zimbabwe), central (Central Africa Republic) and eastern Africa (Kenya and Uganda), probably because they have share evidence of four unique intra-strain recombination events (4, 8, 17, and 29).

Back in West Africa, the occurrence of MSD and MSV was reported in Ghana as early as 1929 (McKinney, 1929), and again in 1988 (Pinner *et al*., 1988) and 1997 (CABI and EPPO, 1997). This was well before the introduction of the -A_1 strain from Nigeria in-between 2002 and 2005, as mentioned earlier on. From 1990 onwards, the release of MSV resistant maize seed was undertaken by more than 36 agricultural research institutes across Africa (Buddenhagen and Bosque-Perez, 1999; Hughes and Odu, 2003). However, release and adoption of these MSV-resistant hybrids took longer than anticipated from Nigeria to other parts of Africa. For example, Obatanpa, is an open pollinated variety that is tolerant to MSV and represents 95% of all maize grown in Ghana but it was only released in the country in 1992, and between 1999 and 2006 no improved varieties were released in the country by the International Institute for Tropical Agriculture (IITA), the International Maize and Wheat Improvement Center (CIMMYT), the Centre for Scientific and Industrial Research - Crop Research Institute (CSIR-CRI Kumasi) and/or CSIR-Savanna Agricultural Research Institute (CSIR-SARI Tamale) (IITA, 2013). Twelve new varieties were only released later on between 2007 and 2012 (IITA, 2013), but these were mostly drought tolerant and not MSV tolerant or resistant, which may explain the dispersal of the virus from Ghana into Kenya and more recently, the occurrence an MSD epidemic reported in 2010 in Ghana (Oppong *et al*., 2015).

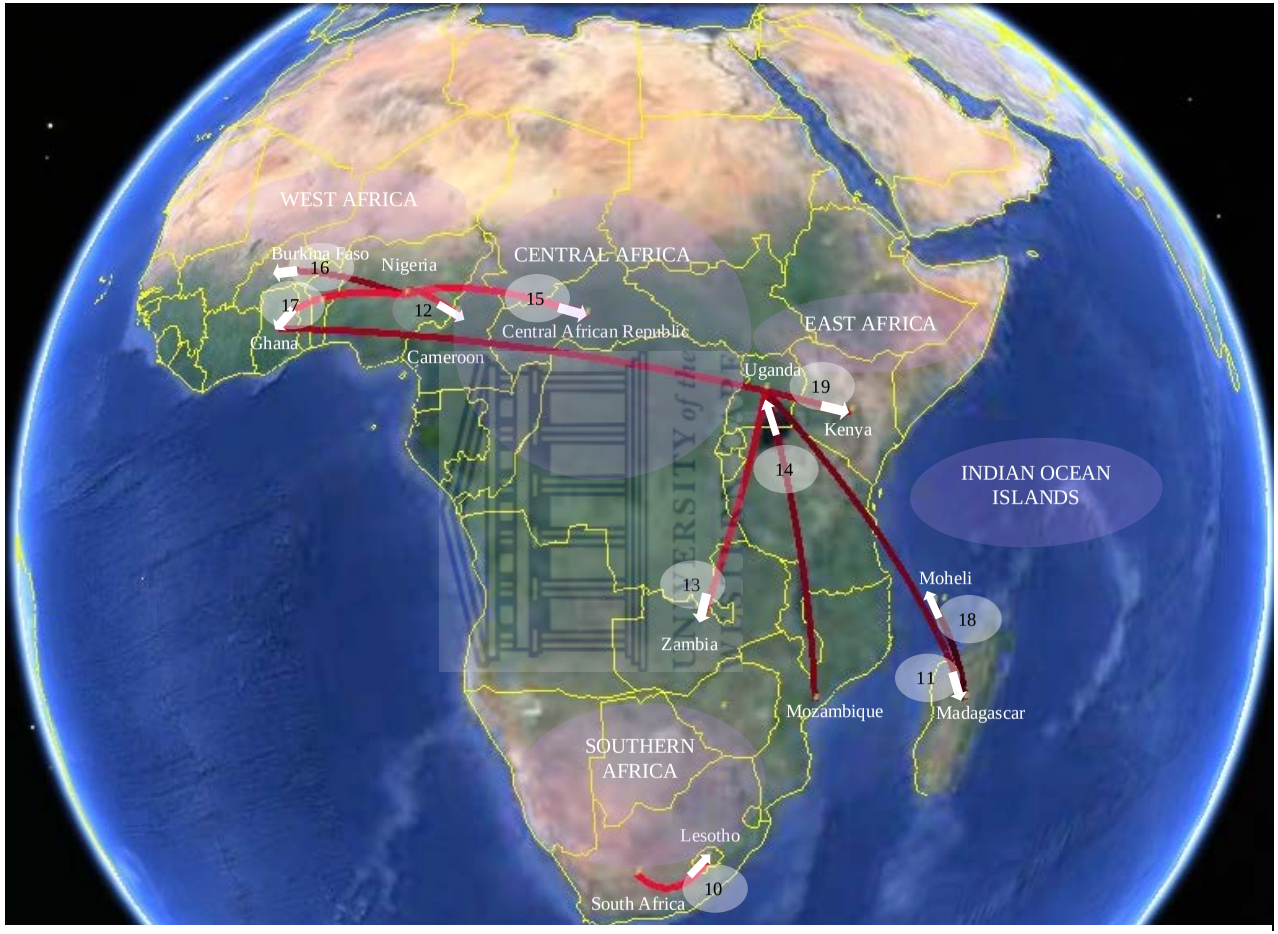


Figure 3.7 Ten of the 19 Bayes factor (BF) supported migration pathways through which MSV-A emigrated from South Africa into Lesotho (10); from Uganda into Madagascar (11); from Nigeria into Cameroon (12); from Uganda into Zambia (13); from Mozambique into Uganda (14); from Nigeria into the Central African Republic (15); from Nigeria into Burkina Faso (16); from Nigeria into Ghana (17); from Madagascar into Moheli (18); and from Ghana into Kenya (19), from the late 1990s into the 21st century. Migration event numbers are enclosed in white circles and in brackets. White arrows indicate direction of movement. Light red line = significant BF support, Dark red line = strong BF support, and Black line = decisive BF support.

3.7. MSV-A Source-Sink Dynamics

3.7.1 Identifying and Characterizing Source and Sink Locations

Locations between which viral pathogens such as MSV-A diffuse, provide conditions that will either support persistence or extinction of variants of the pathogen at different time points. The relationship between these diffusion locations and the conditions regulating persistence and extinction amounts to the source-sink dynamics (Pulliam, 1988; Pulliam and Danielson, 1991). As the theory of source-sink dynamics states, the term source loosely refers to locations with a demographic surplus of viral pathogen as well as conditions ideal for persistence while the term sink refers to locations with a demographic deficit of viral pathogens and conditions that generally promote extinction of the virus over the long term. But, because the conditions prevailing within a given location are not always constant, for example changes in economic policy, which can transform trade practices, a single location may be found acting as both a source and a sink within different time points. Moreover, emigration of pathogens with high multiplication rates from source into sink locations can rescue populations with low multiplication rates from extinction, and enable persistence in otherwise harsh environments and time periods (Amaresekare, 2004). Therefore, understanding and interpreting source-sink dynamics is one way of logically explaining the spatio-temporal patterns in the distribution and abundance of MSV-A variants across Africa and the adjacent Indian Ocean Islands. Consequently, I used the generalized linear model (GLM) to identify and characterize source and sink locations for MSV-A.

The trunk rewards that estimate the contribution that each location makes to the persistence of the trunk lineage estimated from the posterior distribution of trees revealed that MSV-A persistence was predominantly longer in Zimbabwe, followed by Uganda, then South Africa and Kenya respectively compared to other locations considered. The generalized linear model (GLM) inferred that between 1938 and 2011, Zimbabwe occupied 52% of the tree trunk-time, followed by Uganda with 29%, then South Africa and Kenya with 12% and 7% respectively (Figure 3.8 and 3.9). Persistence of MSV-A was highest in Zimbabwe between 1938 and 1974; then in

South Africa from 1974 to 1983; followed by Uganda from around 1984 to 2003; and then finally in Kenya from 2003 until 2011.

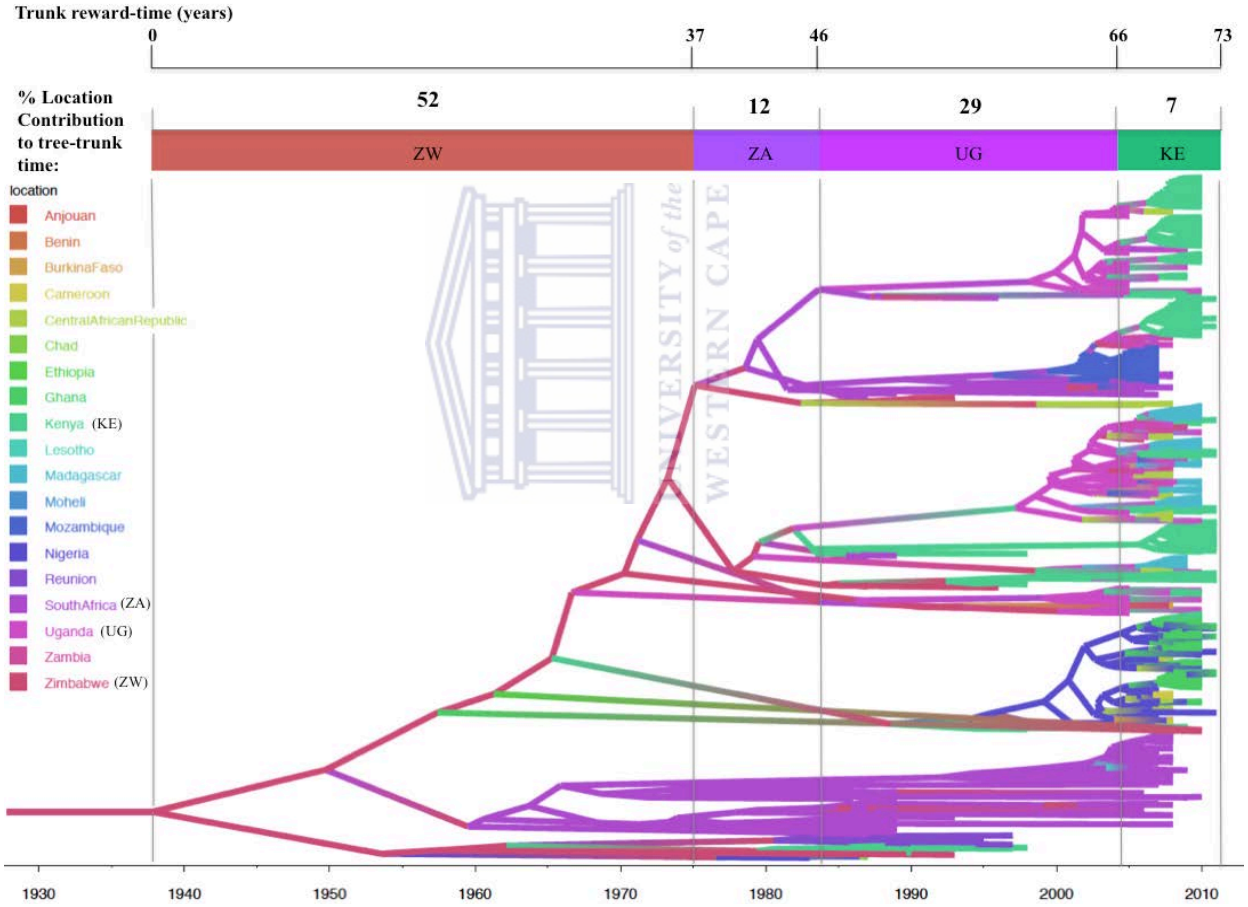


Figure 3.8 Maximum clade credibility (MCC) tree colored according to the time MSV-A persisted in each of the sampling locations as inferred by the GLM diffusion model. Branches are colored according the Markov reward estimates for each location. The trunk lineage in the tree is represented by the upper branch path in the back-borne of the tree from the root to the nodes that represent the ancestors of samples that are exclusively from December 2011.

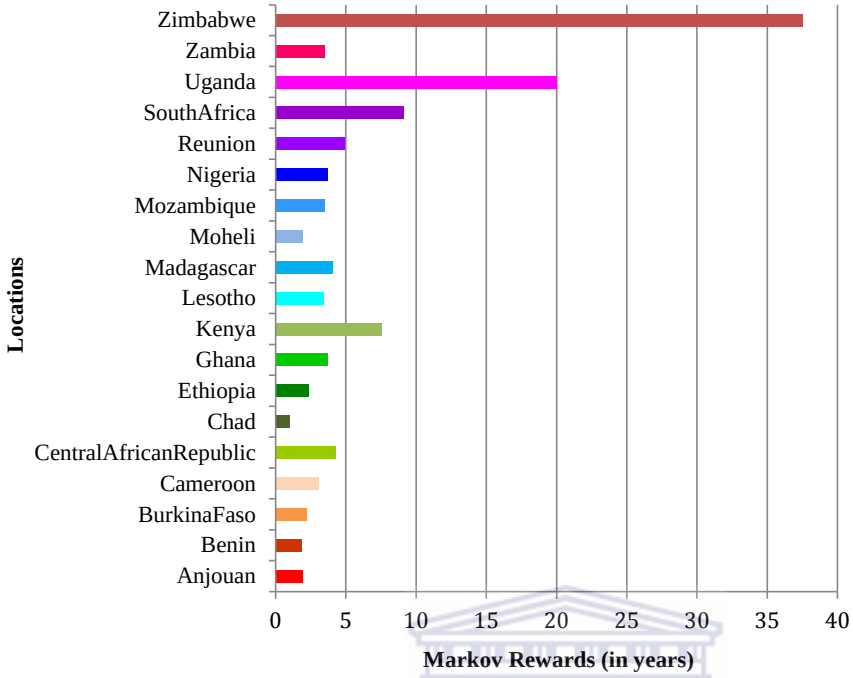


Figure 3.9 Markov Rewards representing the waiting time or persistence of the virus (in years) in each of the 19 sampling locations investigated using the generalized linear model (GLM). Zimbabwe, Uganda, South Africa and Kenya were inferred to have the highest number of Markov rewards suggesting that MSV-A persisted longer in those countries and that the same countries may be important sources and epidemiological links in the persistence of the virus across Africa and the adjacent Indian Ocean Islands.

The highest net Markov jumps were estimated for Uganda, followed by South Africa, Nigeria, Zimbabwe, Kenya, Ghana and Lesotho (Figure 4.0). For the most part, the inference made by the generalized linear model regarding immigrations and emigrations of MSV-A from Uganda, South Africa, Nigeria, Zimbabwe, Kenya, Ghana, and Lesotho is consistent first with the number of Bayes factor supported movements identified for these countries using the discrete phylogeographic model. The inference is also coincides with economic reforms, political and agronomic developments, drought, pathology, and MSD epidemic reports in these countries between 1850 and 2011. Furthermore, while this analyses suggests that Uganda, South Africa, Nigeria, Zimbabwe, Kenya and Ghana were major sources of MSV-A dispersal, locations such as the Anjouan, Burkina Faso, Benin, Cameroon, Central African republic, Chad, Lesotho, Ethiopia, Moheli, Mozambique, Madagascar,

Reunion and Zambia, were characterized as mostly sinks. However, South Africa, Zimbabwe, Nigeria, and Uganda probably acted as both source and sink locations in different time points.

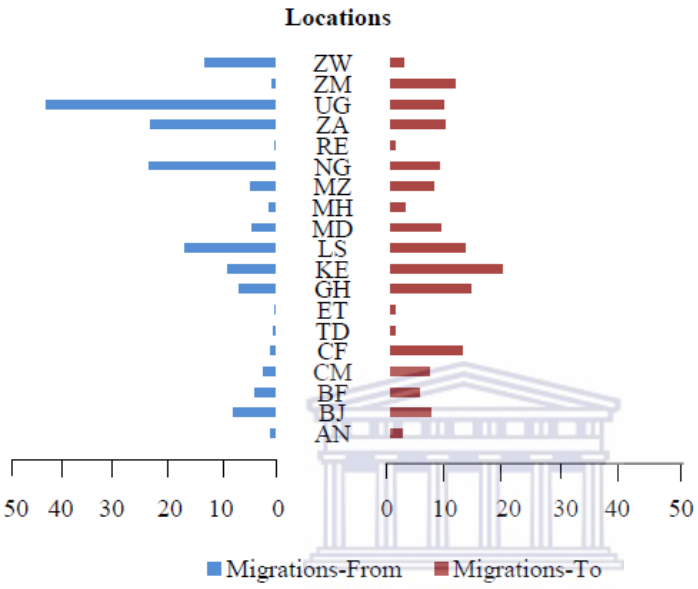


Figure 4.0 Net Markov jumps from and towards each of the 19 locations considered as inferred using the generalized linear model (GLM). The highest net Markov jumps to and from are inferred for Uganda, followed by South Africa, Nigeria, Zimbabwe, Kenya, Ghana and Lesotho.

3.8 Predictor variables possibly determinant of MSV-A dispersal

Of the 27 predictor variables investigated as possible determinants of MSV-A dispersal across Africa and the adjacent Indian Ocean Islands amongst source and sink locations using the generalized linear model (GLM), only five were inferred with high Bayes factor support (Table 3.5 - Appendix 16; Figure 4.1 - Appendix 17) to have influenced the spatio-temporal diffusion of MSV-A since its emergence. This analysis also permitted the characterization of sampling locations into sources and sinks based on the inferred inclusion probabilities and the predictor variable contributions to the dispersal process (Figure 4.2 - Appendix 18). Compared to sink locations, locations inferred as sources with the GLM, were characterized by high average annual precipitation and moderately high average annual temperatures and experienced high seasonal changes. Sources also produced high maize yield but experienced high prevalence of undernourishment, had lower trade imports and exports, high GDP per capita, lower vector control pesticide usage, high percentage forest land area, lower percentage arable land, high population densities, and were in close proximity to sink locations. The characterization of the variation in the 27 predictive variables, classified into four categories (climatic; socio-political and economic; ecological; and geographical distance) amongst source and sink locations investigated (Table 3.6) as inferred by the generalized linear model (GLM), and how this possibly influenced the spatio-temporal diffusion of MSV-A across Africa and the adjacent Indian Ocean Islands is described below.

Table 3.6 Characterization of sources and sinks based on the 27 predictive variables investigated

SOURCES			SINKS		
high	**moderate**	**low**	**high**	**moderate**	**Low**
precipitation seasonal changes	temperature			temperature	Precipitation seasonal changes
yield undernourishment GDP per capita Population density		trade imports trade exports	trade imports trade exports	yield undernourishment	GDP per capita Population density
forest land area pesticide usage		arable land	arable land pesticide usage		forest land area
proximity					Proximity

Key:

source sink **Predictive Variables Classes**

- Climatic factors
- Sociopolitical and economic factors
- Ecological factors
- Geographical factors

3.8.1 Climatic factors

Across Africa, maize production thrives in tropical and subtropical, and savanna climates, mostly in mid-altitude areas (Buddenhagen and Bosque-Perez, 1999; Sleper and Poehlman, 2006) where leafhopper species are also widely distributed (CABI, 1986; Dabrowski, 1987; Mylonas, *et al.*, 2014; Oluwafemi *et al.*, 2007; Reynaud *et al.*, 2009; Rose, 1978). The GLM inferred that sources largely received high average annual precipitation levels (BF = 0.96) and had moderately high average annual temperatures (BF = 0.04) throughout each agricultural season between 1960 and 2013 compared to sinks. Although intensively grown in mid-altitude areas which receive high and moderate average annual precipitation and temperatures, by the mid-1970s maize was prone to MSD in these areas, and it was also absent in areas were conditions best suited its cultivation such as the savanna lands of West Africa (Buddenhagen and Bosque-Perez, 1999). By 1999, Buddenhagen and Bosque-Perez

(1999) reported that the majority of farmers in East Africa were still growing MSV susceptible varieties, which probably explains the spatio-temporal diffusion of the virus from that region to the rest of the continent and into the Islands in the Indian Ocean.

The model also suggests that the position of source locations from the equator [latitude predictive variable (BF = 3.03) - Appendix 17] significantly influence the dispersal of MSV-A into sinks. A latitude and longitudinal of a geographical location influences changes in seasons and climatic conditions experienced in that location (Khavrus and Shelevytsky, 2010). A season is a division of the year, marked by changes in rainfall densities, mean temperature, ecology and hours of daylight, and more importantly, seasons result from the yearly orbit of the earth around the Sun and the tilt of the earth's rotational axis relative to the plane of the orbit (Khavrus and Shelevytsky, 2010; 2012). Following evaluation of the H3N2 influenza epidemic, Lemey *et al*. (2014) converged on the position that disease dispersal is higher between the most connected countries that share similar seasons. Therefore, for a virus such as MSV-A to persist, it must infect multiple hosts, and the rate at which infection and transmission occurs is expected to be higher when locations occupy proximal regions of the equator and share seasons. In a related case, the frequent occurrence of drought, followed by high rainfall has been reported to promote the dispersal of leafhopper species and result in high MSD incidences particularly in the second season for locations with a bi-seasonal maize agronomic calendar (Martin and Shepherd, 2009). Between 1900 and 2011, several drought periods appear to have been followed by excess rains in countries across Africa and the adjacent Indian Ocean Islands (Masih *et al*., 2014). A good example is Zimbabwe where following drought spells from 1990 to 1992 and again between 1994 and 1995 (Unganai, 1994), an increase in irrigated maize production and heavy rains around early 1993 up to late 1993, and early 1994, is reported to have resulted in high MSD incidences in the country (Caulfield, 1994; Unganai, 1994).

