Genetic Diversity and Geographic Distribution of Maize Streak Virus in Kenya

Daniel Pande
Department of Biological and Biomedical Sciences and Technology, Laikipia University, Kenya

Abstract
Maize streak virus (MSV) is one of the most important pathogens responsible for poor maize yields in Africa. For over the past 50 years, the MSV-A_1 genotype has continually been moving back and forth between southern and eastern Africa and from East to West Africa. Despite Kenya being a maize producing country, very little is known of its MSV genetic diversity and geographical distribution of the circulating variants. In this study, a sampling survey was undertaken in the farmers' fields to collect MSV prevalence and symptom severity, where a total of 178 complete MSV genomes were sequenced from both grass and maize. Both phylogenetic and phylogenetic tools were used to illustrate the genetic diversity and geographical distribution. The results showed that the MSVA lineages had a distinct but overlapping geographical distributions in the country and noticeable relationship between the MSV symptom severity and the percentage infectivity.

Keywords: Genetic distribution, geographical distribution, Kenya, Maize streak virus, Mastrevirus

Introduction

Maize streak virus (MSV), one of the most important pathogens responsible for poor maize yields in Sub Saharan Africa is ubiquitously found in Africa; from Egypt in the north (Ammar, 1983) to Sudan in the East, Senegal in the west and South Africa in the south. It is also found on the Atlantic and Indian Ocean islands adjacent to Africa including Madagascar, Mauritius, Rėunion, and Sao Tomė (Monjane et al., 2011; Storey, 1936).

Maize streak virus-A, the maize infecting MSV is considered to have originated in southern Africa in mid-1800s and based on full genome data analysed with a range of Bayesian phylogeographic approaches. It is believed to have spread to La Reunion Island between 1888 and 1970 (in the process of diverging into the MSV- A_6 isolate), to East African countries between 1915 and 1969 (in the process of diversifying into the MSV -A_3 genotype), and to West African countries between 1930 and 1968 (in the process of diversifying into the MSV-A_2 genotype) (Monjane et al., 2011). Whereas the MSV-A_4 genotype has apparently remained in southern Africa, over the past 50 years, the MSV-A_1 genotype has continually been moving back and forth between southern and East Africa and from East to West Africa (Monjane et al., 2011).

In the latest study on genetic diversities, geographical distribution and natural host ranges of MSV-A together with other MSV strains, MSV–A apparently moves throughout Africa far more rapidly than both grass-adapted MSV strains (Varsani et al., 2008) and the related African streak virus, PanSV (Varsani et al., 2009). This increased movement rate could be attributable to MSV-A having probably a broader host range or higher probability of being spread by humans through

transport of infected leaf material or viruliferous insects than its grass-adapted relatives (Martin & Shepherd, 2009).

Based on more than 94 percent sequence identity strain demarcation threshold, MSV is classified into eleven strains (named MSV-A to-K) (Muhire et al., 2013). Despite large scale sampling efforts, all indicators are that only MSV-A- causes severe disease in Maize. MSV-B apparently primarily infects grass species in the genus *Digitaria*, MSV -F and -G primarily infect species in the genus *Urochloa*, MSV-J has been established to be infecting a grass in the genus *Pennisetum* and MSV -C, -H, and -K have primarily been found infecting grass species in the genus *Setaria* (Rybicki et al., 1998; Varsani et al., 2008). 'Grass-adapted' strains such as MSV – B, -C, -D and -E cannot symptomatically infect any maize genotypes except the most susceptible ones (Martin et al., 2001; Schnippenkoetter et al., 2001; Willment et al, 2001). Although the grass-adapted MSV strains may not be a noticeable threat to crops, they could nevertheless provide invaluable information on the evolution and emergence of MSV-A (Kraberger et al., 2017).

To determine the nucleotide sequences relationships between MSV viruses in different groups, Martin et al. (2001) analysed eight MSV-A isolates, and other grass adapted MSV isolates which were cloned, made agroinfectious and fully sequenced. Besides defining an MSV strain as a group of MSV isolates sharing >93% genome-wide sequence similarity, MSV-A isolates, sharing > 98% sequence identity, were classified into different subtype groupings named MSV- A_1-A_6. Owor et al. (2007a) later further refined the MSV-A classification to account for recombination patterns and characterised MSV -A_1 isolates occurring in Uganda into eight distinct recombinant lineages named MSV- A_1I-MSV- A_1VIII.

Diversity studies in Kenya had been conducted on the East African Cassava Mosaic Virus (EACMV) (Bull et al., 2006). However, not much is known about the MSV diversity in Kenya. Therefore there was need to conduct studies on MSV diversity in the country similar to the diversity study done by Owor et al. (2007a) in Uganda.

This report describes the population structure of MSV isolates sampled from maize genotypes and grass species in Kenyan between the year 2008 and 2011. I used analysis of full-genome sequence data, to identify the major MSV variants circulating in Kenya and then to analyse their population genetic, phylogenetic, and phylogeographic characteristics.

Materials and Methods

Virus Sampling

A range of 122 uncultivated grass specimens and a total of 170 Maize samples displaying symptoms characteristic of MSV infection were collected during the first cropping season (May and June) and second cropping season (November and December) between 2008 and 2011; from the maize growing areas in Kenya, viz: The Kenyan Coast, Central, Rift Valley, Western Kenya and parts of Eastern region. For each sample, the geographical coordinates (determined using a global positioning system; GPS, to increase the precision of sampling location), sampling dates (to determine the inter-seasonal fluctuations in MSV populations and sizes) and host species (to determine the host ranges of various MSV pathotypes (recombinant lineages)) were recorded. There was, however, no criterion used in terms of areas to avoid, so long as the fields had maize or grass displaying MSV symptoms.

Three samples from each field were collected where possible. Every field was divided into two diagonals and the three samples were picked from each field. Two from the first diagonal and the third sample from the second diagonal. Fields with fewer than three diseased plants on the two diagonals, up to any three symptomatic plants were sampled from within the field. Samples were

fresh leaves. The fields were approximately 20 kilometres apart. The first was position determined by the nearest maize field from the county headquarters along the main road and contained maize plants of ages ranging from one month to three and a half months old. Symptomatic grass was of any species and age. The samples were press dried between pages of a ledger book later packaged in small envelopes for long-term preservation before processing.

Field Symptom and Degrees of Virulence Scores

Disease severity was scored using IITA's subjective six-point MSV resistance rating system (where plants rated 0 are immune, where as those rated 5 are highly sensitive) in the field. In addition, a field MSV degree of incidence was determined by the percentage number of maize plants infected against the total number of maize plants along the sampling field diagonal.

Cloning and Sequencing of Full Genomes

Viral genomes were isolated from leaf material, circular viral DNA molecules were amplified from a crude total DNA extract using Phi29 DNA polymerase (TempliPhi™, GE Healthcare, USA) as previously described by Owor et al. (2007b) and Shepherd et al. (2008). Briefly, the amplified concatemers were digested (cut) with *Kpn*I or *Bam*HI to yield ~2.7-Kb, potentially linearised viral genomes which were gel purified (invisorb spin DNA extraction kit; invitek) and were subsequently ligated to the *Kpn*I and *Bam*HI sites of pGEMZf+ (Promega Biotech, USA). The clones were fully sequenced by primer walking at Macrogen Inc (Korea).

Phylogenetic and Recombination Analyses of Virus Isolates

A total of 163 MSV-A_1 sequences representing isolates determined in this study, 391 MSV-A_1 sequences from GenBank and one representative of the MSV-A_4 subtype (Pande, 2014) were included in the analyses. A total of 45 genome sequences from grass adapted MSV isolates from Kenya that were determined in this study and 148 MSV B-K genome sequences from GenBank were also included in a separate analysis (Pande, 2014).

All the sequences were aligned by MUSCLE (Edgar, 2004) with manual editing using MEGA 5 (Kumar et al., 2008) and using the CLUSTAL-W (Thompson et al, 1994) based sequence alignment tool which is implemented in MEGA 5. In total, 553 MSV-A_1 sequences and one MSV-A_4 (used to re-root the tree) were firstly analysed for recombination using the computer program RDP4 (with default settings and sequences auto-masked for optimal recombination detection (Martin et al., 2010)).

Maximum likelihood (ML) phylogenetic trees were constructed either using PHYML v1. (Guindon & Gascuel, 2003) with the GTR+I+G_4 nucleotide substitution model (selected as the most appropriate by RDP3 (Martin et al., 2010) or with FastTree2 (with the GTR-CAT nucleotide substitution model and an approximate likelihood ratio test for branch support (Price et al., 2010).

Recombination was analysed using the RDP (Martin & Rybicki, 2000), GENECNOV (Padidam et al., 1999), BOOTSCAN (Martin et al, 2005), MAXCHI (Smith, 1992), CHIMAERA (Posada & Crandall, 2001) SISCAN (Gibbs et al., 2000) and 3SEQ (Boni et al, 2007) recombination detection methods implemented in RDP3 (Martin et al., 2010). Default settings in all programmes were used throughout. Only potential recombination events detected by two or more of these methods, together with phylogenetic evidence of recombination were considered as strong evidence of recombination. Phylogeographic analyses of full viral genome datasets were carried out using visual nested clade analyses of geographically labelled sequences in different clades within the phylogenetic tree.