3.8.2 Sociopolitical and economic factors

Several sociopolitical and economic factors may play a significant role in the spread of vector-borne plant viruses such as MSV-A. Factors under this category investigated with the GLM include the per capita gross domestic product (GDP per capita), maize yield per hectare, maize trade imports and exports, and the population density. My results show that migrations of MSV-A were highest from locations with high GDP per capita (BF = 3960), maize yield (BF = 0.77), population densities (BF = 0.01), and prevalence of undernourishment (BF = 3960). GDP per capita is the gross domestic product divided by the midyear population. The discriminatory power of this predictive variable was significant strong between source and sink locations most probably because countries with higher GDP per capita have more resources to spend and invest per capita on research into disease prevention and control (Meo *et al*., 2013), resulting in lower MSD outbreak frequencies compared to countries with lower GDP per capita. High GDP per capita is usually a result of good political and economic governance, health, agricultural, mining, and educational systems. The islands of Anjouan and Moheli, are good examples of sinks, as both endure political and economic instability and have low GDP per capita (Comoros, 2014), whereas South Africa, Nigeria, Uganda and Kenya are ideal sources as they currently have stable political and economic environments.

Although the GLM inferred that sources had both low maize trade imports and exports compared to sinks, exports of MSV-A susceptible varieties by these countries into sinks probably elevated the spatio-temporal diffusion of the pathogen. The higher imports and exports of maize into and from sinks corresponds with the susceptibility of these countries to frequent droughts (Masih *et al.*, 2014), their localization in either arid to semi-arid regions, the high prevalence of informal and unregulated maize trade routes across borders and ports, the emergence of several new maize seed producers of the continental market; political unrest and high corruption and also the growing integration of these nations into regional and continental (Africa) economic communities which enabled the creation of free trade zones, harmonization of trade tariffs, and reduction of duty and taxes (OECD, 2001; Ahmed, 2012; Barka, 2012; Comoros, 2014). For example the increase in trade between EAC, WAEMU, COMESA, IOC, and SADC member states from the late 1990s onwards (Adar, 2011; Ahmed, 2012; Barka, 2012), may have promoted an increase in MSV-A dispersal into

within east Africa, into west, central, and southern Africa and the adjacent Indian Ocean Islands.

3.8.3 Ecological factors

Locations inferred as sources with the GLM were identified to have a high percentage of forest land area (BF = 3.92) and a high pesticide usage index (BF = 0.04) whereas the percentage arable land in these locations was lower than that in sinks (BF = 0. 19). Forests harbour a wide diversity of flora and fauna, and can be reservoirs of both plant and animal viral pathogens. For example, the African streak viruses, which are close relatives of MSV and strains of MSV, from MSV-B to MSV-K have been reported to infect graminaceous weeds and wild grasses, and some of these viruses thrive in forest lands (Shepherd *et al*., 2010; Oluwafemi *et al*., 2014; Oppong *et al*., 2015). In an effort to maximize food production, most sinks, which already have limited land, most of which is either not suitable for maize cultivation as is the case with Lesotho, Benin, Burkina Faso, and the islands of Reunion, Anjouan, Madagascar or is used to cultivate timber under forest plantations as is the case in Ghana, it is possible that these countries may have expanded the proportion of arable land by encroaching into lands designated as forest lands, which were already infested with variants of MSV and through time, this may have allowed virus to establish itself in maize as a serious pathogen. Most countries in west and east Africa, for example Nigeria, Uganda, and Kenya, are reported to grown maize in highland areas, formerly forest lands, and this probably explains the high MSD outbreaks incidences reported in these countries before and after the release of MSV resistant varieties in the mid-1990s (IITA, 1986; Leke *et al*., 2009; Wangia *et al.*, 2004). This result also corresponds to the inferred long-term persistence of MSV-A in Zimbabwe, South Africa, Uganda and Kenya, countries that although they have had high percentage forest land area, most of this was converted to infrastructural and industrial developments such as housing and tilling, or has been reduced significantly as foreign demand for local tobacco and other commercial crops had gradually increased faster in these countries compared to others on the continent.

3.8.4 Geographical factors

The great circle distance, used here as a proxy for the geographical distance, represents the shortest distance between two points on the surface of a sphere, measured along the surface of the sphere, as opposed to a straight line through the sphere's interior. Although this predictive variable did not discriminate between source and sink locations, migrations of MSV-A were inferred with the GLM to have occurred more between geographically proximate locations (BF = 3690). This finding is reasonable given that the geographical distance between locations influences the geographical dispersal range of both the leafhopper vectors infested with MSV as well as people carrying MSV-infested plant material across district, provincial, country and/or regional borders or ports. This is an increasing possibility with, the number of civil wars, corruption, border posts as well as economic free trade zones and liberalization of maize trade are currently on the increase across Africa since the late 1990s (Comoros, 2014; Sitko, *et al*., 2014; Ahmed, 2012).

Comment [G1]: This doesn't make sense. Are civil wars increasing, is corruption, is the number of border posts. Or have they always been high?? This is a very weak concluding sentence

4.0 CONCLUSIONS

MSV-A isolates used had very low levels of genetic diversity as indicated by the very small overall mean pairwise genetic distance (0.017) estimated using MEGA and ≥ 98 % nucleotide sequence similarity estimated using SDT.

Both inter-strain and intra-strain recombination events occur frequently in the MSV genome and while both events rarely occur within gene coding regions, they appear to be ubiquitous towards the periphery of the gene encoding regions in the MSV genome. Detectable intra-strain recombination events marginally outnumbered detectable inter-strain recombination events, and 15 of these events have not been reported in previous analyses (Owor *et al*., 2007, Varsani *et al*., 2008 and Monjane *et al*., 2011). Most recombination breakpoints were detected in the Rep and RepA gene fragments compared to other gene regions and the majority of intra-strain MSV-A recombination events detected have occurred within the last six decades, the oldest and most conserved of these being events 19, 26 and 28 whereas the most recent are events 8, 16, 17, 21, 23, and 29. Viruses displaying evidence of intra-strain recombination events 20, 25 and 33, are widely distributed in East Africa, whereas those displaying evidence of events 16, 21 and 23, occur more frequently in West and Central Africa. However, viruses displaying evidence of the majority of events (1, 4, 8, 10, 14, 17, 19, 22, 24, 25, 26, 28, and 29) are more spread out across East, West, Central and Southern Africa and Islands in the Indian Ocean. These wide distribution and high prevalence of viruses carrying some of these intra-strain recombination events suggests that these events may have played a significant role in MSD epidemiology and that they have contributed to the well structured geographical distribution pattern of MSV-A variants across Africa and the adjacent Indian Ocean Islands.

While codon positions in the movement protein (MP) and the RepA protein encoding genes appear to have mostly evolved neutrally, or under negative selection, two codons positions (positions 12 and 19) within motif I in the coat protein (CP) encoding gene, constituting 0.82 % of the genome and seven codon positions (147, 166, 195, 203, 242, 260, 267) in the Rep gene within the Rb motif and motif IV site A

and site B, constituting 2.57 % of the genome, were positively selected. The majority of codon positions in the CP and Rep genes evolved neutrally or under negative selection pressure. The difference in selection pressure exerted on the individual MSV-A coding regions is most likely linked to the functional constraints imposed on the same regions that probably influences virus-vector, virus-host, or virus-virus interactions.

Zimbabwe was identified as the most probable location of the most recent common ancestor of the MSV-A strain (posterior state probability = 0.52). The estimated mean nucleotide substitution rate for the full genome recombination-free MSV-A dataset H of 9.922×10^{-4} (95% HPD 8.54×10^{-4} to 1.1317×10^{-3}) substitutions per site per year, was much higher than the 2×10^{-4} to 7×10^{-4} range of estimates obtained from short term (<60 days: Shepherd *et al*., 2005, 2006; Walt *et al*., 2009) and long-term (between 1-6 years: Isnard *et al*., 1998; Harkins *et al.*, 2009; van der Walt *et al*., 2008b) experiments. This indicates that my study has very likely over-estimated the actual substitution rate which should be lower over the more than 80 years of MSV evolution represented here than the short term estimates yielded in experiments. This is perhaps why the estimate of the time to the most recent common ancestor obtained under the combined relaxed uncorrelated molecular clock and constant population size model of 73 years (95% HPD 54.9 - 107.5), which is equivalent to a calendar date of 1938 (95% HPD 1904 - 1956), is more recent and consistent with a tMRCA of 1933 (95% HPD, 1867 to 1950) obtained by Monjane *et al*. (2011) in previous analyses with a partial genome (MPCP) dataset but inconsistent with their tMRCA estimate of 1863 (95% HPD, 1809 to 1935) inferred using a full genome (FG) dataset (Monjane *et al*., 2011).

The spatio-temporal diffusion of MSV-A across Africa and the adjacent Indian Ocean Islands occurred through a network of 34 intra-and-inter regional migration events, 19 of which were Bayes factor (BF) supported. Persistence of MSV-A was highest in Zimbabwe, followed by South Africa, Uganda, and Kenya and these countries were inferred as the primary source locations in the spatio-temporal diffusion of variants of this strain across Africa and the adjacent Indian Ocean Islands. Climatic, sociopolitical, economic, ecological, and geographical proximity were amongst the

factors that strongly influenced the MSV-A dispersal process and enabled the characterization of sampling locations into sources and sinks respectively.

5.0 REFERENCES

Adane, A., and Albrechtsen, S. E. 1998. Seed-borne virus infecting some important crops in Ethiopia. *Pest Management Journal of Ethiopia*, 2(1 & 2): 94 - 101.

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, edited by Finizio, G., Levi, L., and Vallinoto, N. pp. 1 - 39.

Ahmed, M. 2012. Analysis of incentives and disincentives for maize in Uganda. Technical notes series, Monitoring African Food and Agricultural Policies (MAFAP). Food and Agricultural Organization, Rome. pp. 6 - 28.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6): 716 - 723. doi:10.1109/TAC.1974.1100705

Alemu, L., Adane, A., Berhanu, B., and Mesfin, T. 1997. Identification of viruses in maize and grass hosts. *Pest Management Journal of Ethiopia*, 1: 73 - 76.

Allicock, O. M., Lemey, P., Tatem, A. J., Pybus, O. G., Bennett, S. N., Mueller, B. A., Suchard, M. A., Foster, J. E., Rambaut, A., and Carrington, C. V. F. 2012. Phylogeography and Population Dynamics of Dengue Viruses in the Americas. *Molecular Biology and Evolution*, 29 (6): 1533 - 1543. doi:10.1093/molbev/msr320

Amarasekare, P. 2004. The role of density-dependent dispersal in source-sink dynamics. *Journal of Theoretical Biology*, 226: 158 - 168.

Ammar, E. D. 1975. Biology of the leafhopper *Cicadulina chinai Ghauri* (Homoptera, Cicadellidae) in Giza, Egypt. *Zeischrift fnr Angewandte Entomologie*, 79 (4): 337 - 345.

Amstader, I and Eriksen, J. 1994. Evaluation of U. S. Government Response to 1991/92 Southern Africa Drought. Prepared for AID Bureau for Humanitarian Response. 600 Water Street, S. W., Washington, D. C. USA. *Management Systems International*, 202: 484 - 7170

Andreu, M. L., Peterson, H. H., Grunewald, O., Norman, D. 2006. Biotechnology and economic development: The economic benefits of *maize streak virus* tolerant maize in Kenya. Southern Agricultural Economics Association Annual Meeting, Orlando, Florida February 5 - 8.

Anonymous. 1983. Plant infectious agents. Viruses, viroids, virusoids, and satellites. *Plant infectious agents Viruses, viroids, virusoids, and satellites*, ix.

Asanzi, C. M., Bosque-Perez, N. A., Buddenhagen, I. W., Gordon, D. T., and Nault, L. R. 1994. Interactions among *maize streak virus* disease, leafhopper vector populations and maize cultivars in forest and savanna zones of Nigeria. *Plant Pathology*, 43(1): 145 - 157.

Asanzi, M. C., Bosque-Perez, N. A., Nault, L. R., Gordon, D .T., and Thottapilly, G. 1995.Biology of *Cicadulina* species (Homoptera: Cicadellidae) and transmission of *Maize streak virus*. *African Entomology*, 3 (2): 173 - 179.

Assefa, T. 1999. Survey of maize diseases in western and northwestern Ethiopia. In: Maize Production Technology for the future: Challenges and Opportunities. Proceedings of the 6th Conference. CIMMYT/EARO. pp. 121 - 124.

Autrey, L. J. C. 1983. Maize mosaic virus and other maize virus diseases in the islands of the western Indian Ocean. In: Proceedings International maize Virus Disease Colloquium and Workshop. Gordon, D. T., Knoke, J. K., Nault, L. R., and Ritter, R. M. Ohio Agricultural Research and Development Center, Wooster, USA. pp. 167 - 181

Azarian, T., Lo Presti, A., Giovanetti, M., Cella, E., Rife, B., Lai, A., Zehender, G., Ciccozzi, M., and Salemi, M. 2015. Impact of spatial dispersion, evolution,

and selection on Ebola Zaire Virus epidemic waves. *Scientific Reports*, 5: 10170. doi:10.1038/srep10170

Azzam, O., Frazer, J., De La Rosa, D., Beaver, J. S., Ahlquist, P., and Maxwell, D. P. 1994. Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. *Virology*, 204 (1): 289 - 296.

Baele, G., Lemey, P., Bedford., T., Rambaut., A., Suchard, M. A., Alekseyenko, A. V. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 30: 2157 - 2167.

Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., and Lemey, P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian Phylogenetics. *Molecular Biology and Evolution*, 30 (2): 239 - 243.

Balirwa, E. K. 1992. Maize research and production in Uganda. PN-ABT-623. pp. 1 - 19.

Barka, H. B. 2012. Border Posts, Checkpoints, and intra-African Trade: Challenges and Solutions. African Development Bank (AfDB). *Chief Economist Complex*.

Barrow, M. R. 1992. Development of maize hybrids resistant to *Maize streak virus*. *Crop Protection*, 11 (3): 267 - 271. doi:10.1016/0261-2194(92)90048-A

Barrow, M. R. 1993. Increasing maize yield in Africa through the use of *maize streak virus r*esistant hybrids. *African Crop Science Journal*, 1: 139 - 144.

Bedford, T., Cobey, S., Beerli. P., Pascual, M. 2010. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathogens*, 6: e1000918. doi.10.1371/journal.ppat.1000918

Bielejec, P., Rambaut, A., Suchard, M. A. and Lemey, P. 2011. “Spread: spatial phylogenetic reconstruction of evolutionary dynamics”, *Bioinformatics*, 27 (20): 2910 - 2912.

Biek, R., Henderson, J. C., Waller, L. A., Rupprecht, C. E., Real, L. A. 2007. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceeding of the National Academy of Sciences*, USA, 104: 7993 - 7998.

Bjarnason, M. S. 1986. Progress in breeding for resistance to the *maize streak virus* disease. *In:* Gelow, B. (ed.) *1st Eastern, Central, Southern Africa Regional Maize Workshop To Feed Ourselves.* Lusaka Zambia, March 10-17, 1985: CIMMYT, Mexico D.F.

Bock, K. R., Guthrie, E. J., and Woods, R. D. 1974. Purification of *maize streak virus* and its relationship to viruses associated with streak diseases of *Sugarcane* and *Panicum maximum*. *Annals of Applied Biology*, 77: 289 - 296.

Bofkin, L., and Goldman, N. 2007. Variation in evolutionary processes at different codon positions. *Molecular Biology and Evolution*, 24: 513 - 521.

Boulton, M. I., Steinkellner, H., Donson, J., Markham, P. G., King, D. I., and Davies, J. W. 1989. Mutational analysis of the virion-sense genes of *maize streak virus*. *Journal of General Virology*, 70: 2309 - 2323.

Boulton, M. I., King, D. L., Markham, P. G., Pinner, M. S., and Davies, J. W. 1991a. Point substitutions in a promoter-like region and the V1 gene affect the host range and symptoms of *maize streak virus*. *Virology*, 183: 114 - 121.

Boulton, M. I., King, D. L., Markham, P. G., Pinner, M. S., and Davies, J. W. 1991b. Host range and symptoms are determined by specific domains of the *maize streak virus* genome. *Virology*, 181: 312 - 318.

Bousque-Perez, N. A., and Alam, M.S. 1992. Mass rearing of Cicadulina leafhoppers to screen for *maize streak virus* resistance. IITA, Ibadan, Nigeria.

Bousque-Perez, N. A., Olojede, S. O., and 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*, 101 (3): 307-317.

Bousque-Perez, N. A. 2000. Eight decades of *maize streak virus* research. *Virus Research*, 71: 107 - 121.

Briddon, R. W., Pinner, M. S., Stanley, J., Markham, P. G. 1990. Geminivirus coat protein gene replacement alters insect specificity. *Virology*, 177 (1): 85 - 94.

Brunt, A., Crabtree, K., Dallwitz, M. J., Gibbs, A., Watson, L. 1990. Viruses of Tropical Plants- Descriptions and Lists from the VIDE Database. Wallingford, UK: CAB International/ACIAR, pp. 446 - 447.

Bua, B., and Chelimo, B. M., Adipala, E., Tusiime, G., and Majaliwa, J. G. M. 2010. The reaction of maize genotypes to maize streak virus disease in central Uganda. In: *Second RUFORUM Biennial Regional Conference on "Building capacity for food security in Africa"*, Entebbe, Uganda, 20 - 24 September 2010. pp. 293 - 297. *RUFORUM.*

Buddenhagen, I., W., and Bosque-Perez, N. A. 1999. Historical Overview of breeding for durable resistance to maize streak virus for tropical Africa. *South African Journal of Plant and Soil*, 16 (2): 106 - 111. doi.10.1080/02571862.1999.10634856.

CABI, and EPPO. 1997. *Maize streak monogeminivirus*. [Distribution Map]. Distribution Maps of Plant Diseases, December (Edition 1). Wallingford, UK: CAB International, Map 789.

CABI. 1986. *Cicadulina mbila*. Distribution Maps of Plant Pests, 1986, December, Map 481.

Cardwell, K. F., Schulthess, F., Ndemah, R., Ngoko, Z. 1997. A systems approach to assess crop health and maize yield losses due to pests and diseases in Cameroon. *Agricultural Ecosystems and Environment*, 65: 33 - 47.

Caulfield, M. J. 1994. A field selection method to screen for resistance to *maize streak virus.* In: Proceedings of the Tenth South African Maize Breeding Symposium, Potchefstrom, South Africa, 17-19 March 1992. du Plessis, J. G., van Rensburg, J. B. J., McLaren, N. W., and Flett, B. C (eds.). pp. 46 - 49.

Coe, E. H. J., Nueffer, M. G., and Hoisington, D. A. (eds.) 1988. The genetics of maize. American Society of Agronomy. Madison, Wisconsin. *Agronomy Monographs* 18: 81 - 236.

Comoros. 2014. Mining Laws and regulations Handbook, Volume 1: Strategic Information and Basic Laws. ISBN 1-4330-719-1 pp. 130 - 190.

Conte, J. Le. 1974. A virus disease of maize in Dahomey. *Agronomie Tropicale*, I, 29 (8): 831 - 832.

Dabrowski, Z. T. 1987.*Cicadulina ghaurii* (*Hem., Euscelidae*): distribution, biology and *maize streak virus* (MSV) transmission. *Journal of Applied Entomology*, 103: 489 - 496.

Dabrowski, Z. T. 1987a. Two new species of *Cicadulina China* (*Hemiptera*: *Euscelidae*) from West Africa, *Bull. Ent. Res*, 77: 53 - 56.

Dabrowski, Z. T., Nwilene, F., and Kumar, R. 1991. First regular observations on leafhoppers, cicadulina spp., vectors of Maize *streak virus* (MSV) in Southeastern Nigeria. *Insect Sci. Appl*, 12: 249 - 261.

Dabrowski, Z. T., Wilson, M. R., and Nault, L. R. 1987a. Comparative studies of Cicadulina leafhoppers in West Africa. *In:* Wilson, M. R. (ed.) 2*nd International Workshop on Leafhoppers and Planthoppers of Economic Importance. 28th July-1st August 1986.* Provo, UT: Bringham Young University.

Damsteegt, V. D. 1980. Investigations of Maize vulnerability of U.S. maize to *maize streak virus*. *Protection Ecology,* 2 (3): 231 - 238.