Results and Discussions

Phylogeography of the Maize Adapted MSV-A Strain

Out of the 292 symptomatic maize and grass leaf samples collected, 178 (142 from maize and 36 from grass samples) yielded clones of full MSV genomes which, once sequenced, were all found to belong to the MSV-A_1 subtype. An additional 15 previously sampled MSV full genomes from Kenya were obtained from GenBank (accession numbers: HQ693329, HQ693330. HQ693331, HQ693332, HQ693333, HQ693334, FJ882090, FJ882092, FJ882093, FJ882094, EU152256, EU152257, AF329878, AF329879 and AF329880) along with 375 MSV- A_1 genomes obtained from elsewhere in Africa.

Monjane et al. (2011) classified MSV-A_1 subtypes into 15 recombinant lineages (MSV-A_1I – MSV-A_1XV), retaining both the recombinant lineage naming convention and the recombinant lineage names designated by who had identified. He named just six lineages, and called them haplotypes.

MSV-A_1 subtype has a continent-wide distribution, but its recombinant lineages display a discernible degree of geographical clustering (Monjane et al., 2011). The study focused on the differences in the MSV-A_1 recombinant lineage demographics in different regions of Kenya. This is because the most significant influence on the appearance of virus epidemics is diversity. From the surveyed sequences, seven major Kenyan MSV-A_1 recombinant lineages were identified (MSV-A_1II, MSV-A_1III, MSV-A_1IV, MSV-A_1V, MSV-A_1VI, MSV-A_1VII, and MSV-A_1XIII). As was found in a 2005 survey of Ugandan MSVs (Owor et al., 2007a), the most widely distributed MSV genotypes in Kenya in my 2008-2011 survey were MSV-A_1III and MSV-A_1V which were found in at least half of the regions.

The predominant recombinant lineages per region included: MSV-A_1V in Nairobi (3/3 sequenced MSVA_1 viruses), MSV-A_1XIII on the coast (10/12 sequenced MSV-A_1 viruses), MSV-A_1V in the eastern region (4/8 sequenced MSV-A_1 viruses), MSV-A_1III in Nyanza (37/76 sequenced MSV-A_1 viruses), MSV-A_1III in the western region (9/12 sequenced MSV-A_1 viruses) and MSV-A_1III and MSV-A_1V in the Rift Valley (17/43 and 10/43 Sequenced MSVA_1 viruses respectively). These results mirror a phylogeographical finding on the East African Cassava Mosaic Virus (EACMV) strains by Bull et al. (2006) in Kenya where the isolates had a distinct but overlapping geographical distributions.

Based on the subjective field visual symptom assessment scores (see fFig1), the most and least virulent of these recombinant MSV-A_1 lineages were MSV-A_1III and MSV-A_1V. It should be noted, however, that the high degrees of MSV virulence in fields from which MSV-A_1III was isolated may have been influenced by the hot and wet climatic conditions around the Great Lakes region where very high virulence scores were recorded in most surveyed fields irrespective of the recombinant lineage present. Fields from which MSV-A_1III isolates were sampled elsewhere in the country were found to be not as severely affected as those around the lake region. The error bars on figure 1 represent 95 percent confidence intervals of the means.

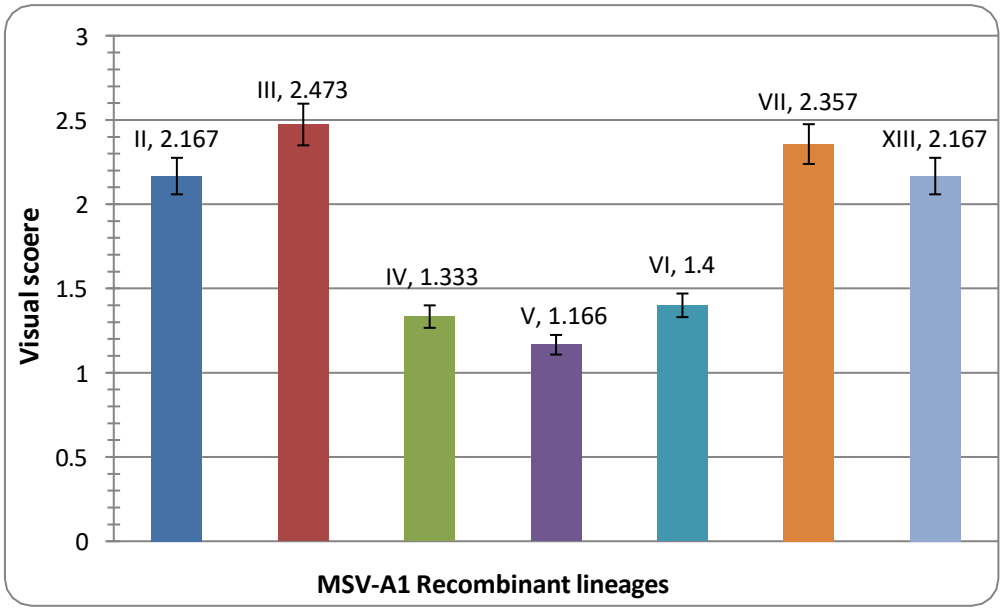

Fig. 1: Subjective MSV-A1 Symptom Severity Ratings (on a scale of 0 to 5)

There was an apparent association between the symptom severity and the incidence (percentage field infected plants, Fig. 2), save for MSV-A_1V and MSV-A_1XIII, which respectively had an elevated and decreased incidence as compared to their relative symptom severities (which were respectively low and high). This may have been due to the fewer numbers of surveyed maize samples showing low incidence: particularly in the coastal region where predominantly MSV-A_1XIII isolates were found. Conversely, in regions where MSVA_1-V was found, infected fields frequently had a 100 percent incidence. The error bars in figure 2 represent 95 percent confidence intervals of the means.

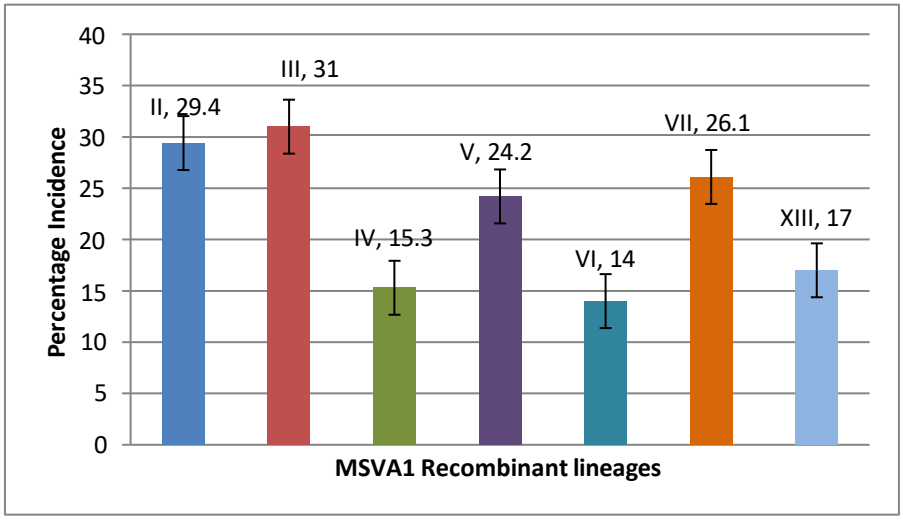

Fig. 2: Average MSV-A1 Percentage Field Incidence

When focusing on the phylogenetic relationships and sampling locations of the 178 MSV-A_1 sequences sampled in Kenya between 2008 and 2011, it is again apparent that MSV-A populations in Kenya might display a slightly greater degree of geographical clustering than those reported in Uganda by Owor et al. (2007a), and was very similar to what was observed in the diversity of cassava mosaic virus in Kenya by Bull et al. (2006). In particular, Nyanza and the Rift valley have very similar MSV population structures and display evidence of extensive mixing between MSV-A_1III and MSV-A_1XIII lineages (Table 1; Figure 3). Similarly, the Eastern and Central regions also have similar MSV population structures and display evidence of frequent movements of MSV-A_1V and MSV-A_1IV lineages between the regions. This implies that there are substantial impediments to the movement of MSV throughout the country, which could be attributed to wind patterns. The distribution of the MSV-A_1XIII lineages in Coastal and Nyanza, Rift Valley regions is intriguing and suggests that there have been some movements of the virus between the three regions but since MSV is only transmitted by leaf hoppers, it remains to be established how this happened because of the vast distance separating the in Coastal and Nyanza, Rift Valley regions.