Damsteegt, V. D. 1981. Exotic virus and virus-like diseases of maize *In:* D. T. Gordon, J. K. K. a. G. E. S. (ed.) *Virus and Virus-like Diseases of Maize in the United States*. Southern Cooperative Series Bulletin 247, June 1981. 210 pages.

Damsteegt, V. D. 1983. *Maize streak virus*: I. Host range and vulnerability of maize germplasm. *Plant Disease*, 67: 734 - 737.

Damsteegt, V. D. 1984. *Maize streak virus*: effect of temperature on vector and virus. *Phytopathology*, 74: 1317-1319.

Damsteegt, V. D. and Igwegbe, E. C. K. 2005. Epidemiology and control of *maize streak virus*. *In:* Hadidi, A., Khetarpal, R. K., and Koganezawa, H. (eds.). *Plant Virus Disease Control*.APS Press copyright 2005.

Darriba, D; Taboada, G. L; Doallo, R and Posada, D. 2012. jModeltest2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772. doi:10.1038/nmeth.2109

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., Adoul-Karime, A. L., Abdou-Chakour, C., Reynaud, B., Harkins, G. W., Varsani, A., Martin, D. P., and Lett, J-M. 2012. East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite

begomovirus. *BMC Evolutionary Biology*, 12: 228. http://www.biomedcentral.com/1471-2148/12/228

De Carvalho, I. 1948. Relacao preliminar de dofncas encontradas em plantes e insectos com anatocoes fitopatologicas. Coloia de Mocambique, *Reporticao de Agricultura, Seccao de Micologia*.

Delport, W., Poon, A., Frost, S. D. W., Kosakovsky Pond, S. L. 2010. DATAMONKEY: A suite of phylogenetic tools for evolutionary biology. *Bioinformatics*, 26: 2455 - 2457.

Delport, W. Poon, A., Frost, S. D. W., Kosakovsky Pond, S. L. 2010. DATAMONKEY Webserver. http://www.datamonkey.org

Denic, M. P., Chauque, C., Jose, M., Langa, D., Mariote, P. Fato., and Haag, W. 2001. Maize screening for multiple stress tolerance and agronomic traits. *Eastern and Southern Africa Regional Maize Conference*, 7: 88 - 91.

Desbiez, C., David, C., Mettouchi, A., Laufs, J., and Gronenborn, B. 1995. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proceedings of Natural Academic Science, USA*, 92: 5640 - 5644.

Doebley, J. F., Goodman, M. M., and Stuber, C. W. 1984. Isoenzymatic variation in Zea (graminae). *Systems Botanical*, 9: 203 - 218.

Drummond, A. J., and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7: 214. http://tree.bio.ed.ac.uk/software/beastv1.8.1

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4: e88.

Drummond, A. J., and Bouckaert, R. R. 2014. Bayesian evolutionary analyses with BEAST 2. Cambridge University Press.

Duchene, S., Duchene, D., Holmes, E. C., and Ho, S. Y. W. 2015. The Performance of the Date-Randomization Time-Structured Virus Data. *Molecular Biology Evolution*, 32 (7): 1895 - 1906. doi:10.1093/molbev/msv056

Duignan, P., and Gann, L. H. 1975. Economic achievements of the colonizers: An assessment. In: *Colonialism in Africa, 1870-1960 Volume 4., The Economics of Colonialism*, 1870-1960, ed. Gann, L. H and Duignan, P, pp. 330 - 696. Cambridge University Press.

Duffy, S., and 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 (pt 6): 1539 - 1547. doi:10.1099/vir.0.009266-0.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32: 1792 - 1797.

Efron, J. M., Kim, S. K., Fajemisin, J. M., Mareck, J. H., Tang, C. Y., Dabrowski, Z. T., Rossel, H. W., Thottapilly, G., and Buddenhagen, I. W. 1989. Breeding for resistance to *maize streak virus*: A multidisciplinary team approach. *Plant Breed*, 103 (1): 1 - 36.

European and Mediterranean Plant Protection Organization (EPPO). 2014. PQR database. Paris, France: European and Mediterranean Plant Protection Organization. http:// www.eppo.int/DATABASES/pqr/pqr.htm

Esenam, U. 1966. Observations on the maize streak disease in Western Nigeria. *Nigerian Agricultural Journal*, 3: 38 - 41.

Etienne, J. Rat, B. 1973. Stripe virus: an important maize disease of maize in Reunion. *Agronomie Tropicale, I*, 28 (1): 11 - 17.

Fajemisin, J. M., Cook, G, E., Okusanya, F., and Shoyinka, S. A. 1976. *Maize streak virus* epiphytotic in Nigeria. *Plant Disease Reports*, 60: 443 - 447.

Fajemisin, J. M. 1984. Basic constraints to maize productivity in Tropical Africa. *U.S. Universities/CIMMYT Maize Conference*, August 8-15, 1984. CIMMYT. Mexico.

Fajemisin, J. M., Kim, S. K., Efron, Y., and Alam, M. S. 1984. Breeding for durable disease resistance in tropical maize with special reference to maize *streak virus*. *FAO Plant Production and Protection Paper*, 55: 71.

Fajemisin, J. M., and Shoyinka, S. A. 1976. Maize streak and other maize virus disease in West Africa. In: Williams, L. E., Gordon, D. T., Nault, L.R. (ed.) *International maize virus disease colloquium and workshop*. Wooster USA: Ohio Agricultural Research and Development Center, pp. 52 - 60.

Food and Agriculture Organization of the United Nations. 2007. Food balance sheets. Food and Agriculture Organization of the United Nations, Rome, Italy. http://faostat.fao.org/site/368/desktopdefault.aspx?pageid=368#ancor. Accessed 17th August 2014.

Food and Agriculture Organization Corporate Statistical Database (FAOSTAT). 2012. Statistical databases and data-sets of the food and agriculture organization of the United Nations http://faostat3.fao.org/browse/rankings/countries_by_commodity/E. Accessed 14th June 2015.

Fakorede, M. A. B., Alofe, C. O., Adenola, O. A., Kim, S. K., Nweke, F. I., Iken, J. E., and Akinwumi, J. A. 1997. Research priorities for the improvement of maize in Nigeria: farmers ' views. Pages 382 - 387 *in* Contributing to food self-sufficiency: maize research and development in West and Central Africa, edited by B. Badu-Apraku, M. O. Akoroda., M. A. B. Fakorede, M. Ouedraogo and R. J. Carsky. Proceedings of a regional maize workshop., 4 - 7 May 1999, IITA, Cotonou, Benin.

Fakorede, M. A. B. 2002. Adoption of recommended maize production technologies in Nigeria: a farm-level survey. (Report in progress).

Faria, N. R., Suchard, M. A., Rambaut, A., and Lemey, P. 2011. Towards a quantitative understanding of viral phylogeography. *Current Opinion in Virology*, 1 (15): 423 - 429. doi:10.1016/j.coviro.2011.10.003

Faria, N. R., Hodges-Mameletzis, I., Silva, J. C., Erasmus, S., Paolucci, S., Ruelle, J., Pieniazek, D., Taveira, N., Trevino, A., Goncalves, M. F., Jallow, S., Xu, L., Camacho, R. J., Soriano, V., Goubau, P., de Sousa, J. D., Vandamme, A-M., Suchard, M. A., and Lemey, P. 2012. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. *Journal of General Virology*, 93: 889 - 899. doi:10.1099/vir.0.038638-0

Faria, N. R., Suchard, M. A., Rambaut, A., Streicker, D. G., and Lemey, P. 2013. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philosophical Transactions of the Royal Society B*, 368 (1614): 20120196. http://dx.doi.org/10.1098/rstb.2012.0196

Fennah, R. G. 1960. A new species of *Cicadulina* (*Homoptera*: *Cicadelidae*) from East Africa. *Annual Magazine of Natural History*, 2: 757 - 758.

Famine Early Warning Systems (FEWS) Network Zambia. 2001. Monthly Report: October - November 2001. pp 1 - 4. http://www.fews.net

Frischmuth, T., and Stanley, J., 1998. Recombination between viral DNA and the transgenic coat protein gene of African cassava mosaic geminivirus. *Journal of General Virology*, 79: 1265 - 1271.

Frost, S. D. W., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoefffer, S., and Bedford, T. 2015. Eight challenges in phylodynamic inference. *Elsevier Epidemics*, 10: 88 - 92.

Fuller, C. 1901. Mealie variegation.First entomology report of the Government of Natal (South Africa) 1899-1900, p. 17-19. P. Davis and Sons, Government Printers. Pietermaritzburg, South Africa.

Gelaw, B. 1985. CIMMYT's maize improvement role in East, Central and Southern Africa. *In:* To Feed Ourselves. A Proceedings of the First Eastern, Central and Southern Africa Regional Maize Workshop, Lusaka, Zambia, March 10-17 1985. *pp208-228*. Downloaded from http://ufdc.ufl.edu//uf00080089/00001 on 3 February 2015.

Gelman, A., Meng, X. L. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat Sci*, 13: 163 - 185.

Ghauri, M. S. K. 1961. A new East African *Cicadulina* [*Cicadellidae: Homoptera*) resembling *C. zeae* China infesting maize in Southern Africa. *Ann. Mag. Nat. Hist* (Ser 13) 4: 369 - 370.

Ghauri, M. S. K. 1964. A new species of *Cicadulina* China (*Homoptera: Cicadelloidea*) from Kenya. *Ann. Mag. Nat. Hist*, 13 (7): 205 - 208.

Ghauri, M. S. K. 1971. A remarkable new species of *Cicadulina* China (*Homoptera: Cicadelloidea*) from East Africa. Bull. *Ent. Res*, 60: 631 - 633.

Goodman, R. M. 1981. Geminiviruses. In *Handbook of Plant Virus Infections*, pp. 879 - 910. Edited by E. Kurstak. Amsterdam: Elsevier/North-Holland.

Google Earth: available at https://earth.google.com

Grenfell, B. T. *et al*. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303: 327 - 332.

Griffiths, R. C., and Tavares, S. 1994. Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society London B Biological Sciences* 344: 403 - 410.

Guthrie, E. J. 1976. Virus diseases of maize in East Africa. In: L. E. Williams D. T. Gordon and N. R. Nault (ed.) *International Maize Virus Disease Colloquium and Workshop.* Aug 16-19, 1976, Wooster, OH.

Guthrie, E. J. 1977. Virus diseases of maize in East Africa. In: Proceedings of the International Maize Virus Disease Colloquium and Workshop. Wooster, Ohio, USA: Ohio Agricultural Research and Development centre, 62 - 68.

Guthrie, E. J. 1978. Measurement of yield losses caused by maize streak disease. *Plant Disease Reporter*, 62 (10): 839 - 841.

Gutierrez, C. 2000. DNA replication and cell cycle in plants: learning from geminiviruses. *EMBO Journal*, 19 (5): 792 - 799.

Hanson, S. F., Hoogstraten, R. A., Ahlquist, P., Gilbertson, R. L., David, R., Russell, D., and Maxwell, P. 1995. Mutational Analysis of a Putative NTP-Binding Domain in the Replication-Associated Protein (AC1) of *Bean Golden Mosaic Geminivirus*. *Virology*, 211 (1): 1 - 9.

Harkins, G. W., Martin, D .P., Duffy, S., Monjane, A. L., Shepherd, D. N., Windram, O. P., Owor, B. E., Donaldson, L., van Antwerpen, T., Sayed, R. A., Flett, B., Ramusi, M., Rybicki, E. P., Peterschmitt, M., and 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.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., and Gronenborn, B. 1995. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Research*, 23 : 910 - 916.

Hilbeck, A., and Andow, D. A. 2004. Environmental Risk Assessment of Genetically Modified Organisms. Volume 1. A Case Study of *Bt* Maize in Kenya. CAB International, Wallingford, UK. ISBN 0 85199 8615. pp 36.

Ho, S. Y. W., and Duchene, D. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23: 5947 - 5975.

Hofer, J. M. M. I., *et al*. 1992. Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *Plant Cell*, 4: 213 - 223.

Hopkins, J. C. F. 1936. Annual report of the Senior Plant Pathologist for the year ending 31st December, 1936. *Rhodesia Agricultural Journal*, 33 (6): 413 - 421.

Hopkins, A. G., 1973. An Economic History of West Africa. London: Longman. doi:10.2307/3601532

Horvath, G. V., Pettko-Szandtner, A., Nikovics, K., Bilgin, M., Boulton, M., Davies, J. W., Gutierrez, C., and Dudits, D. 1998. Prediction of functional regions of the *maize streak virus* replication-associated proteins by protein-protein interaction analysis. *Plant Molecular Biology*, 38: 699 - 712.

Howell, S. H. 1984. Physical structure and genetic organization of the genome of maize streak virus (Kenyan isolate). *Nucleic Acids Research*, 12: 7359 - 7375.

Hughes, F. L., Rybicki., E. P., Wechmar, M. Bvon. 1992. Genome typing of southern African subgroup 1 geminiviruses. *Journal of General Virology*, 73 (5): 1031 - 1040.

Hughes, J. d'A., and Odu, B.O. 2003. An Overview of Maize viruses in Sub-Saharan Africa. In: Plant Virology in Sub-Saharan Africa: Proceeding of a Conference Organized by IITA, 4-8 June 2001, International Institute of Tropical Agriculture, Ibadan, Nigeria pp. 162

ICTVdb Management. 2006. http://www.ncbi.nlm.gov/ICTVdB/ICTV/dB/29010006.htm. Accessed 17th October 2014.

Ingvarsson, P. K. 2008. Molecular evolution of codon usage in *Populus*. *BMC Evolutionary Biology*, 8: 307. doi:10.1186/1471-2148-8-307

Iliffe, J. 1987. The African Poor: A History. Cambridge University Press. pp. 159. ISBN 0 521 344 158.

International Institute of tropical Agriculture (IITA). 1984. IITA Research Highlights 1983. Ibadan, Nigeria, pp. 124.

International Institute of Tropical Agriculture (IITA). 1986. Solving the problem of *maize streak virus*: a breakthrough to increase maize production in sub-Saharan Africa. Submission for the King Baudouin award 1986. IITA, Ibadan, Nigeria.

International Institute of Tropical Agriculture (IITA). 2013. New Drought Tolerant Maize Varieties for Ghana. In: Drought Tolerant Maize for Africa: *A Quarterly Bulletin of the Drought Tolerant Maize for Africa Project*, 2 (1): 1 - 4.

Improved Alignment Editor (IMPALE): available at http://www.cbio.uct.ac.za/~arjun/IMPALE

IPPC-Secretariat. 2005. Identification of risks and management of invasive alien species using the IPPC framework. Proceedings of the workshop on invasive and the International Plant Protection Convention, 22-26 September 2003. xii + 301 pp.

Janes, G. R., Hutwanger, L., Cates, Jr W., Stroup, D. F., and Williamson, G. D. 2000. Descriptive Epidemiology: Analyzing and Interpreting Surveillance Data. In: Teutsch S. M., Churchill, R. E., eds. *Principles and Practice of Public Health Surveillance*. New York, NY: Oxford University Press, Inc. pp. 112 - 167.

Jeffreys, M. D. W. 1963. How ancient is West African Maize? *J. Int. African Inst*, 33: 115 - 131.

Jeske, H., Lutgenmeier, M., and Preiss, W. 2001. Distinct DNA forms indicate rolling circle and recombination-dependent replication of *Abutilon mosaic virus*. *EMBO Journal*, 20: 6158 - 6167.

Kaitisha, G. C. 2003. Some virus diseases of crop plants in Zambia. In: Hughes J d'A, Odu BO (eds) Plant Virology in Sub-Saharan Africa. Proceedings of a conference organized by IITA. 4-8 June 2001, International Institute of Tropical Agriculture, Ibadan, Nigeria.

Karavina, C. 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. doi:10.5897/AJAR2014.8897

Kass, R. E., and Raftery, A. E. 1995. Bayes factors. *Journal of the American Statistical Association*, 90: 773 - 795.

Khavrus, V., and Shelevytsky, I. 2010. Introduction to the solar motion geometry on the basis of a simple model. *Physics Education*, 45 (6): 641. doi:10.1088/0031-9120/45/6/010

Khavrus, V., and Shelevytsky, I. 2012. geometry and the physics of seasons. *Physics Education*, 47 (6): 680. doi:10.1088/0031-9120/47/6/680.

Kim, S. K., Efron, Y., Fajemisin, J. M., and Buddenhaggen, I. W. 1989. Mode of gene action for resistance in maize to maize *streak virus*. *Crop Science*, 29 (4): 890 - 894.

Kingman, J. F. C. 1982a. The coalescent. *Stochastic Processes and Their Applications*, 13: 235 - 248.

Kishino, H., Thorne, J. L., and Bruno, W. J. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology of Evolution*, 29: 170 - 179.

Konate, G., and Traore, O. 1992. Reservoir hosts of maize streak virus (MSV) in the Sudan-Sahel zone: identification and spatiotemporal distribution. Phytoprotection 73 (3): 111 – 117.

Kosakovsky Pond, S. L., Frost, S. D., and Muse, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21 (5): 676 - 679.

Kosakovsky Pond, S. L., and Frost, S. D.W. 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, 22 (5): 1208 - 1222. doi:10.1093/molbev/msi105

Kosakovsky Pond, S. L., Frost, S. D. W., Grossman, Z., Gravenor, M. B., Richman, D. D., and Leigh Brown, A. J. 2006a. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comparative Biology*, 23: 2993 - 3000.

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. 2006b. GARD: a genetic algorithm for recombination. *Bioinformatics*, 22: 3096 - 3098.

Kuhnert, D., Wu, C. H., Drummond, A. J. 2011. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infection Genetics and Evolution*, 11: 1825 - 1841.

Kuzmina, N. A., Lemey, P., Kuzmin, I. V., Mayes, B. C., Ellison, J. A., Orciari, L. A., Hightower, D., Taylor, S. T., Rupprecht, C. E. 2013. The Phylogeography and Spatiotemporal Spread of South-Central Skunk Rabies Virus. *PLoS ONE*, 8 (12): e82348. doi:10.1371/journal.pone.0082348

Kyetere, D. T., Ming, R., Mcmullen, M., Pratt, R. C., Brewbaker, J., Musket, T., Pixley, K., and Moon, H. 1995. Monogenic tolerance to *maize streak virus* maps to the short arm of chromosome 1. *Maize Genet. Coop. News Lett*, 69: 136 - 137.

Kyetere, D. T., Ming, R., Mcmullen, M. D., Pratt, R. C., Brewbaker, J., and Musket, T. 1999.Genetic analysis of tolerance to maize *streak virus* in maize. *Genome*, 42: 20 - 26.

Lagat, M., Danson, J., Kimani, M., Kuria, A. 2008. Quantitative trait loci for resistance to maize streak virus disease in maize genotypes used in hybrid development. *African Journal of Biotechnology*, 7: 2573 - 2577.

Lartillot, N., and Philippe, H. 2006. Determining Bayes factors using thermodynamic integration. *Systematic Biology*, 55: 195 - 207.

Larson, D. W., and Mbowa, S. 2004. Strategic marketing Problems in the Uganda Maize Seed Industry. *International Food and Agribusiness Management Review*, 7: 86 - 93.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. 1995a. Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. *Biochimie*, 77 (10): 765 - 773. doi:10. 1016/0300-9084(96)88194-6

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rodgers, S. G., Schell, J., and Gronenborn, B. 1995b. *In-vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceeding National Academic Science*, USA 92: 3879 - 3883.

Lazarowitz, S. G. 1987. The molecular characterization of geminiviruses. *Plant Molecular Biology Reporter*, 4: 177 - 192.

Lazarowitz, S. G. 1988. Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus. *Nucleic Acids Research*, 16: 229 - 249

Lazarowitz, S. G., Pinder, A. J., Damsteegt, V. D., and Rogers, S. G. 1989. *Maize streak virus* genes essential for systemic spread and symptom development. *EMBO Journal* 8: 1023 - 1032.

Lefeuvre, P., Lett, J-M., Varsani, A., Martin, D. P. 2009. Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses. *Journal of General Virology*, 83 (6): 2697 - 2707. doi:10.1128/JVI.02152-08.

Lefeuvre, P., Martin, D. P., Harkins., G. W., Lemey, P., Gray, A. J. A., Meredith, S., Lakay, F., Monjane, A., Lett, J-M., Varsani, A., and Heydarnejad, J. 2010. The Spread of Tomato Yellow Leaf Curl Virus from the Middle East to the World. *PLoS Pathogens*, 6 (10): e1001164. doi:10.1371/journal.ppat.1001164

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

Leke, W. N., Njualem, D. K., Nchinda, V. P., Ngoko, Z., Zok, S., Ngeve, J., Brown., J. K., and Kvarnheden, A. 2009. Molecular identification of *Maize streak virus* reveals the first evidence for a subtype A1 isolate infecting maize in Cameroon. *New Disease Reports*, 19: pp 36.

Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, A. M. 2009. Bayesian phylogeography finds its roots. *PLoS Computational Biology* 5 (9): e1000520. doi:10.1371/journal.pcbi.1000520

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

Lemey, P., Rambaut, A., Bedford, Faria, N. R., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D. and Suchard, M. A. 2012. The

seasonal flight of influenza: a unified framework for spatiotemporal hypothesis testing. *arXiv*: 1210.5877.

Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., and Suchard, M. A. 2014. Unifying Viral Genetics and Human Transportation Data To predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS Pathogens*, 10 (2): e1003932. doi:10.1371/journal.ppat.1003932

Lett, J.-M., Granier, M., Hippolyte, I., Grondin, M., Royer, M., Blanc, S., Reynaud, B., and Peterschmitt, M. 2002. Spatial and temporal distribution of geminiviruses in leafhoppers of the genus *Cicadulina* monitored by conventional and quantitative polymerase chain reaction. *Phytopathology*, 92: 65 - 74.

Liu, H., Boulton, M. I., Davies, J.W. 1997. Maize streak virus coat protein binds single- and double-stranded DNA in vitro. *Journal of General Virology*, 78: 1265 - 1270.

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W., and Stanley, J. 1999. Been yellow dwarf virus RepA, but not Rep, binds to maize retinoblastoma protein and the virus tolerates mutations in the consensus binding motif. *Virology*, 256: 270 - 279.

Liu, H., Boulton, M. I., Oparka, K. J., Davies, J. W. 2001. Interaction of the coat proteins of Maize *streak virus*: implications for the transport of viral DNA. *Journal of General Virology*, 82: 35 - 44.

Mafura, K. 2015. Lesotho national Action Programme in Natural Resource Management, Combating Desertification, and Mitigating the Effects of Drought. pp 1 - 2.

Magenya, O. E. V., Mueke, J., and 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.

Magee, D., Beard, R., Suchard, M. A., Lemey, P., and Scotch, M. 2015. Combining Phylogeography and Spatial Epidemiology to Uncover Predictors of H5N1 Diffusion. *Archives of Virology*, 160 (1): 215 - 224. doi:10.1007/s00705-014-2262-5

Makoni, R. D. 2000. Effects of the economic structural adjustment programme (1991-1993) on the participation of secondary school girls in Zimbabwe.

Malik, P. S., Kumar, V., Bagewadi, B., Mukherjee, S. K. 2005. Interaction between coat protein and replication initiation protein of *Mung bean yellow mosaic India virus* might lead to control of viral DNA replication. *Virology*, 337: 273 - 283.

Manyong, V. M., Kling, J. G., Makinde, K. O., Ajala, S. O., and Menkir, A. 2000. Impact of IITA improved germplasm on maize production in West and Central Africa. Impact document. IITA, Ibadan, Nigeria.

Maphosa, B. 1994. Lessons from the 1992 Drought in Zimbabwe: The Quest for Alternative Food Policies. *Nordic Journal of African Studies*, 3(1): 53 - 58.

Martin, D, P., and Rybicki, E. P. 1998. Microcomputer-based quantification of *maize streak virus* symptoms in *Zea mays*. *Phytopathology*, 88: 422 - 427.

Martin, D, P., Willment, J. A., and Rybicki, E. P. 1999. Evaluation of *Maize streak virus* pathogenicity in differentially resistant *Zea mays* genotypes. *Phytopathology*, 89 (8): 695 - 700.

Martin, D, P., Rybicki, E. P. 2000. Improved efficiency of *Zea mays* agroinoculation with *Maize streak virus*. *Plant Disease*, 84: 1096 - 1098.

Martin, D. P., Willment, J. A., Bilharz, R., Velders, R., Odhiambo, B., Njuguna, J., James, D., and Rybicki, E. P. 2001. Sequence diversity and virulence in *Zea mays* of *Maize streak virus* isolates. *Virology*, 288 (2): 247 - 255.

Martin, D. P., and Rybicki, E. P. 2002. Investigation of *Maize streak virus* pathogenicity determinants using chimaeric genomes. *Virology*, 300: 180 - 188.

Martin, D. P., and Shepherd, D. N. 2009. Review: The epidemiology, economic impact and control of maize streak disease. *Food Security* 1(3): pp. 305 - 315.

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

Masih, I., Maskey, S., Mussa, F.E.F., and Trambauer, P. 2014. A review of droughts on the African continent: a geospatial and long-term perspective. *Hydrology and Earth System Sciences*, 18: 3635 - 3649. doi:10.5194/hess-18-3635-2014

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J., Buckler, E., and Doebley, J. F. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of National Academy of Science*, 99: 6080 - 6084.

McCann, J. 2001. Maize and grace: history, corn and Africa's new landscapes, 1500-1999. Comp. *Stud. Soc. Hist*, 43: 246 - 272.

McClean, A. P. D. 1947. Some forms of *streak virus* occur in maize, sugarcane and wild grasses. Pretoria: South African Government Printers. *Union of South African Department of Agriculture and Forestry Science Bulletin*, 265: 1 - 39.

McKinney, H. H. 1929. Mosaic diseases in the Canary Islands, West Africa and Gibraltar. *Journal of Agricultural Research*, 39: 557 - 578.

Mesfin, T., Den Hollander, J., and Markham, P. G. 1991.*Cicadulina* species and *Maize streak virus* in Ethiopia. *Tropical Pest Management*, 37 (3): 240 - 244.

Meo, S. A., Al Masri, A. A., Usmani, A. M., Memon, A. N., and Zaidi, S. Z. 2013. Impact of GDP, Spending on R and D, Number of Universities and Scientific Journals on Research Publications among Asian Countries. *PLoS ONE*, 8(6): e66449. doi:10.1371/journal.pone.0066449

Monjane, A. L., Harkins, G. W., Martin, D. P., Lemey, P., Lefeuvre, P., Shepherd, D. N., Oluwafemi, S., Simuyandi, M., Zinga, I., Komba, E. K., Lakoutene, D. P., Mandakombo, N., Mboukoulida, J., Semballa, S., Tagne, A., Tiendrebeogo, F., Erdmann, J. B., van Antwerpen, T., Owor, B. E., Flett, B., Ramusi, M., Windram, O. P., Syed, R., Lett, J-M., Briddon, R. W., Markham, P. G., Rybicki, E. P., and 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.

Morato, D. G., Barreto, F. R., Braga, J. U., Natividade, M. S., Costa, M. da' C. N., Morato, V., Teixeria, M. da' G. L. C. 2015. The spatiotemporal trajectory of dengue epidemic in a medium-sized city. *Mem Inst Oswaldo Cruz*, Rio de Janeiro, 110 (4): 528 - 533. doi: 10.1590/0074-0276140388

Morris, M. L., Tripp, R., and Dankyi, A. A. 1999. Adoption and impacts of improved maize production technology: A case study of Ghana grains development project. *Economics Program Paper* 99-01., Mexico, D. F: CIMMYT ISSN: 1405 - 7735.

Muhire, B. M., Varsani, A., and Martin, D. P. 2014. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* 9(9): e108277. doi: 10.1371/journal.pone.0108277

Mullineaux, P. M., Donson, J., Morris-Krisnich, B. A., Boulton, M. I., and Davies, J. W. 1984. The nucleotide sequence of *Maize streak virus* DNA. *EMBO Journal*, 3: 3063 - 3068.

Mullineaux, P. M., Boulton, M. I., Bowyer, P., van der Vlugt, R., Marks, M., Donson, J., and Davies, J. W. 1988. Detection of a non-structural protein of MW 11 000 encoded by the virion DNA of maize streak virus. *Plant Molecular Biology*, 11: 57 - 66.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S., and Scheffler, K. 2013. FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Molecular Biology Evolution*, 30: 1196 - 1205.

Mwangi, S. F. M. 1998. Status of Northern Leaf Blight, *Phaeosphaeria maydis* Leaf Spot, Southern leaf blight, Rust, *Maize streak virus* and physiologic specialization of *Exserohilum turcium* in Kenya. Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University. Blacksburg. Virginia. http://scholar.lib.vt.edu/theses/available/etd-02199-125602. Accessed 6th October 2015.

Mylonas, P., Yonow, T., and Kriticos, D. J. 2014.*Cicadulina mbila* (*Naude*). *Harvest Choice Pest Geography*, St. Paul, MN: InSTePP-HarvestChoice.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Iban˜ez, J. T., and Hanley-Bowdoin, L. 2011. Functional Analysis of a Novel Motif Conserved across Geminivirus Rep Proteins. *Journal of Virology*, 85 (3): 1182 - 1192.

Nicholson, I. F. The administration of Nigeria, 1900-1961. *Oxford*, Clarendon Press, 1969. In: The Transfer of Agricultural Techniques. Yudelman, M. pp. 338 - 350.

News24. 2001. Zambia names winners of maize tender. http://www.news24.com/xArchive/Archive/Zambia-names-winners-of-maize-tender-20011026.

Newton, M. A., and Raftery, A. E. 1994. Approximate Bayesian Inference by Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society Ser. B*, 56: 3 - 48.

Ngoko, Z., Cardwell, K. F., Marasas, W. F. O., Wingfield, M. J., Ndemah, R., Schulthess, F. 2001. Biological and physical constraints on maize production in the Humid Forest and Western Highlands of Cameroon. *European Journal of Plant Pathology*, 108: pp. 893 - 902.

Nunes, E., Sousa, D., and Sataric, I. 1985. Research on the constraints to maize production in Mozambique. In: To Feed Ourselves: A Proceedings of the first east, central, and southern Africa regional Maize Workshop, Lusaka, Zambia, March 10 - 17, 1985. pp. 67 - 69

Nunes, M. R. T., Palacios, G., Faria, N. R., Sousa, E. C. Jnr., Pantoja, J. A., Rodrigues, S. G., Carvalho, V. L., Medeiros, D. B. A., Savji, N., Baele, G., Suchard, M. A., Lemey, P., Vasconcelos, F. C., Lipkin, W. I. 2014. Air Travel Is Associated with Intracontinental Spread of Dengue Virus Serotypes 1-3 in Brazil. *PLoS Neglected Tropical Diseases* 8(4): e2769. doi:10.1371/journal.pntd.0002769

Ogata, Y. 1989. Monte Carlo method for high dimensional integration. *Numer Math*, 55: 137 - 157

Okoth, V. A. O., and Dabrowski, Z. T. 1987.Population density, species composition and infectivity with *Maize streak virus* (MSV) of *Cicadulina* spp. leafhoppers in some ecological zones in Nigeria. *Acta Oecologica/Oecologia Applicata* 8 (3): 191-200.

Okoth, V. A. O., Dabrowski, Z. T., Thottapilly, G., and Emden, H. F. 1987. Comparative analysis of some parameters affecting Maize *streak virus* (MSV) transmission of various *Cicadulina* spp. populations. *Insect Science and its Application*, 8(3): 295 - 300.

Oluwafemi, S., Jackai, L. E. N., and Alegbejo, M. D. 2007. Comparison of transmission abilities of four *Cicadulina* species vectors of maize *streak virus* from Nigeria. *Entomologia Experimentalis et Applicata*, 124 (3): 235 - 239.

Oluwafemi, S., Kraberger, S., Shepherd, D. N., Martin, D. P., and Varsani, A. 2014. A high degree of African streak virus diversity within Nigerian maize fields includes anew *mastrevirus* from *Axonopus compressus*. *Archives of Virology*, 159 (10): 2765 - 70. doi:10.1007/s00705-014-2090-7.

Omede, A. J. 2006. Nigeria's Relations with Her Neighbours. *Stud. Tribes Tribals*, 4(1): 7 - 17.

Oppong, A., Offei, S. K., Adu-Dapaah, H., Lamptey, J. N., Kurenbach, B., Walters, M., Shepherd, D. N., Martin, D. P., Varsani, A. 2015. Mapping the distribution of *maize streak virus* genotypes across the forest and transition zones of Ghana. *Archives of Virology*, 160 (2): 483 - 492. doi: 10. 1007/s00705-014-2260-7.

Organization for Economic Cooperation and Development (OECD). 2001. Prospects for Trade Between Nigeria and Its Neighbours. Drafted by Soule, B. G., and Obi, C. OECD Publications 2. pp 7 - 41.

Orozco, B. M., and Hanley-Bowdoin, L. 1998. Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *Journal of Biological Chemistry*, 273: 24448 - 24456.

Owor, B. 2008. *Maize streak virus* (MSV) diversity in Uganda and the assessment of gene silencing as a tool for the development of resistance to MSV. PhD Thesis. University of Cape Town.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thompson, J. A., Martin, D. P., and Varsani, A. 2007. Successful application of FTA Classic Card technology and use of bacteriophage Q29 DNA polymerase for large

scale field sampling and cloning of complete *Maize streak virus* genomes. *Journal of Virology Methods*, 140: 100 - 105.

Padidam, M., Sawyer, S., and Fauquet, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265: 218 - 225.

Pagel, M., Meade, A., and Barker, D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53: 673 - 684.

Palmer, K. E., and Rybicki, E. P. 1998. The Molecular Biology of *Mastreviruses*. *Advances in Virus Research*, 50: 183 - 234.

Pande, D., Kraberger, S., Lefeuvre, P., Lett, J.-M., Shepherd, D., Varsani, A., and Martin, D. 2012. A novel maize infecting mastrevirus from La Reunion Island. *Archives of Virology*, 157: 1617 - 1621.

Peterschmitt, M., Quiot, J. B., Reynaud, B. 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: 641 - 653.

Peterschmitt, M., Granier, M., Frutos, R., and Reynaud, B. 1996. Infectivity and complete nucleotide sequence of the genome of a genetically distinct strain of maize streak virus from Reunion Island. *Archives of Virology*, 141: 1637 - 1650.

Phimister, I. R. 1991. Secondary industrialization in Southern Africa: the 1948 customs agreement between Southern Rhodesia and South Africa. *Journal of Southern African Studies*, 17 (3): 430 - 443.

Pingali, P. L. 2001.CIMMYT 1999-2000 World maize facts and trends. Meeting world maize needs: Technological opportunities and priorities for the public sector. Mexico, d.F. CIMMYT.

Pinner, M. S., Markham, P. G., Markham, R. H., Dekker, E. L. 1988. Characterization of *Maize streak virus*: description of strains; symptoms. *Plant Pathology*, 37 (1): 74 - 87.

Piperno, D. R., and Flannery, K. V. 2001. The earliest archeological maize (*Zea mays* L.) from highland Mexico: New accelerator mass spectrometry dates and their implications. *Proceedings of National Academy of Science*, 98: 2101 - 2103.

Potts, D. H. 2010. Circular Migration in Zimbabwe and Contemporary Sub-Saharan Africa. *James Curry*, Woodbridge.UK. pp. 149 - 150.

Pulliam, H.R., and Danielson, B.J. 1991. Sources, sinks and habitat selection: a landscape perspective on population dynamics. *Am. Nat*, 137: s50 – s66.

Pulliam, R. 1988. Sources, sinks and population regulation. *Am. Nat*, 132: 652 – 661.

Pratt, R. C., and Gordon, S. G. 2006. Breeding for resistance to maize foliar pathogens. *Plant Breed rev*, 27: 119 - 174.

Prasanna, B. M .P., Vasal, S, S. K., Kassahun, B., and Singh, N. N. 2001. Quality protein maize. *Current Science*, 81: 1308 - 1319.

Preiss, W., and Jeske, H. 2003. Multitasking in replication is common among geminiviruses. *Journal of Virology*, 77: 2972 - 2980.

Pybus, O. G., and Rambaut, A. 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews*, 10: 540 - 550.

Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. A., Rambaut, A., Crawford, F. W., Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P., Delwart, E. L. 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *PNAS*, www.pnas.org/cgi/doi/10.1073/pnas.1206598109

Rambaut, A., Suchard, M., and Drummond, A. 2009. FigTree version 1.4.2: available at http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A., Suchard, M., and Drummond, A. 2009. Tracer version 1.6: available at http://tree.bio.ed.ac.uk/software/tracer/.

Rambaut, A., Suchard, M., and Drummond, A. 2009. *Path-O-Gen* version 1.4: Temporal Signal Investigation Tool. Institute of Evolutionary Biology, University of Edinburg. Available at: http://treebio.ed.ac.uk/software/pathogen/.

Ramsden, C., Melo, F. L., Figueiredo, L M., Holmes, E. C., Zanoto, P. M . A. 2008. High rates of molecular evolution in hantaviruses. *Molecular Biology Evolution*, 25: 1488 - 1492.

Republic of Kenya. 2004. National Policy on Disaster Management, (Revised Draft) p4. In: Kenya Natural Disaster Profile by de la Porte, P. A. United Nations Development Program, Enhanced Security Unit. Nairobi, Kenya.

Reynaud, B., Delatte, H., Peterschmitt, M., and Fargette, D. 2009. Effects of temperature increase on the epidemiology of three major vector-borne viruses. *European Journal of Plant Pathology*, 123(3): 269 - 280.

Rodier, A., Assie, J., Marchand, J-L., and Herve, Y. 1995. Breeding maize lines for complete and partial resistance to Maize *streak virus* (MSV). *Euphytica*, 81: 57 - 70.

Rose, D. J. W. 1972. Times and sizes of dispersal flight by *Cicadulina* spp. vectors of maize *streak virus* disease. *Journal of Animal Ecology*, 41: 494 - 506.

Rose, D. J. W. 1973. Management of Cicadulina leaf hopper populations to reduce streak disease in maize crops in the Highveld in Rhodesia. *Rhodesia Agricultural Journal*, 70 (3): 63 - 64.

Rose, D. J. W. 1973a. Laboratory observations on the biology of *Cicadulina* spp. (Hem., Cicadellidae), with particular reference to the effects of temperature. *Bulletin of Entomological Research*, 62 (3): 471 - 476.

Rose, D. J. W. 1974. The Epidemiology of Maize streak disease in relation to population densities of *Cicadulina* spp. *Annals of Applied Biology*, 76: 199 - 207.

Rose, D. J. W. 1978. Epidemiology of maize streak virus disease: Virus spread by several species of leafhoppers, *Cicadulina*. *Annual Review of Entomology*, 23: 259 - 282.

Rosegrant, H. W. 2008. Biofuels and grain prices: Impacts and policy responses. *International Food Policy Research Institute* (IFPRI), pp 2 - 4.

Rossel, H. W., and Thottapilly, G. 1985. Virus diseases of important food crops in Tropical Africa. Ibadan, Nigeria: International Institute of tropical Agriculture (IITA), pp. 61

Rubaihayo, E., Kaahwa, E. R., and Kabeere, F. 1985. "Maize Research and Seed production in Uganda." *in* Proceedings of the first Eastern, Central and Southern Africa Regional Maize Workshop, Lusaka, Zambia, March 10 - 17. Government of Zambia and CIMMYT.

Ruppel, R. F. 1965. A review of the genus *Cicadulina* (Hemiptera, Cicadellidae). Publications of the Museum, Michigan State University. *Biological Series*, 2 (8): 385 - 428.

Saunders, K., Lucy, A., and Stanley, J. 1991. DNA forms of geminivirus African cassava mosaic virus consistent with rolling circle mechanism of replication. *Nucleic Acids Research*, 19: 2325 - 2330.

Selection Map version 1.0. Available at: http://www.cbio.uct.ac.za/~brejnev/ComputationalTools.html

Settlage, S. B., Miller, Band Hanley-Bowdoin, L. 1996. Interactions between geminivirus replication proteins. *Journal of Virology*, 70: 6790 - 6795.

Scallan, E., Griffin, P. M., Angulo, F. J., Tauxe, R. V., and Hoekstra, R. M. 2011a. Foodborne illness acquired in the United States - unspecified agents. *Emerging Infectious Diseases*, 17: 16 - 22.

Scheffler, K., Martin, D. P., Seoighe, C. 2006. Robust inference of positive selection from recombining coding sequences. *Bioinformatics*, 22: 2493 - 2499.

Schierup, M. H., and Hein, J. 2000a. Consequences of Recombination on Traditional Phylogenetic Analysis. *Genetics*, 156: 879 - 891.

Schierup, M. H., Hein, J. 2000b. Recombination and the Molecular Clock. *Molecular Biology and Evolution*, 17: 1578 - 1579.

Schluter, D., Price, T., Mooers, A., and Ludwig, D. 1997. Likelihood of ancestor states in adaptive radiation. *Int J. Org Evol*, 51: 1699 - 1711.

Schnippenkoetter, W. H., Martin, D. P., Hughes, F., Fyvie, M., Willment, J. A., James, D., von Wechmar, B., and Rybicki, E. P. 2001.The biological and genomic characterization of three *mastreviruses*. *Archives of Virology*, 146: 1075 - 1088.

Shepherd, D. N., Martin, D. P., Van Der Walt, E., Dent, K., Varsani, A., and Rybicki, E. P. 2010. *Maize streak virus:* an old and complex 'emerging' pathogen. *Molecular Plant Pathology*, 11(1): 1 - 12.

Shepherd, D. N., varsani, A., Windram, O., Lefevre, P., Monjane, A. L., Owor, B., Martin, D. P. 2008. Novel *sugarcane streak Reunion virus*-like *Mastrevirus* isolates from Southern Africa and *La* Reunion. *Archives of Virology*, 153: 605 - 609.

Sitko, N. J., Kuteya, A. N., and Chisanga, B. 2014. Analysis of the Effects of maize Trade Restrictions in the COMESA Region on Food Prices and Market Development. *Indaba Agricultural Policy Research Institute*, (IAPRI).

Sleper, A. D., and Poehlman, M. J. 2006. Breeding field crops. 5th Edition. *Blackwell* Publishing, pp. 277 - 296.

Soto P. E. A. 1978.A new vector of maize streak virus. *East Afric Agric For J*, 44:70 - 71.

Spatial Phylogeographic Reconstruction of Evolutionary Dynamics (SPREAD) version 1.0.6: available at: http://www.phylogeography.org/SPREAD

Stack, J. C., Welch, J. D., Ferrari, M. J., Shapiro, B. U., Grenfell, B. T. 2010. Protocols for sampling viral sequences to study epidemic dynamics. *J. R. Soc Interface* 7: 1119 - 1127.

Stenger, D. C., Revington, G. N., Stevenson, M. C., Bisaro, D. M. 1991. Replication release of geminivirus genomes from tandemly repeated copies: evidence for rolling circle replication of a plant viral DNA. *Proceeding of Natural Academic Science, USA*, 88: 8029 - 8033.

Storey, H. H. 1925.The transmission of streak disease of maize by the leafhopper *balclutha mbila Naude*. *Annals of Applied Biology*, 12: 422 - 439.

Storey, H. H. 1928.Transmission of maize streak virus. *Annals of Applied Biology*, 15: 1 – 25.

Storey, H. H., and McClean, A. P. D. 1930. The transmission of streak disease between maize, sugarcane and wild grasses. *Annals Applied Biology*, 17: 691-719.

Storey, H. H. 1936. Virus diseases of East African Plants. V. Streak disease of maize. *East African Agricultural Journal*, 1: 471 - 475.

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 B* 125: 455 - 479.

Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology Evolution*, 18: 1001 - 1013.

Sylvester, E. S., Richardson, J., Nickel, J. L. 1973. An additional note on virus-like particles associated with maize streak disease. Plant Disease Reporter 57 (5): 414 - 416.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2013. MEGA version 6.06: molecular evolutionary genetic analyses using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology Evolution*, 28: 2731 - 2739.

Thommes, P., Osman, T.A.M., Hayes, R. J., and Buck, K. W. 1993. TGMV replication protein AL 1 preferentially binds to single-stranded DNA from the common region. *FEBS Lett*, 319: 95 - 99.

Thottappilly, G., Bosque-Perez, N. A., and Rossel, H. W. 1993. Viruses and virus diseases of maize in tropical Africa. *Plant Physiology*, 42(4): 494 - 509. doi:10.1111/j.1365-3059.1993.tb01529.x

Torok, M., and Anderson, M. 2008. Introduction to Public Health Surveillance. *Focus on Field Epidemiology*, *5* (5): 1 - 5.

Townsend, R., Stanley, J., Curson, S. J., Short., M. N. 1985. Major polyadenylated transcripts of *cassava latent virus* and location of the gene encoding the coat protein. *EMBO Journal*, 4: 33 - 37.

Trifonov, V., and Khiabanian, H., and Rabadan, R. 2009. Geographical dependence, surveillance and origins of the 2009 influenza A (H1N1) virus. 361 (2): 115 - 119.

Ubeda, F., and Wilkins, J. F. 2011. The Red Queen theory of recombination hotspots. *Journal of Evolutionary Biology*, 24: 541 - 553.

Unganai, L. S. 1994. Drought and Southern Africa: A Note from the Harare Regional Drought Monitoring Centre. *Drought Network News*.

Unseld, S., Hohnle, M., Ringel, M., Frismuth, T. 2001.Subcellular targeting of the coat protein of African cassava mosaic geminivirus. *Virology*, 286: 373 - 383.

USAID/ZIMBABWE. 1993. Assessment of Program Impact. pp 1 - 5.

van der Merwe, C. P. 1926. The maize *jassid Balclutha mbila* Naude. *Journal Department of Agriculture, Union of South Africa*, 12: 75 - 77.

van Rensberg, G. D. J. 1982. Laboratory observations on the biology of *Cicadulina mbila* (Naude) (Homoptera: Cicadellidae), a vector of maize streak disease. 1. The effect of temperature. *Phytopathologica*, 14: 99 - 107.

van Rensburg, G. D. J. 1983. Southern African species of the genus *Cicadulina* China (Homoptera: Cicadellidae) with descriptions of new species. *Entomological Mem*. Department of Agriculture, Republic of South Africa. No. 57.

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: 2063 - 2074.

WABS. 2008. Maize value chain study in Ghana, enhancing efficiency and competitiveness. *Draft Report*. http://www.yumpu.com/en/document/view/3462564/maize-value-chain-study-in-ghana-valuechains4poor. Accessed on 18 January 2015 at SANBI University of the Western Cape, South Africa.

Walt, E van der., Rybicki, E. P., Varsani, A., Polston, J. E., Bilharz, R., Donaldson, L., Monjane, A. L., Martin, D. P. 2009. Rapid host adaptation by extensive recombination. *Journal of General Virology*, 90 (7): 734 - 746.

Wangia, C., Wangia., S., and de Groote, H. 2004. Review of Maize Marketing in Kenya: Implementation and Impact of Liberalization, 1989-1999. In: Friesen, D. K., and Palmer, A. F. E (eds). Integrated Approaches to Higher Maize Productivity in the New Millennium. Proceedings of the 7th Eastern and Southern African Regional Maize Conference, Nairobi, Kenya, 11-15 February 2002. Mexico, D. F. CIMMYT. pp. 10 - 20.

Warburton, M., Wilkes, G., Taba, S., Charcosset, A., Mir, C., Dumas, F., Madur, D., Dreisigacker, S., Bedoya, C., Prasanna, B. M., Xie, C. X., Hearne, S. *et al*., 2011. Gene flow among different teosinte taxa and into the domesticated maize gene pool. *Genetic Resources and Crop Evolution*, 58: 1243 - 1261.

Webb, M. D. 1987. Species recognition of *Cicadulina* leafhoppers (Hemiptera: Cicadellidae): vectors of pathogens of Graminae. *Bulletin of Entomological Research*, 77: 683 - 712.

Welz, H. G., Scheschert, A., Pernet, A., Pixley, K. V. and Geiger, H. H. 1998. A gene for resistance to the *Maize streak virus* in the African CIMMYT Maize inbred line CML202. *Molecular Breeding*, 4: 147 - 154.

Willment, J. A. 1999. Investigations of the molecular determinants of maize streak virus. PhD Thesis, Department of Microbiology, University of Cape Town, South Africa

Willment, J. A., Martin, D. P., and Rybicki, E. P. 2001. Analysis of the diversity of African *streak mastreviruses* using PCR-generated RFLPs and partial sequence data. *Journal of Virology Methods*, 93: 75 - 87.

Willment, J. A., Martin, D. P., Van der Walt, E., and Rybicki, E. P. 2002. Biological and genomic characterization of Maize *streak virus* isolates from wheat. *Phytopathology*, 92: 81-86.

Woolston, C. J., Reynolds, H. V., Stacey, N. J., Mullineaux, P. M. 1989. Replication of Wheat dwarf virus DNA in protoplasts and analysis of coat protein mutants in protoplasts and plants. *Nucleic Acids Research*, 17: 6029 - 6041.

Xie, Q., Suarez-Lopez, P., and Gutierrez, C. 1995. Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO*, 1 (14): 4073 - 4082.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., Chen, M. H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. S*ystems Biology*, 60: 150 - 160.

Yang, Z., and Nielsen, R. 2008. Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage. *Molecular Biology and Evolution*, 25 (3): 568 - 579. doi:10.1093/molbev/msm284

Zagre, M. B. 1983. Studies on maize streak virus and its transmission characteristics by the leafhopper vector *Cicadulina triangula*. Master's thesis, Universite Nationale du Benin, Cotonu, Benin.

Zehender, G., Ebrenati, E., Bernini, F., Lo Presti, A., Rezza, G., Delogu, M., Galli, M., and Ciccozzi, M. 2011. Phylogeography and epidemiological history of West Nile virus genotype 1a in Europe and the Mediterranean basin. *Infection, Genetics and Evolution*, 11 : 646 - 653. doi:10.1016/j.meegid.2011.02.003

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. W., and McKenna, R. 2001. Structure of the *Maize streak virus* Germinate Particle. *Virology*, 279: 471 - 477.

Appendix 1

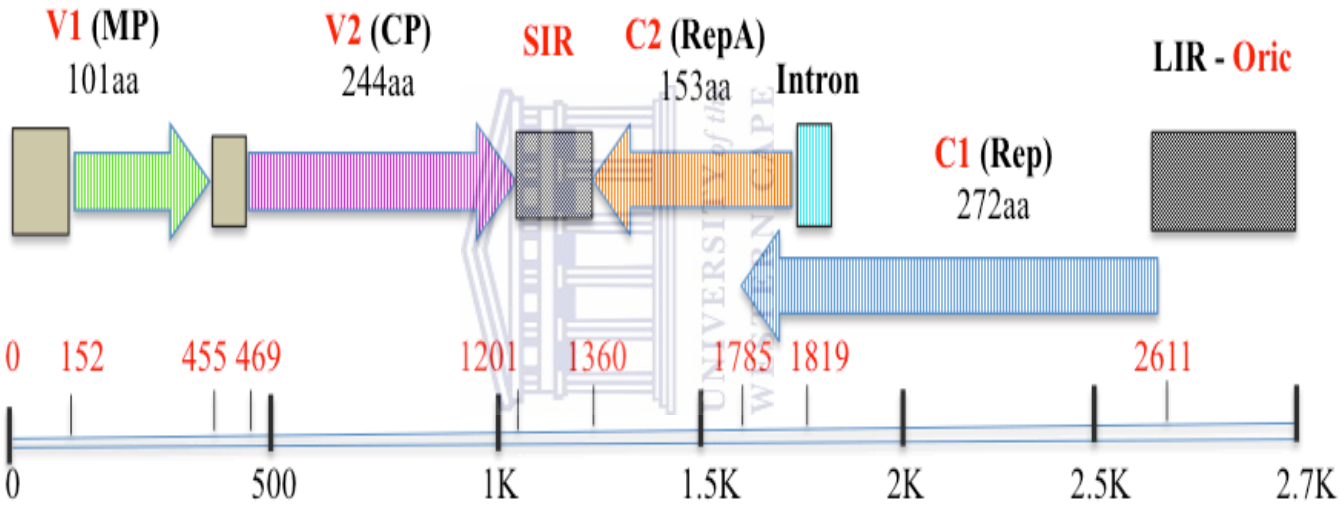


Figure 1.1 The structure of the linearized *Maize streak virus* (MSV) single-stranded DNA (ssDNA) genome. The figure shows the approximate positions of the short and long intergenic regions (SIR and LIR); Origin of replication (Oric), the intron and colored arrows represent genes (V1, V2, C1 and C2) encoding the movement (MP), coat (CP), and replication associated proteins (Rep and RepA), the number of amino acids encoded by each protein and the direction in which transcription occurs. Brown marked regions represent non-coding nucleotide sequences.

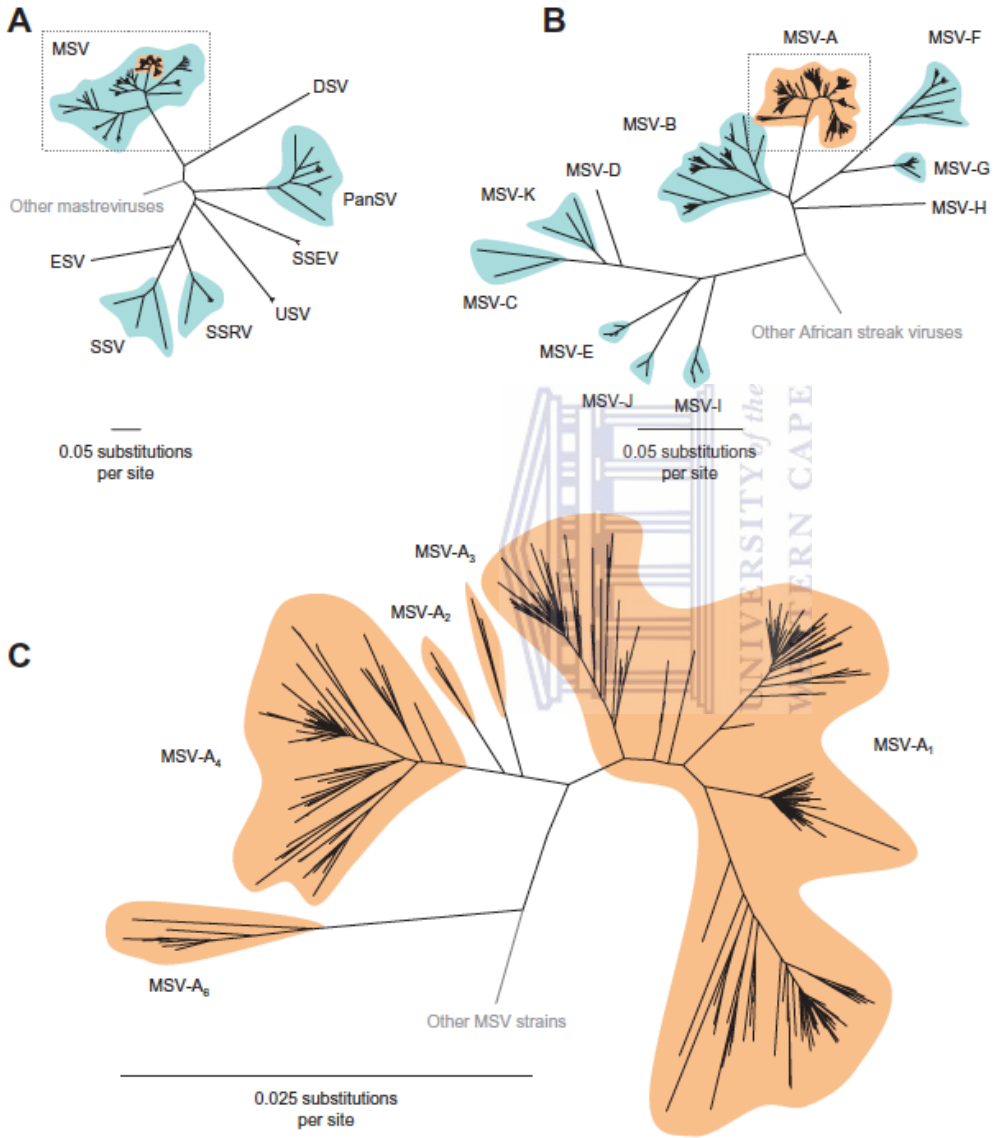
Appendix 2

Table 1.1 Top 20 maize producer countries in the world (FAOSTAT, 2012; http://faostat3.fao.org/browse/rankings/countries_by_commodity/E).

Rank	Area	Production (Int $1000)
1	United States of America	22233636
2	China, mainland	10126214
3	Brazil	2971351
4	Argentina	2635030
5	India	2554046
6	Indonesia	2012638
7	Ukraine	1373511
8	Mexico	1365318
9	France	1336765
10	South Africa	1054543
11	Nigeria	1048014
12	Ethiopia	780289
13	Canada	704334
14	United Republic of Tanzania	667938
15	Philippines	613668
16	Pakistan	542924
17	Kenya	483817
18	Malawi	427153
19	Romania	415331
20	Zambia	390350

* Production is given in metric tonnes per country.

Appendix 3



(**A**) A tree showing the relationship between MSV and the six African streak virus species (ESV, SSV, SSRV, USV, SSEV and PanSV). *Digitaira streak virus* (DSV), from the Island of Vanuatu in the Pacific, is closely related to the African streak viruses and is included for reference purposes. This tree was rooted on the virus *Chloris striate mosaic virus* (not shown). The boxed area is expanded in (B).

(**B**) The eleven known MSV strains (A-K). The MSV-A strain which causes severe MSD is highlighted in orange. The boxed area is expanded in (C).

(**C**) The five MSV-A subtype strains (MSV-A_1, -A_2, -A_3, -A_4, -A_5, -A_6).

Figure 1.2 Phylogenetic relationships of viruses related to *Maize streak virus* (MSV). Adapted from Shepherd *et al.*, 2010.

Appendix 4

Table 1.2 Locations from which full genome MSV sequences have been sampled

Location	Source (s)
Anjouan	Novel (unpublished)
Benin	Conte, 1974; CABI and EPPO, 1997; EPPO, 2014)
Burkina Faso	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014)
Cameroon	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014; Rossel and Thottapilly, 1985)
Central African Republic	(CABI and EPPO, 1997; EPPO, 2014)
Chad	Monjane *et al*., 2011
Ethiopia	(Pinner *et al*., 1988; CABI and EPPO, 1997; EPPO, 2014; Mesfin *et al*., 1991)
Ghana	(Pinner *et al*., 1988; CABI and EPPO, 1997; EPPO, 2014; McKinney, 1929; Oppong *et al*., 2015)
Kenya	Storey, 1936; Pinner *et al*., 1988; CABI and EPPO, 1997; IPPC-Secretariat, 2005; EPPO, 2014)
Lesotho	Monjane *et al*., 2011
Madagascar	(Brunt *et al*., 1990; CABI and EPPO, 1997; EPPO, 2014)
Mauritius	(Guthrie, 1977; Pinner *et* al., 1988; Shepherd, 1925; CABI and EPPO, 1997; EPPO, 2014)
Moheli	Novel (unpublished)
Mozambique	(De Carvalho, 1948; CABI and EPPO, 1997; EPPO, 2014; IITA, 1986)
Nigeria	(Esenam, 1966; Fajemisin *et al*., 1976; Kim *et al*., 1981; Kim *et al*., 1989; Pinner *et al*., 1988;CABI and EPPO, 1997; EPPO, 2014)
Reunion	(CABI and EPPO, 1997; EPPO, 2014; Etienne and Rat, 1973; Malithano *et al*., 1997; Lagat *et al*., 2008)
South Africa	(Fuller, 1901; CABI and EPPO, 1997; EPPO, 2014)
Uganda	(Storey, 1936; CABI and EPPO, 1997; EPPO, 2014; Owor, 2008)
Zambia	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014; IITA, 1986)
Zimbabwe	(IITA, 1986; Rose, 1974; CABI and EPPO, 1997; EPPO, 2014)

* CABI – Commonwealth Agricultural Beaurex International: http://www.cabi.org

* EPPO – European and Mediterranean Plant Protection Organization

Appendix 5

Table 1.3 Countries where no MSV full genome sequences have been collected or are publicly available

Country	Source (s)
Angola	(IITA, 1986; CABI and EPPO, 1997; EPPO, 2014)
Botswana	(CABI and EPPO, 1997; EPPO, 2014)
Burundi	(Pinner *et al*., 1988; CABI and EPPO, 1997; EPPO, 2014)
Congo	(CABI and EPPO, 1997; EPPO, 2014)
Democratic Republic of Congo	(IITA, 1986; CABI and EPPO, 1997; EPPO, 2014)
Cote d'Ivoire	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014)
Egypt	Ammar, 1975; CABI and EPPO, 1997; EPPO, 2014)
Gabon	(CABI and EPPO, 1997; EPPO, 2014)
Guinea	(CABI and EPPO, 1997; EPPO, 2014)
Malawi	(CABI and EPPO, 1997; EPPO, 2014)
Mali	(CABI and EPPO, 1997; EPPO, 2014)
Niger	(CABI and EPPO, 1997; EPPO, 2014)
Rwanda	(Pinner *et* al., 1988; CABI and EPPO, 1997; EPPO, 2014)
Sao Tome and Principe	(CABI and EPPO, 1997; EPPO, 2014)
Senegal	(CABI and EPPO, 1997; EPPO, 2014)
Sierra Leone	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014)
Sudan	(CABI and EPPO, 1997; EPPO, 2014)
Swaziland	(CABI and EPPO, 1997; EPPO, 2014)
Tanzania	(Storey, 1936; CABI and EPPO, 1997; EPPO, 2014)
Togo	(Anon, 1983; CABI and EPPO, 1997; EPPO, 2014)
India	(EPPO, 2014)
Indonesia	(EPPO, 2014)
Yemen	(Brunt *et al*., 1990; CABI and EPPO, 1997; EPPO, 2014)

* CABI – Commonwealth Agricultural Beaurex International: http://www.cabi.org

* EPPO – European and Mediterranean Plant Protection Organization

Appendix 6

Table 2.2 Predictor variables investigated as possible determinants of MSV-A dispersal

Predictor variable	Data source
Absolute Centroid Latitude	Google Earth - https://earth.google.com/
Absolute Centroid Longitude	Google Earth - https://earth.google.com/
Gross Domestic Product (GDP) per capita	http://data.worldbank.org/indicator/NY.GDP.PCAP.CD
Great Circle Geographical Distance	Google Earth - https://earth.google.com/ and R-script (Appendix 8)
Percentage Arable land	http://data.worldbank.org/indicator/AG.LND.ARBL.ZS
Percentage Forest Land	http://data.worldbank.org/indicator/AG.LND.FRST.K2
Pesticide Use	http://data.worldbank.org/indicator/AG.LND.PRCP.MM
Population Density	http://data.worldbank.org/indicator/EN.POP.DNST
Precipitation	http://data.worldbank.org/indicator/AG.LND.PRCP.MM
Prevalence of Undernourishment	http://faostat3.fao.org/dowload/D/FS/E
Temperature	http://www.weatherbase.com/weather/countryall.php3
Yield per hectare	http://faostat3.fao.org/download/Q/QC/E
Total Maize Imports	http://faostat3.fao.org/download/T/TP/E
Total Maize Exports	http://faostat3.fao.org/download/T/TP/E

Appendix 7

Description of the 27 predictor variables investigated as possible determinants of MSV-A dispersal

- **Absolute Centroid Latitude –** represents the centroid latitude geographical coordinate of a given location per country.

- **Absolute Centroid Longitude –** represents the centroid longitude geographical coordinate of a given location per country.

- **GDP per capita (current US$)** – is **gross domestic product** divided by midyear population. **GDP** is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.

- **Great Circle Geographical Distance** – is the shortest distance between two points on the surface of a sphere, measured along the surface of the sphere (as opposed to a straight line through the sphere's interior).

- **Percentage Arable land** (% of land area) - includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded.

- **Percentage Forest area (sq. km)** - is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agro-forestry systems) and trees in urban parks and gardens.

- **Percentage Prevalence of Undernourishment** - refers to the percentage proportion of new cases of undernourishment expressed as a three-year average per country.

- **Pesticide Use (tonnes per 1000 Ha)** – refers to the active ingredient measured in tonnes per 1000 hectares that is used in arable lands and on permanent crops per country.

- **Precipitation in depth (mm per year) -** is the long-term average in depth (over space and time) of annual precipitation in the country. Precipitation is defined as any kind of water that falls from clouds as a liquid or a solid.

- **Population density** – refers to the number of people per square kilometer of land area per country.

- **Temperate (^{o}C)** – is the mean annual temperature per country measured in degrees Celsius.

- **Total Maize Export Quantity (tonnes)** – is the quantity of maize exported per country in tonnes annually.

- **Total Maize Import Quantity (tonnes)** – is the quantity of maize measured in tonnes imported per country annually.

- **Yield (Hg/Ha)** – is the quantity of maize produced measured in tonnes per country per hectare of arable land annually.

Appendix 8

R-script written to estimate great circle distances (as-the-crow-flies).

The script implements the Haversine formula to determine the pair-wise great circle distances between the centroid geographical coordinates of all sampling locations.

Usage : R-package

```
library (fields)

setwd(/output/working/directory/)

latlong_mat <- read.csv(locations_geocordinates.csv,sep=,,na.strings=,row.names=1)
great_dist_mat <- rdist.earth(matrix(c(latlong_mat$Longitude,latlong_mat$Latitude),

ncol=2),matrix(c(latlong_mat$Longitude,latlong_mat$Latitude), ncol=2), miles=FALSE, R=6371)

rownames (great_dist_mat) <- rownames(latlong_mat)

colnames (great_dist_mat) <- rownames(latlong_mat)

diag(great_dist_mat) <- 0

write.table((great_dist_mat), file = great_circ_dist.csv, sep = ';', col.names=NA)
```

Appendix 9

Table 3.1 (a) Inter-strain Recombination Analyses results

Event	Breakpoint		Recombinant
	Start	End	
1	1593	1722	K128_B_MSV_A
2	1722	1764	NG36_pjet_MSV_A NG32_pjet_MSV_A
3	154	303	K209_B_MSV_A
4	1357	1463	GH53_B
5	1971	2033	GH111
6	1535	1555	MSV-A_UG_KasF43-2005-EF015779
7	98*	190	K222_ba
8	331*	398	NG25_B_MSV_A
9	1366	1396	K57_1_2009
10	1631	2576	Rob5
11	152	245	O60_Bethlehem O65_bethlehem [T]
12	1952	30	MSV-A_ZW_Nmg_g168-2006-EU628576
13	2655	1419*	GH3_B
14	1849*	37	O86RC_letsele_1987
15	293	1498*	MSV-A_MZ_Map9_Moz4-2007-HQ693359
16	2661	1499*	MSV-A_ZA_Mak1_M22K-1988-HQ693420
17	543	961	K263_Bh1_as
18	1951*	2655	O56_Kabete_Kenya_1990
19	72	119	GH3_B
20	889	995	g321a_Mau_2008
21	550	960*	MA40_2010

22	145	1494*	MSV-A_ZW_Mas5_Mic7-1993-FJ882146
23	72	1571	MA31_2010
24	1394*	2656	MSV-A_RE_Reu-1997-HQ693399
25	170	1357*	MSV-A_KE_Ken_-1983-X01089
26	418	1824	MSV-A_ZA_ThoE_g132-2006-EU628568
27	106	118	GH3_B
28	1761*	2631*	MSV-A_ZA_MakD-1998-AF329884
29	2659*	2686*	g354_Mau_2008
30	2026	2619	K147_2009
31	1985	1996	MSV-J_Zm-Mic24-1987-EU628641
32	2027	60	K57_1_2009
33	1631*	2643*	GH52_B

Appendix 10

Table 3.1 (b) Intra-strain Recombination Analyses results

Event	Breakpoint		Recombinant
	Start	End	
1	2654	1612	MSV-A_ZA_Omr_g221-2007-EU628573
2	307	1455	MSV-A_ZA_Hec4_O13-1989-FJ882112
3	1632	2342	MSV-A_UG_Jin219-2005-EF547111
4	2663	1552	MSV-A_ZA_VM-1993-AF239961
5	2128	2563	O60_Bethlehem
6	23	1757	MSV-A_CM_Ton_Cam6-2008-HQ693327
7	1807	352	K209_B_MSV_A
8	2660	1951	Mad8_2009
9	1684	2654	K44_1_2009
10	1952	30	MSV-A_ZW_Nmg_g168-2006-EU628576
11	1635	2656	O47b_letsele_1987
12	1997	2619	K57_1_2009 K183
13	1983	126	K95_2009
14	1949	2653	MSV-A_ZA_Let5_M46K-1989-HQ693419
15	1265	141	MSV-A_ZW_Mas5_Mic7-1993-FJ882146
16	2651	1553	GH155_b
17	1959	179	K161B_bam
18	2119	514	MSV-A_ZW_MatC-1998-AF329883
19	2667	1423	MSV-A_ZA_Hei_O9-1979-FJ882115
20	2575	1428	O45RC_Ilanga_1988
21	2655	1609	GH151_b_MSV_A
22	1941	2677	MSV-A_ZA_ThoE_g132-2006-EU628568
23	1635	2656	O77_roude_1984
24	2040	78	g524_B_MSV_A

25	883	2302	K144_B_MSV_A
26	416	1628	O65_bethlehem
27	1112	2011	MSV-A_KE_Ken_-1983-X01089
28	824	1675	MSV-A_ZA_Hec1_O4-1989-FJ882109
29	2637	1405	K110_2009
30	1548	1661	MSV-A_TD_Dja_Mic26-1987-FJ882106
31	2221	194	MSV-A_ZA_Blu4_Ta31-2008-HQ693403
32	430	2092	MSV-A_UG_Mba41-2005-EF547074
33	2648	1635	MSV-A_UG_Luw192-2005-EF547108
34	1836	2620	MSV-A_CF_Bos7_Car7-2008-HQ693313

Appendix 11

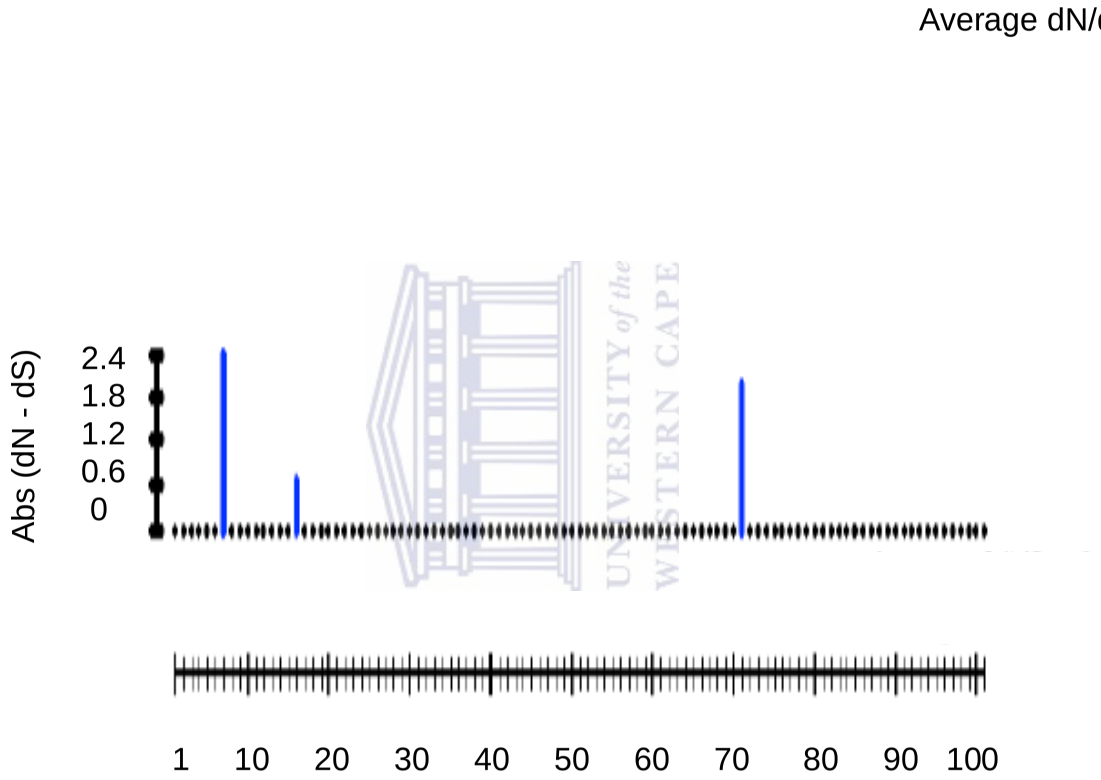


Figure 3.3a Selection Map for dataset D (MP transcript comprising 101 codons). Blue bars highlight 3 codon positions (~3%) evolving under the influence of negative selection. The absence of red bars denotes zero codons positions evolving under the influence of positive selection. The height of each bar (blue or red) approximates to the absolute non-synonymous -synonymous nucleotide substitution rate at a site within each codon position.

Appendix 12

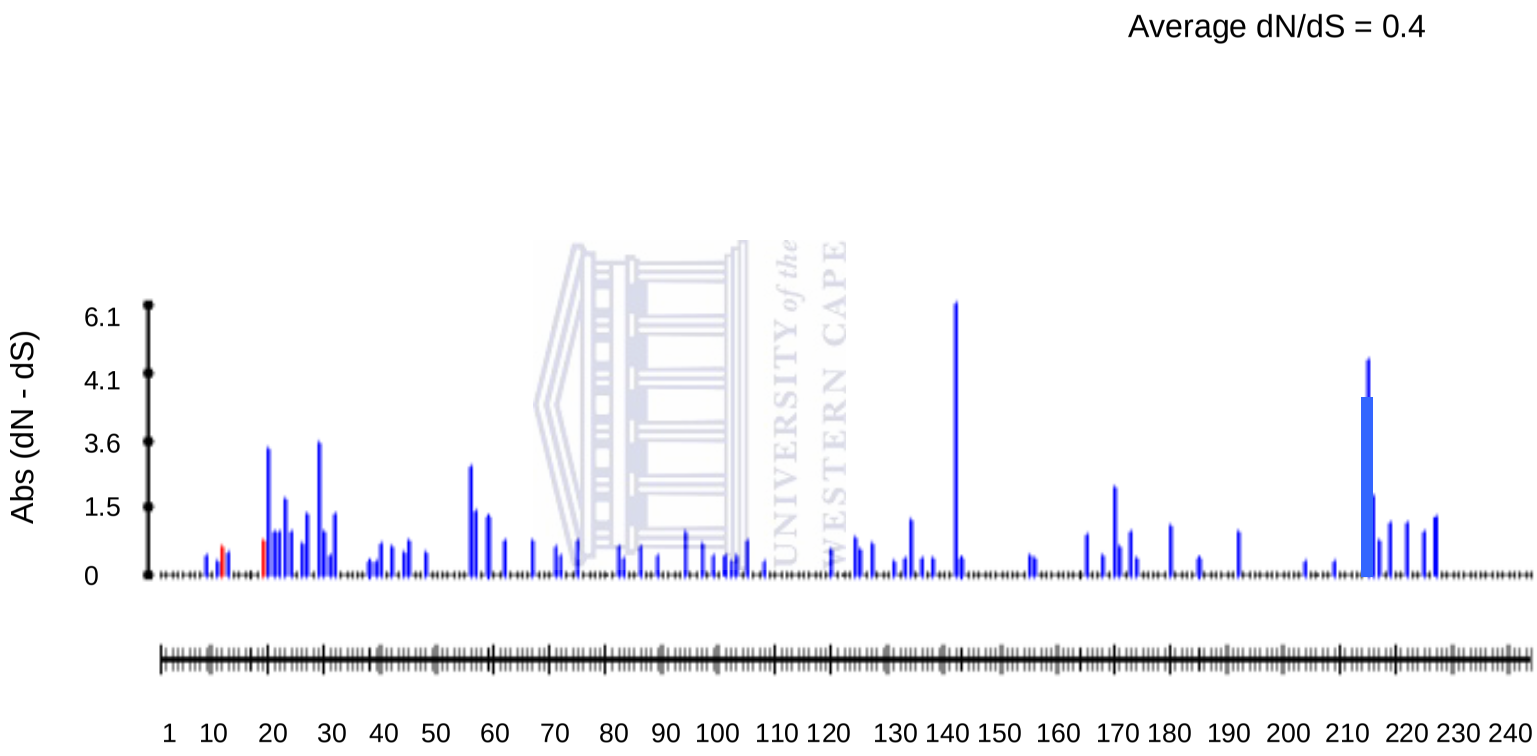


Figure 3.3b Selection Map for dataset E (CP transcript comprising 244 codons). Blue bars highlight 72 codon positions (29.51%) evolving under the influence of negative selection, whereas two red bars denote two codons positions (0.82% - positions 12 and 19) evolving under the influence of positive selection. The height of each bar (blue or red) approximates to the absolute non-synonymous - synonymous nucleotide substitution rate at a site within each codon position.

Appendix 13

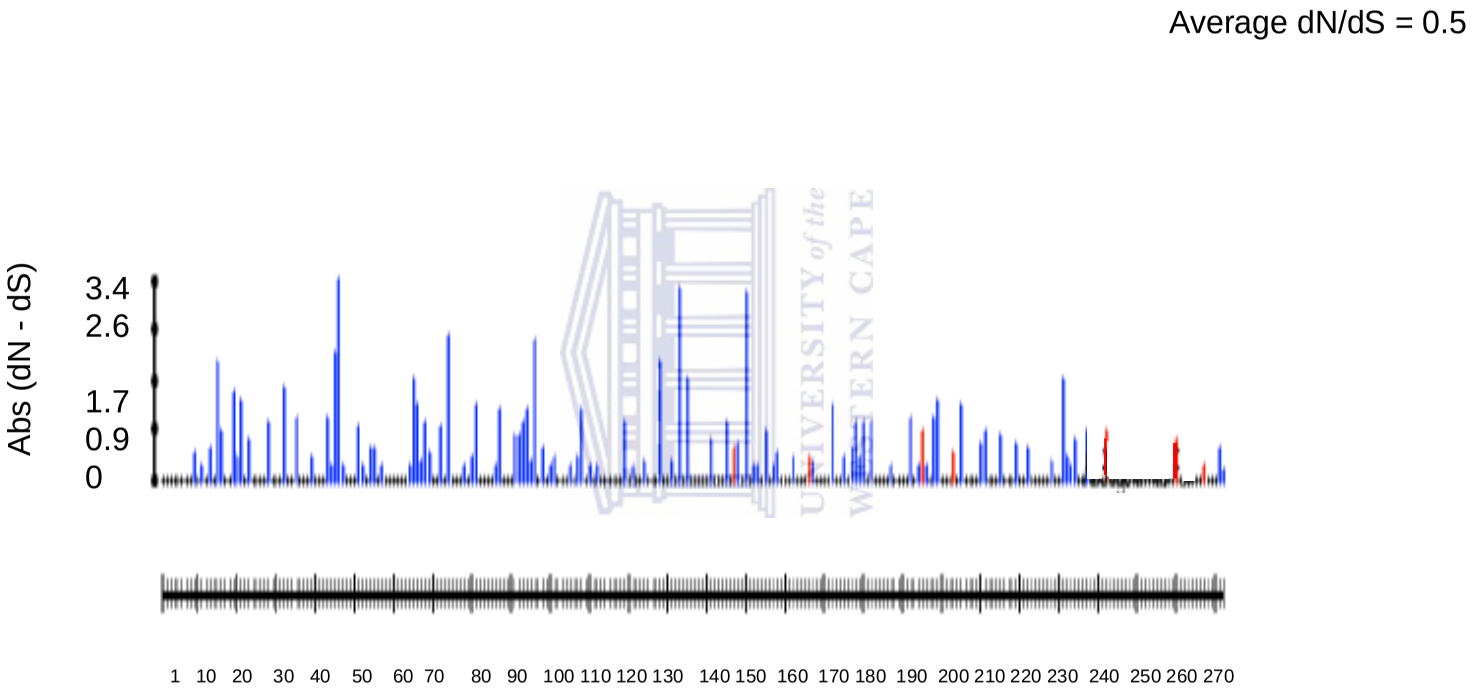


Figure 3.3c Selection Map for dataset F (Rep transcript comprising 272 codons). Blue bars highlight 95 codon positions (34.93%) evolving under the influence of negative selection, whereas the seven red bars denote seven codons positions (2.57% - positions 147, 166, 195, 203, 242, 260, 267) evolving under the influence of positive selection. The height of each bar (blue or red) approximates to the absolute non-synonymous -synonymous nucleotide substitution rate at a site within each codon position.

Appendix 14

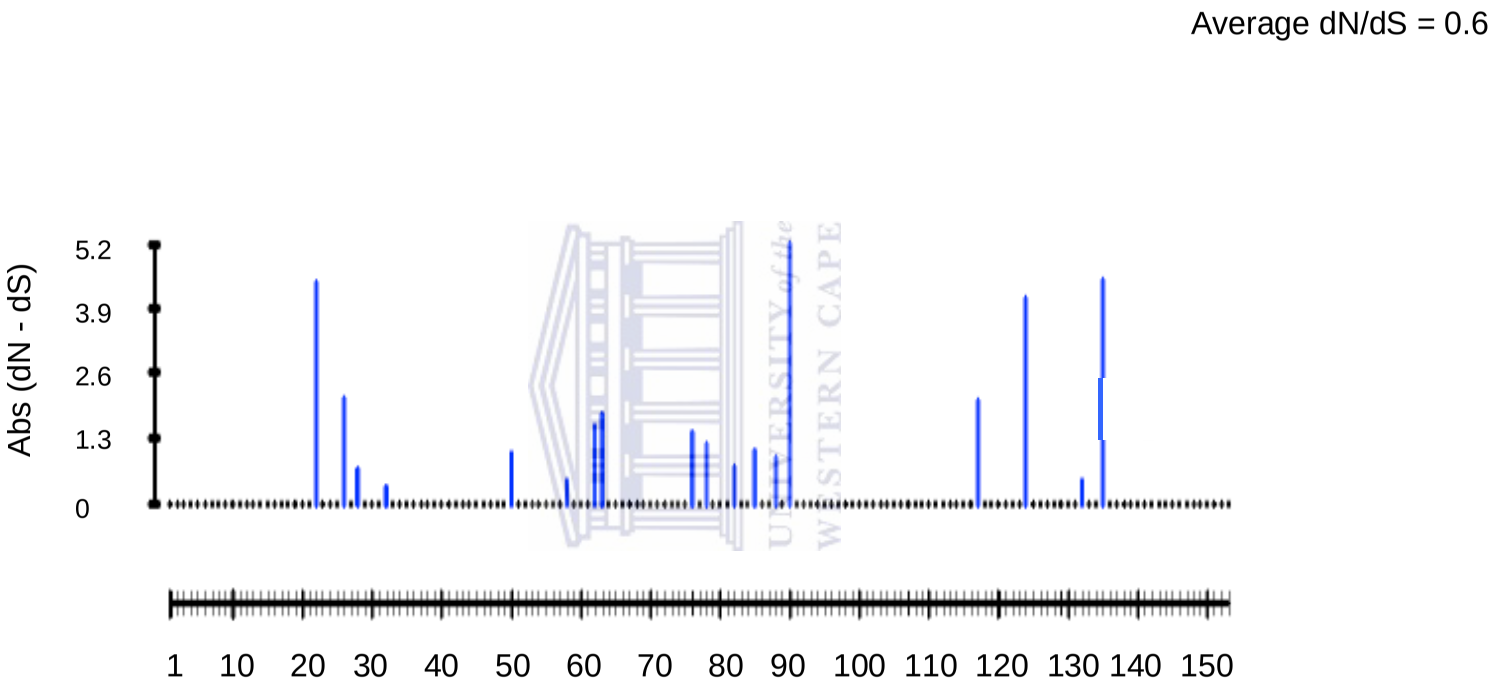


Figure 3.3d Selection Map for dataset G (Rep A transcript comprising 153 codons). The plot shows no significant evidence of codon positions evolving under the influence of positive selection as indicated by the absence of red bars in the plot. Blue bars highlight 18 codon positions (11.8%) evolving under the influence of negative selection. The height of each bar (blue or red) approximates to the absolute non-synonymous -synonymous nucleotide substitution rate at a site within each codon position. A total of 135 codon positions (88.2%) represented by black dots were identified as evolving neutrally.

Appendix 15

Table 3.4 Total number of Bayes factor supported Epidemiological linkages and MSV-A Movements.

Epidemiological linkages	Movement	Bayes Factor Support
1. LS and ZA	from ZA into LS	74722.89892
2. KE and UG	from KE into UG	74722.89892
3. GH and NG	from NG into GH	74722.89892
4. CF and UG	from UG into CF	74722.89892
5. CF and NG	from NG into CF	74722.89892
6. CM and NG	from NG into CM	24902.20115
7. ZA and ZW	from ZW into ZA	14938.06159
8. UG and ZW	from ZW into UG	4973.92204
9. UG and ZM	from UG into ZM	4387.796184
10. GH and KE	from GH into KE	1350.598565
11. BF and NG	from NG into BF	696.8621368
12. MZ and UG	from MZ into UG	226.1188666
13. MD and UG	from UG into MD	186.9724889
14. MZ and ZA	from ZA into MZ	28.95804553
15. MD and MH	from MD into MH	23.27827053
16. BJ and NG	from NG into BJ	14.55996449
17. NG and ZW	from ZW into NG	7.456991575
18. TD and NG	from NG into TD	6.485324493
19. BF and ZW	from ZW into BF	5.189947256

Country code: BF = Burkina Faso; BJ = Benin; CF = Central African Republic; CM = Cameroon; GH = Ghana; KE = Kenya; LS = Lesotho; MD = Madagascar; MH = Moheli; MZ = Mozambique; NG = Nigeria; TD = Chad; UG = Uganda; ZA = South Africa; ZM = Zambia; and ZW = Zimbabwe.

Appendix 16

Table 3.5 Bayes factor support for the 27 predictor variables investigated in the study

Predictor Variable	Bayes factor support
Origin_Precipitation	0.962621607782898
Destination_Precipitation	0.379160315361242
Origin_Temperature	0.04004004004004
Destination_Temperature	0.04004004004004
Origin_location_lat	3.02925989672978
Destination_location_lat	0.18848398991269
Origin_location_long	0.04004004004004
Destination_location_long	0.379160315361242
Origin_GDPpercapita	3960
Destination_GDPpercapita	0.379160315361242
Origin_trade_exports	0.18848398991269
Destination_trade_exports	0.379160315361242
Origin_trade_imports	0.18848398991269
Destination_trade_imports	0.962621607782898
Origin_Pesticide	0.04004004004004
Destination_Pesticide	0.379160315361242
Origin_PercForestLand	3.91743522178305
Destination_PercForestLand	0.04004004004004
Origin_PercArableLand	0.18848398991269
Destination_PercArableLand	0.766408479412964
Origin_PopDensity	0.01
Destination_PopDensity	0.379160315361242
Origin_Yield	0.766408479412964
Destination_Yield	0.766408479412964
Origin_Undernourishment	3960
Destination_Undernourishment	1.96831392298814
GreatCircleDist	3960

Appendix 17

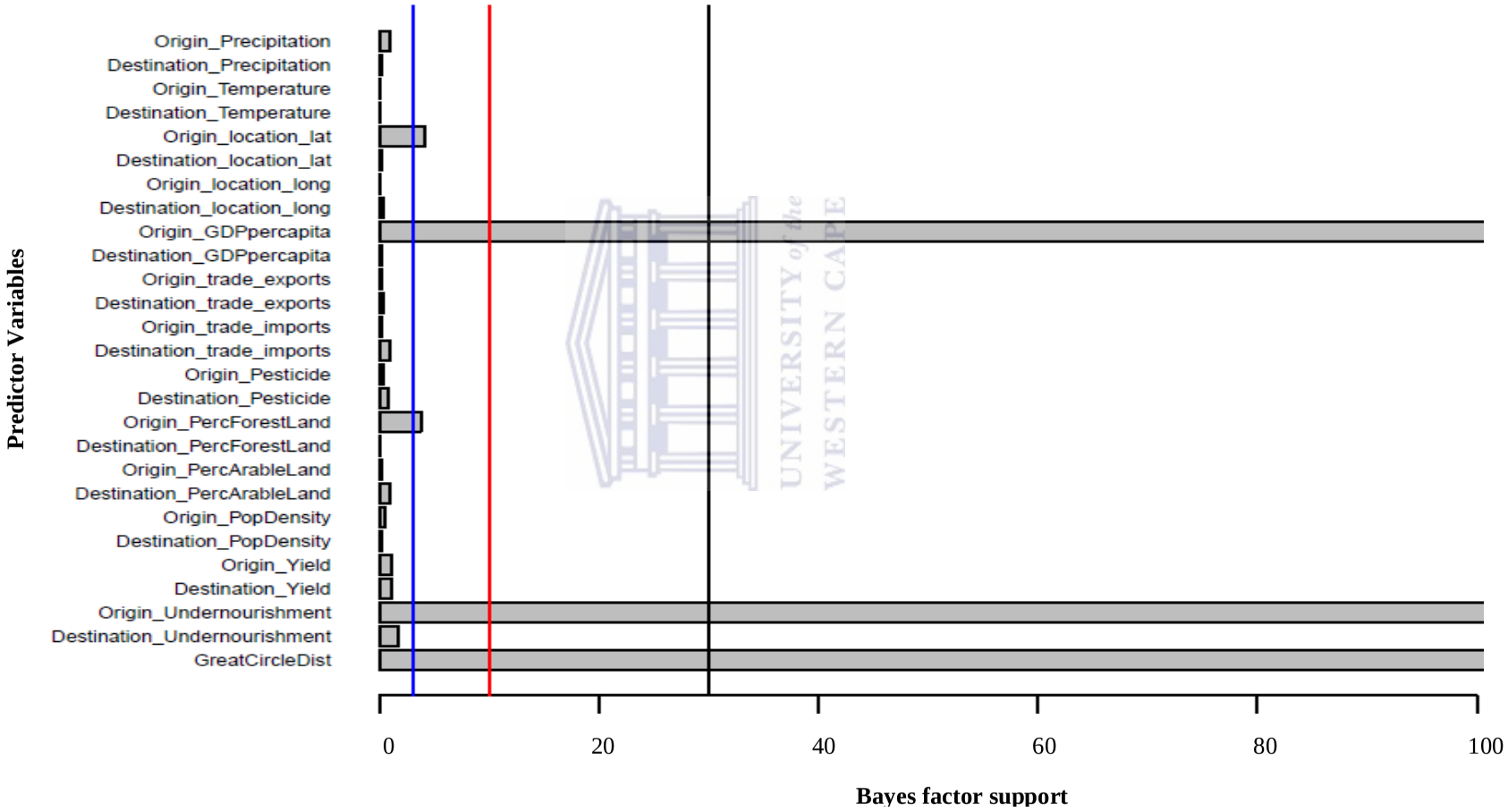


Figure 4.1 Bayes Factor support for the 27 potential predictor variables investigated as possible determinants of MSV-A dispersal across Africa and the adjacent Indian Ocean Islands.

Appendix 18

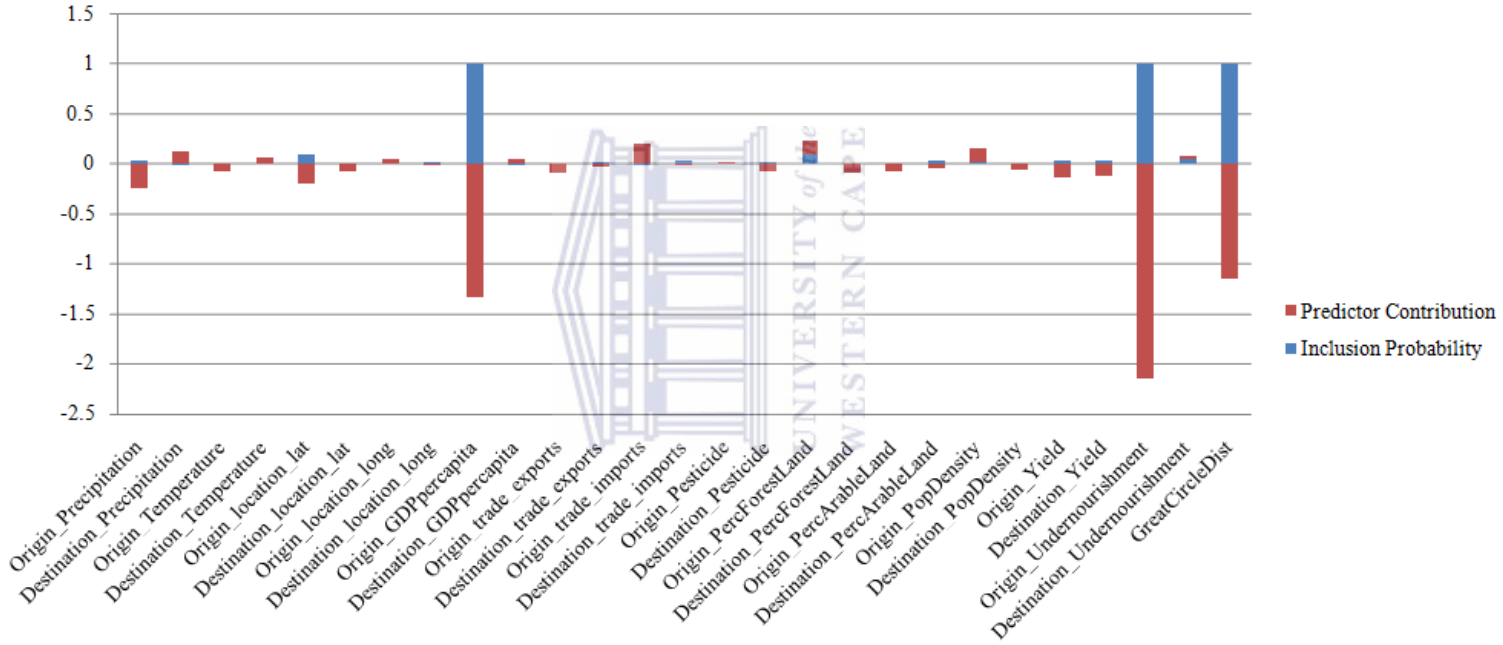


Figure 4.2 Predictor contributions and inclusion probabilities inferred using the generalized linear model for the 27 potential predictive variables.

Appendix 19

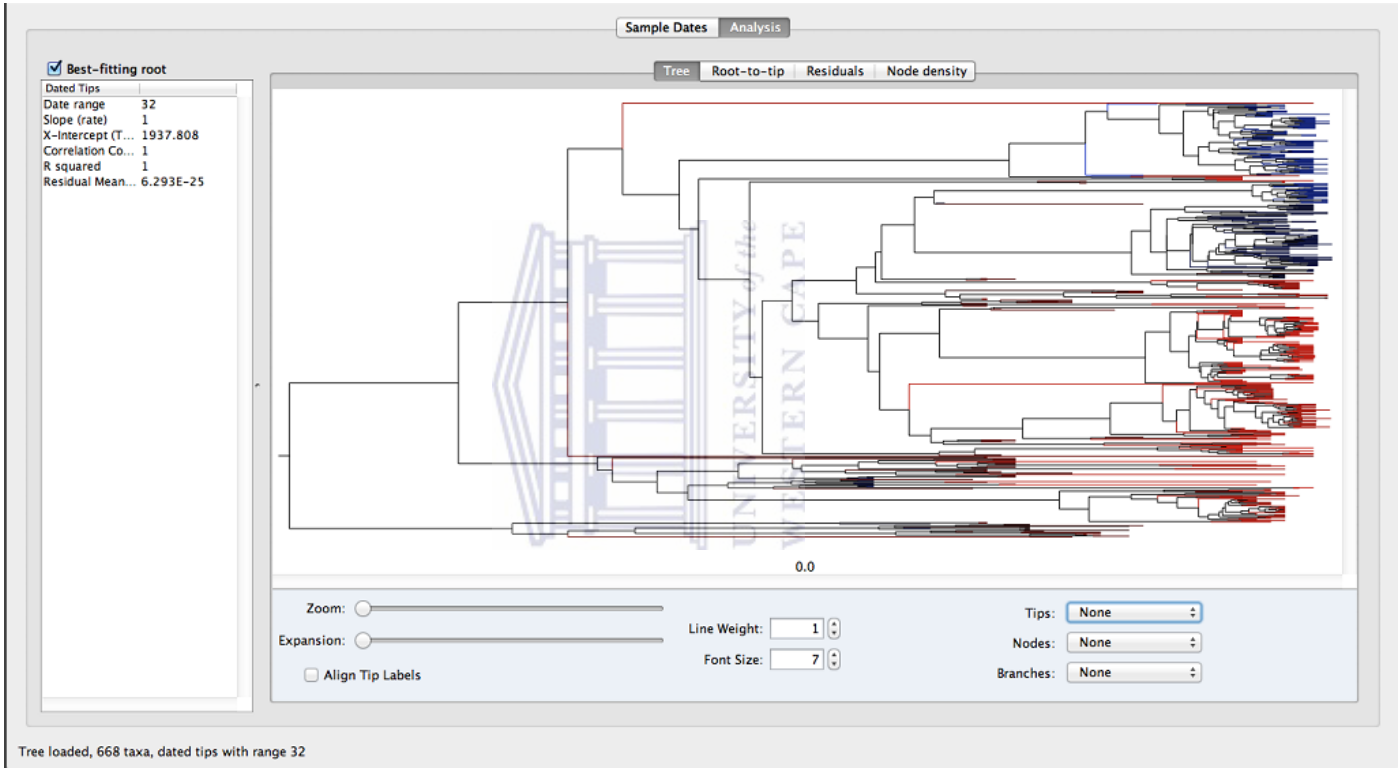


Figure 4.3 Evidence of a high correlation between the sampling times and the genetic distance of the 668 MSV-A samples which was used as a proxy for a strong temporal signal in dataset H.

Appendix 20

Table 3.7 Accession numbers of publicly available MSV-A sequences used

AF329878	KJ699341	HQ693309	FJ882097	EU628566	EF547118
KJ699344	KJ699343	HQ693310	HQ693335	FJ882102	HQ693399
KJ699342	KJ699346	HQ693311	HQ693336	HQ693367	X94330
KJ699303	KJ699347	HQ693312	HQ693337	HQ693368	FJ882106
KJ699304	KJ699353	HQ693313	HQ693338	HQ693369	EF547117
KJ699305	KJ699348	HQ693308	HQ693339	HQ693370	EF547119
KJ699306	KJ699349	HQ693309	HQ693340	HQ693371	EF547075
KJ699307	KJ699350	HQ693316	HQ693341	HQ693372	EF547099
KJ699308	KJ699351	HQ693317	EU628564	HQ693373	EF547100
KJ699309	KJ699352	HQ693318	HQ693342	HQ693374	EF547101
KJ699310	HQ693281	HQ693319	HQ693343	HQ693375	EF547102
KJ699311	HQ693282	HQ693320	HQ693344	HQ693376	EF547103
KJ699312	HQ693283	HQ693321	FJ882098	HQ693377	EF547104
KJ699313	HQ693284	HQ693322	EU628565	HQ693378	EF547105
KJ699314	HQ693281	HQ693323	HQ693345	HQ693379	EF547112
KJ699315	FJ882089	HQ693324	HQ693346	HQ693380	EF547113
KJ699316	HQ693286	HQ693325	HQ693347	HQ693381	EF547114
KJ699317	HQ693287	HQ693326	HQ693348	EU628567	EF547116
KJ699318	HQ693288	HQ693327	HQ693349	HQ693382	EF547115
KJ699319	HQ693289	HQ693328	HQ693350	HQ693383	EF547111
KJ699320	HQ693290	AF329878	HQ693351	HQ693384	EF547081
KJ699321	HQ693291	AF329879	HQ693352	HQ693385	EF015782
KJ699322	HQ693292	FJ882090	HQ693353	HQ693386	EF547121
KJ699323	HQ693293	FJ882091	HQ693354	X01633	EF547122
KJ699325	HQ693294	FJ882092	HQ693355	HQ693387	EF547076
KJ699327	HQ693295	X01089	HQ693356	HQ693388	EF547077
KJ699328	HQ693296	AF395891	HQ693357	HQ693389	EF547080
KJ699329	HQ693297	AF329885	HQ693358	HQ693390	EF547079
KJ699330	HQ693298	FJ882093	HQ693359	HQ693391	EF547080
KJ699331	HQ693299	HQ693329	HQ693360	HQ693392	EF015780
KJ699332	HQ693300	HQ693330	HQ693361	HQ693393	EF015779
KJ699333	HQ693301	HQ693331	HQ693362	FJ882103	EF547096
KJ699334	HQ693302	HQ693332	HQ693363	FJ882104	EF547106
KJ699335	HQ693303	HQ693333	HQ693364	FJ882105	EF547107
KJ699336	HQ693304	HQ693334	HQ693365	HQ693394	EF547084
KJ699337	HQ693305	FJ882094	FJ882099	HQ693395	EF547085
KJ699338	HQ693306	AF329880	FJ882100	HQ693396	EF547087
KJ699339	HQ693307	FJ882095	FJ882101	HQ693397	EF547108
KJ699340	HQ693308	FJ882096	HQ693366	HQ693398	EF547109

Appendix 20: Continued.

EF547094	EU628570	HQ693423	HQ693438	FJ882143	HQ693459
EF547095	HQ693406	HQ693424	AF239961	FJ882144	HQ693460
EF547097	HQ693407	FJ882124	HQ693439	FJ882145	HQ693461
EF547098	EU628571	HQ693425	HQ693440	FJ882146	HQ693462
EF547068	HQ693408	HQ693426	HQ693441	FJ882147	HQ693463
EF547069	FJ882107	EU628572	HQ693442	AF329881	HQ693464
EF547070	HQ693409	EU152254	HQ693443	AF329882	HQ693465
EF547071	HQ693410	EU152255	HQ693444	HQ693471	HQ693466
EF547072	HQ693411	FJ882126	HQ693445	HQ693472	HQ693467
EF547073	FJ882108	HQ693427	HQ693446	FJ882148	HQ693468
EF547074	HQ693412	HQ693428	HQ693447	FJ882149	HQ693469
EF015781	HQ693413	EU628573	HQ693448	HQ693473	FJ882140
EF547123	HQ693414	HQ693429	HQ693449	EU628576	HQ693470
EF547124	FJ882109	HQ693430	FJ882139	HQ693474	FJ882141
EF547066	FJ882110	HQ693431	HQ693450	AF329883	FJ882142
EF547067	FJ882111	FJ882127	HQ693451	FJ882130	FJ882138
EF547065	FJ882112	HQ693432	HQ693452	FJ882131	EU628574
EF547082	FJ882113	HQ693433	HQ693453	FJ882132	EU628575
EF547083	FJ882114	HQ693434	HQ693454	FJ882133	Y00514
EF015783	FJ882115	HQ693435	HQ693455	FJ882134	EU628568
EF547110	HQ693415	HQ693436	HQ693456	FJ882135	HQ693437
EF547086	FJ882116	FJ882128	HQ693457	FJ882136	EU628569
EF547088	FJ882117	FJ882129	HQ693458	FJ882137	HQ693422
EF547089	HQ693416	FJ882121	EF547064	HQ693405	HQ693421
EF547090	FJ882118	FJ882122	EF015778	HQ693402	AF329884
EF547091	FJ882119	FJ882123	HQ693400	HQ693403	HQ693418
EF547092	HQ693417	HQ693419	HQ693401	HQ693404	FJ882120
EF547093	AF003952	HQ693420	EF547063	EF547120	KJ437657
KJ437661	KJ437662	KJ437663	KJ437664	KJ437658	KJ437659
				KJ437660	

Appendix 21

Table 3.8 Taxon labels for Dataset A

K10_FL	MSV-A_UG_Mub89	MSV-A_UG_Masin138	K20_3
K14_2_FL	MSV-A_UG_Hoi156	K67_B_MSV_A	K49_2
K97	K188_B_MSV_A	K190_B_MSV_A	K23_2
K58_2	MSV-A_UG_Iga243	MSV-A_UG_Nak125	K19_1
K206_GH	MSV-A_UG_Mpi8	MSV-A_UG_Bug245	K26_2
MSV-A_UG_Hoi170	K173_b2_as_bgl	K126_B_MSV_A	K163_GH
K231_B_MSV_A	K151_b2_as_bgl	K147	K178_ba
MSV-A_UG_Bus255	K152	K223_ba	K133
K121	K27_B_MSV_A	MSV-A_CF_Bos6_Car6	K21_2
K104	K28_B_MSV_A	K102_B_MSV_A	K230
MSV-A_UG_Mask26	MSV-A_UG_Masin139	K217	K207
K215_ba	MSV-A_UG_Mba38	K214	K44_1
MSV-A_UG_Mpi14	K150	K170	K193_B_MSV_A
K142_B_MSV_A	MSV-A_UG_Hoi167	K198	K279_Bh1_as
K164_ba	K143_B_MSV_A	K197	K262_B_MSV_A
K171_ba	K86_B_MSV_A	K210	K225_ba
K94_B_MSV_A	K129	K172	O87RC_letsele
MSV-A_UG_Kas71	K105	MSV-A_UG_Iga224	MSV-A_ZA_Hec4_O13
MSV-A_UG_Kas63	K77_B_MSV_A	GH144_b_MSV_A	Mad8
g520_K_ptz_MSV_A	K221_B_MSV_A	GH123_b_MSV_A	Mad28
K64_2009	K100A	K136	Mad20
K56_B_MSV_A	K149_B_MSV_A	K216_B_MSV_A	Mad19
MSV-A_UG_Kas76	K69_B_MSV_A	K226	Mad13
K135_B_MSV_A	K132_B_MSV_A	K122_B_MSV_A	Mad21
K218_ba	K145	K146_B_MSV_A	Mad10
K85_B_MSV_A	MSV-A_UG_Mask25	MSV-A_CF_Bang1_Car39	Mad23
K78_b2_as_bgl	MSV-A_UG_Nak129	K99	Mad11
MSV-A_UG_Kas70	MSV-A_UG_Luw196	MSV-A_CF_Bang10_Car49	Mad5
K92_B_MSV_A	MSV-A_UG_Nak118	K95	Mad2
K89_B_MSV_A	MSV-A_UG_WakF56	K183	MSV-A_ZA_Hec5_O17
K87_B_MSV_A	MSV-A_UG_Hoi158	K162_bam	MSV-A_ZA_Mak2_M49
MSV-A_UG_Kib182	MSV-A_UG_Kib179	K162_ba	MSV-A_ZA_Let5_M46K
MSV-A_UG_Iga244	K65_B_MSV_A	MSV-A_UG_KasF43	MSV-A_ZA_Mak1_M22K
K36_B_MSV_A	MSV-A_NG_Ogb1_N36a	K57_1	MSV-A_ZA_Joz_Riz33
MSV-A_UG_Mbal304	MSV-A_UG_Kas75	K103	MSV-A_ZA_MakD
K101	K100	K13_FL	MSV-A_ZA_Tra_D4
K113	K38_B_MSV_A	K83_B_MSV_A	MSV-A_ZA_TreA_g141
MSV-A_UG_Muk203	K180_bam	K52_1	K80_B_MSV_A
	K82_B_MSV_A	K49_1	MSV-A_UG_Masin144

Appendix 21: Continued..

MSV-A_UG_Nak111			
MSV-A_UG_Hoi165	MSV-A_UG_Luw192	K185_ba	K125_B_MSV_A
MSV-A_MZ_Pem3_Moz39	MSV-A_MZ_Map10_Moz5	MSV-A_ZA_Mal4_T9	MSV-A_ZA_Boo_D1
MSV-A_UG_Kib188	K115	MSV-A_ZA_SA	MSV-A_ZA_BooB_g145
K47_B_MSV_A	K295_bh1_as	MSV-A_ZA_Emp5_Ta24	MSV-A_ZA_Cat1_D6
K30_B_MSV_A	K242	MSV-A_ZA_Nat1_g195	MSV-A_ZA_Cat2_D3
MSV-A_ZM_Kas_Z23	K165_B	MSV-A_ZA_Kom	MSV-A_ZA_Har_Riz35
MSV-A_MZ_Bil4_Moz23	MSV-A_MZ_ChiA_g200	MSV-A_ZW_Mas1_Bet43	MSV-A_ZA_Jac_D7
MSV-A_MZ_Lib2_Moz31	MSV-A_MZ_Bil6_Bet25	O86RC_letsele	MSV-A_ZA_Kwa_Riz25
MSV-A_MZ_Map6_Moz27	MSV-A_MZ_Map3_Moz24	O78_MPE	MSV-A_ZA_War1_T5
K189_B_MSV_A	MSV-A_MZ_Map5_Moz26	O66_greytown	MSV-A_ZA_Emp1_T5
MSV-A_MZ_Xai2_Moz7	MSV-A_ZA_Pie1_Ben3a	MSV-A_ZA_Hec6_O18	MSV-A_ZA_Emp3_Ta22
MSV-A_MZ_Inh2_Moz1	MSV-A_ZA_Pie2_Ben3b	MSV-A_ZA_Let1_O8	MSV-A_ZA_Emp4_Ta23
MSV-A_MZ_Bil2_Moz21	MSV-A_ZA_Umz_D10	O59_greytown	MSV-A_ZA_War11_Ta11
MSV-A_MZ_Bil7_Bet16	MSV-A_MZ_Nha_Moz15	MSV-A_ZW_Har1_g186	MSV-A_ZA_War2_Ta1
MSV-A_UG_Nak119	MSV-A_MZ_Bil1_Moz20	MSV-A_ZA_Let2_O10	MSV-A_ZA_War8_Ta8
MSV-A_UG_Nak120	MSV-A_MZ_Chi1_chimoz	MSV-A_ZA_Har_M11	MSV-A_ZA_War3_Ta2
MSV-A_ZW_Maz1_Bet49	MSV-A_MZ_Map1_Moz16	MSV-A_ZA_Let3_O14	MSV-A_ZA_War4_Ta3
K60K	MSV-A_MZ_Map8_Moz3	MSV-A_ZA_Pas_M24	MSV-A_ZA_War9_Ta9
K60	MSV-A_MZ_Map9_Moz4	MSV-A_ZW_Maz3_g265	MSV-A_ZA_War5_Ta4
K109_B_MSV_A	MSV-A_ZA_Por1_D14	MSV-A_ZW_Hel2_Bet36	MSV-A_ZA_War6_Ta6
MSV-A_ZM_Kab2_Z25	MSV-A_ZW_Hel1_Bet36R	MSV-A_ZA_Koe1_O15	MSV-A_ZA_War7_Ta7
MSV-A_MZ_Bob_g204	MSV-A_ZA_ThoE_g132	MSV-A_ZA_Koe2_O21	MSV-A_ZA_Ros_D2-2006
MSV-A_MZ_Chi4_g210	O85B_unk	MSV-A_ZA_Koe3_M42K	MSV-A_ZA_RosE_g131
MSV-A_MZ_Map7_Moz29	O83_letsele	MSV-A_LS_Mal1_Les1	MSV-A_ZA_Not_D5
MSV-A_MZ_Map2_Moz17	O71_PE	MSV-A_ZA_Blu4_Ta31	MSV-A_ZA_Por5_Ta19
MSV-A_UG_Luw110	MA31	MSV-A_ZA_Por2_Ta13	MSV-A_ZA_Mal5_T11
MSV-A_MZ_Lib1_Moz30	MSV-A_MZ_Xai1_xaimoz	MSV-A_ZA_Por4_Ta16	MSV-A_ZA_Mal6_T12
MSV-A_MZ_Nam_Moz34	K128_B_MSV_A	MSV-A_ZA_Mal1_T3	MSV-A_ZA_War10_Ta10
K123_B_MSV_A	MSV-A_ZA_Emp2_T8	MSV-A_ZA_Mal3_T6	MSV-A_ZA_MitC_g129
K181_ba	O45RC_Ilanga	MSV-A_ZA_Mal2_T4	MSV-A_ZA_Oho_Sa26
MSV-A_MZ_Chi3_Moz13	K232	O90B_Komatiport_moz	MSV-A_ZA_New_D9
MSV-A_MZ_Mac_Moz19	K222_ba	O49_letsetele	MSV-A_LS_Mal2_Les2
MSV-A_MZ_Chi2_Moz11	MSV-A_KE_Nye2_Ken12	O80_CT_oudvelt	MSV-A_ZA_RosB_g142
MSV-A_MZ_Bil5_Moz6	MSV-A_ZW_MatC	MSV-A_ZA_Let4_O16	MSV-A_ZA_Blu2_Ta29
K208_GH	o95_5B_MPE	MSV-A_ZA_Pot1_Riz48	MSV-A_ZA_Por6_Ta35
MSV-A_MZ_Bil3_Moz22	MSV-A_ZA_Cpt_M50	g521_B_MSV_A	O93B_MPE
MSV-A_ZW_Chi_Bet64R	MSV-A_ZA_Pot3_O27	Ama_D18	O92B_MPE
MSV-A_MZ_Map4_Moz25	MSV-A_ZA_Pot4_O28	MSV-A_LS_Mal3_Les3	O84_Letsele
MSV-A_ZM_Chi3_Z8	O46RC_Bergenshall	MSV-A_ZA_Blu1_Ta28	O82_letsele
K134_B_MSV_A	MSV-A_ZA_Pot7_O31	MSV-A_ZA_Blu3_Ta30	O50_riversonderend

Appendix 21: Continued..

MSV-A_MZ_Inh1_Moz9	MSV-A_ZA_Pot8_O33	MSV-A_ZA_Blu5_Ta32	MSV-A_ZA_Geo_M51
MSV-A_ZA_Wil_O38	Ma54_2	K254_B_MSV_A	MA01
MSV-A_ZA_VM	Ma54_1	MSV-A_CF_Bai3_Car148	MA11
MSV-A_ZA_Hec2	Ma68	MSV-A_CF_Ban3_Car38	MA32
MSV-A_ZA_Hec3_O12	Ma69	MSV-A_CF_Boss1_Car34	K281_bh1_as
O60_Bethlehem	K160	MSV-A_CF_Boss2_Car35	MSV-A_KE_Kar_K1
MSV-A_ZA_Nat2_g194	Ma45_bh1_as	MSV-A_CF_Bos3_Car3	MSV-A_KE_Nan2_Ke2
MSV-A_ZA_Por3_Ta14	MA35	MSV-A_CF_Bos4_Car4	K287_bh1_as
MSV-A_ZA_Fer_D16	MA29	MSV-A_CF_Bang2_Car40	K285_Bh1_as
O65_bethlehem	MA20	MSV-A_UG_Mpi11	K261_b2_as_bgl
g525_B_MSV_A	K139	MSV-A_CF_Bai2_Car13	MSV-A_KE_Km
g524_B_MSV_A	MSV-A_ZW_Mvu_Bet45	MSV-A_CF_Bang8_Car46	MSV-A_KE_Oyu_K8
MSV-A_ZA_Pot10_O24	MA06	MA35B	MSV-A_KE_Nan1_Ke3
MSV-A_ZA_Pot6_O29	MSV-A_ZM_Chi11_Z20	MSV-A_CF_Bak_Car5	K270_B_MSV_A
MSV-A_ZA_Pot9_O34	MSV-A_ZW_Chi_Zim3	MSV-A_CF_Ban2_Car37	K282_bh1_as
MSV-A_ZA_Hei_O9	K35	MA12	K264_B_MSV_A
MSV-A_ZA_Jou1_O25	MSV-A_UG_Hoi154	MSV-A_CF_Yal1_Car32	K184_ba
MSV-A_ZA_Jou2_O32	MSV-A_ZM_Chi1_Z6	K246_Bh1_as	K268_BH1
MSV-A_ZA_Pot2_O26	MSV-A_MZ_Pem6_Moz42	MSV-A_KE_Nye3_g377	K257_BH1
MSV-A_ZA_Pot5_O28k	MSV-A_ZM_Chi6_Z10	MSV-A_KE_Nye1_Ken11	K265_BH1
MSV-A_ZA_Hec1_O4	MSV-A_ZM_Lus1_Z3	MSV-A_ZM_Chi8_Z12	K278_bh1_as
O89B_Zim	MSV-A_CF_Bim3_Car19	MSV-A_ZM_Chi7_Z11	K277_bh1_as
O79_Zim	MSV-A_ZM_Chi9_Z17	MSV-A_UG_MbaF27	K283_bh1_as
O52_MSV_mat_KEP	MSV-A_UG_Bug248	MSV-A_CF_Ban1_Car20	K175
MSV-A_ZW_MatB	MSV-A_ZM_Lus2_Z4	MSV-A_CF_Bai4_Car15	K127_B_MSV_A
MSV-A_KE_Kag_K14b	MA23	MSV-A_CF_Bang5_Car43	K166_ba
MSV-A_KE_Nak_K7	MA22	MSV-A_CF_Bang6_Car44	K176_ba
MSV-A_KE_Sag	MSV-A_MZ_Pem2_Moz37	MSV-A_UG_Mask23	K219_v2
MSV-A_UG_Mbal308	MA18	MSV-A_UG_KasF42	K219_v1
MSV-A_KE_Gat	MSV-A_ZM_Chi10_Z19	Ma62	K269_BH1
K249_bh1_as	MSV-A_MZ_Pem5_Moz41	Ma53_bh1_as	MSV-A_KE_Nan3_Ke8
K248_bh1_as	MSV-A_UG_MubF49	MSV-A_UG_Luw107	K275_Bh1_as
K174	MSV-A_UG_Mask18	MSV-A_ZM_Chi2_Z7	K177_B_MSV_A
MSV-A_CF_Bang4_Car42	MSV-A_ZW_Maz2_Bet33	MSV-A_UG_Mask21	K251_Bh1_as
MSV-A_UG_Mba41	MSV-A_UG_Wak1	Ma51_bh1_as	Ma74
Ma89	MSV-A_ZM_Chi4_Z9b	MA37	Ma67_bh1_as
Ma81	MSV-A_ZM_Chi5_Z9a	MA09	Ma60
Ma73_bh1_as	K274_Bh1_as	MA04	Ma59
Ma72_bh1_as	MA41	MA24	K245_BH1
Ma50_bh1_as	K81_B_MSV_A	MA21	K267_B_MSV_A
Ma54_3	K191_B_MSV_A	MA26	MSV-A_UG_Iga231

Appendix 21: Continued..

K280_B_MSV_A	MSV-A_UG_Kap289	GH143_b_MSV_A	MSV-A_CM_Baf6_Cam23
K200_GH	MSV-A_UG_Mub94	MSV-A_CF_Bim1_Car16	MSV-A_NG_Ile_N34
MSV-A_CF_Bos7_Car7	K284_bh1_as	MSV-A_CM_Baf2_Cam17	GH4_B
MSV-A_NG_Abu1_NG2	K252_Bh1_as	MSV-A_CM_Baf4_Cam19	GH118_b
K84_B_MSV_A	K209_B_MSV_A	GH141_b_MSV_A	GH113_b
MSV-A_UG_Kap292	g563_B_MSV_A	MSV-A_CF_Bang9_Car47	MSV-A_NG_Lal_N1
MSV-A_UG_Masin149	MSV-A_ZW_Maz4_Mic3	MSV-A_CF_Bos1_Car1	NG6_b_MSV_A
K140	GH111	MSV-A_NG_Eji2_N35a	MSV-A_CM_Baf3_Cam18
MSV-A_UG_Bush53	GH53_B	MSV-A_CF_Bos2_Car2	GH52_B
MSV-A_UG_Kab82	K263_Bh1_as	MSV-A_NG_Ipe_N24	MSV-A_NG_Iba3_N16
MSV-A_UG_KabF48	Mad35_2009	GH19_B	MSV-A_NG_Odo_N18
K141	MSV-A_CM_Ton_Cam6	MSV-A_CF_Bim2_Car18	GH159_b
MSV-A_BF_Lou2_BF6	O57_Kimathi_est	MSV-A_NG_Abe_g239	MSV-A_NG_Iwo_N28a
MSV-A_BF_Oua1_BF4	O55_ILRAD_kenya	GH112_b_MSV_A	K76_B_MSV_A
MSV-A_ZW_MatA	057_b_MSV_A	MSV-A_NG_Igb_N12	K106_B_MSV_A
MSV-A_ZW_Zi_M41	055_b_MSV_A	MSV-A_CF_Bang7_Car45	K74_B_MSV_A
K110	056_b_MSV_A	G32_Bam	K120_B_MSV_A
K39_B_MSV_A	MSV-A_KE_Kan_K4	MSV-A_NG_Ond_N27a	K114_B_MSV_A
MSV-A_CF_Bam1_Car50	MSV-A_KE_MtKA	MSV-A_CF_Bang3_Car41	GH142_b_MSV_A
MSV-A_ZW_Har2_Mic22	O56_Kabete_Kenya	MSV-A_NG_Eji1_N35b	GH110_b_MSV_A
MSV-A_ZW_Mas4_Mic6	o54b_Lerosho_kenya	MSV-A_RE_Pie6_Mic30	MSV-A_NG_Ife_N22
MSV-A_ZW_Mas3_Mic5	MSV-A_KE_Ken	MSV-A_RE_Reu2	GH103_b_MSV_A
MSV-A_ZW_Mas6_Mic8	Mic26_b_MSV_A	MSV-A_RE_Pie2_Mic16	MSV-A_NG_Eji6_N11
MSV-A_ZW_Mas2_Mic4	MSV-A_TD_Dja_Mic26	MSV-A_RE_Pie4_Mic13	MSV-A_CM_Yau_M12
K144_B_MSV_A	MSV-A_NG_Ns	MSV-A_RE_Pie5_Mic14	MSV-A_NG_Iba1_N4
MA10	MSV-A_ZW_Mas5_Mic7	MSV-A_RE_Pie3_Mic17	MSV-A_BF_Lou1_BF5
MA02	MSV-A_ZA_Omr_g221	MSV-A_RE_Reu	MSV-A_BF_Oua2_BF3
MSV-A_MZ_Pem1_Moz36	MSV-A_ZW_Nmg_g168	MSV-A_RE_Jos2_Mic19	K229
MSV-A_ZM_Kab1_Z24	Rob5	MSV-A_RE_Jos1_Mic18	GH157
MSV-A_ZM_Lit_Z14	EW_Rob6	MSV-A_CF_Bai1_Car8	MSV-A_CM_Ema_Cam5
MSV-A_MZ_Pem4_Moz40	MSV-A_RE_Pie1_Mic1	MSV-A_NG_Job_N20a	GH129_b
mad26_2009	O47b_letsele	GH64_B	MSV-A_CM_Baf1_Cam11
K244_bh1_as	g562_B_MSV_A	GH1_B	MSV-A_NG_Oyo_N26
K273_bh1_as	MSV-A_UG_Iga235	G1_Bam	MSV-A_KE_Ama
MA8_2010	NG8_b_MSV_A	MSV-A_NG_Iba5_N22b	MSV-A_UG_Hoi159
Ma75_bh1_as	GH58_B	GH55_B	MSV-A_UG_Nak123
K247_Bh1_as	GH106_b_MSV_A	NG5_B_MSV_A	GH135
MA40_2010	NG13_b_MSV_A	GH127_b_MSV_A	GH104_b
K161B_bam	GH153_b	MSV-A_CM_Omb_Cam10	GH60_B
O77_roude	GH132_b_MSV_A	GH57_B	MSV-A_CF_Bou1_Car25
ET_Rob7	MSV-A_NG_Eji3_N29a	GH154_b	GH105_b_MSV_A

Appendix 21: Continued..

K194_v2	GH120_b_MSV_A	MSV-A_NG_Oko_N32a	GH146_b
MSV-A_UG_Jin219	MSV-A_NG_Iba2_N22	GH128_b_MSV_A	NG4_H_MSV_A
MSV-A_UG_Luw103	MSV-A_NG_Iba4_N37	G13_Bam	GH24_B
MSV-A_UG_Tor271	MSV-A_NG_Abu2_NG1	GH136_b_MSV_A	MSV-A_NG_Ile_g82
MSV-A_UG_Wak4	MSV-A_NG_Bau_NG3	GH130_b	G29_Bam
NG25_B_MSV_A	MSV-A_NG_Eji5_N33	GH139_b_MSV_A	GH140_b_MSV_A
MSV-A_BJ_BenN_Mic20	MSV-A_NG_Ogb2_N30a	GH61_B	GH155_b
K237	MSV-A_NG_Ogb3_N30b	GH5_B	GH162_b
GH148_b	MSV-A_NG_Eji4_N31a	MSV-A_CM_Baf5_Cam22	GH131_b_MSV_A
GH151_b_MSV_A	GH3_B	K90_B_MSV_A	MSV-A_BF_Gol_BF1
GH114_b_MSV_A	MSV-A_CF_Yal2_Car33	NG36_pjet_MSV_A	GH119_b
			NG32_pjet_MSV_A

Appendix 22

Table 3.9 Fifteen MSV-A movements concordant to those reported in the Monjane *et al*. (2011) analyses

Movement	From	To
1	South Africa	Lesotho
2	South Africa	Mozambique
3	Zimbabwe	Burkina Faso
4	Zimbabwe	Central African Republic
5	Zimbabwe	Reunion Island
6	Zimbabwe	Kenya
7	Zimbabwe	Uganda
8	Zimbabwe	South Africa
9	Uganda	Zambia
10	Uganda	Kenya
11	Uganda	Central African Republic
12	Nigeria	Benin
13	Nigeria	Burkina Faso
14	Nigeria	Central African Republic
15	Nigeria	Cameroon