Table 1: The Regional Distributions of Each of Seven MSV-A1 Genotypes Found in Kenya

	Western	Nyanza	Rift Valley	Central	Eastern	Coast	Nairobi
MSVA-1II	2	8	6	2	0	0	0
MSVA-1III	9	37	17	0	0	2	0
MSVA-1IV	0	1	1	3	2	0	0
MSVA-1V	0	7	10	19	4	0	3
MSVA-1VI	0	2	2	1	0	0	0
MSVA-1VII	0	10	0	0	1	0	0
MSVA-1XIII	0	11	7	0	1	10	0

The figures shown in table 1 represent the counts of the individual MSV-A_1 genotypes.

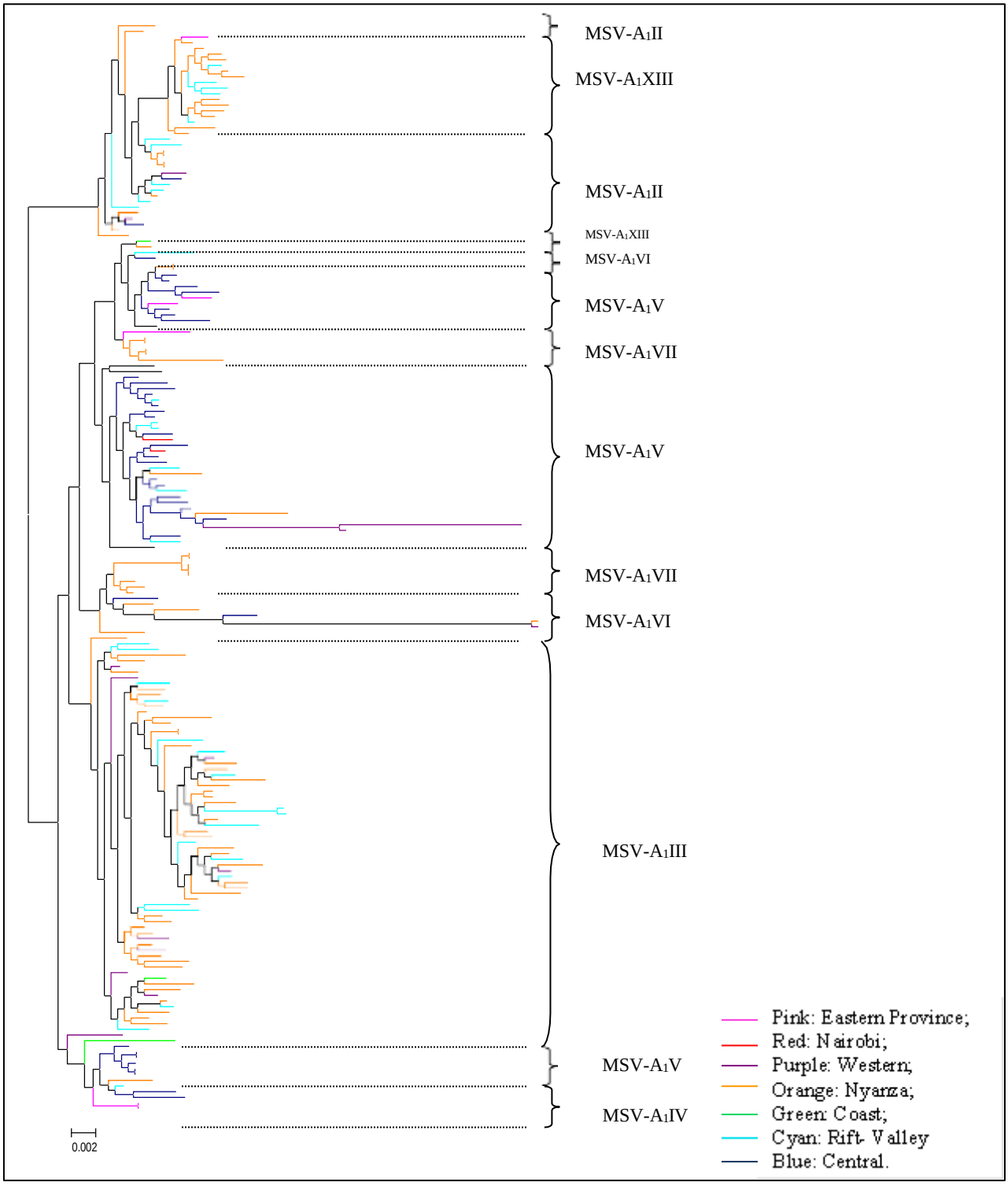

Fig. 3: Maximum Likelihood Phylogenetic Tree of 178 Full Genome Sequences of MSV-A1 Isolates from Kenya

Phylogeography of the Grass Adapted MSV Strains

Out of the 45 grass adapted viruses determined in this study, only the MSV-B, -C, -E, -F, -J and –K strains were identified (see figure 4).

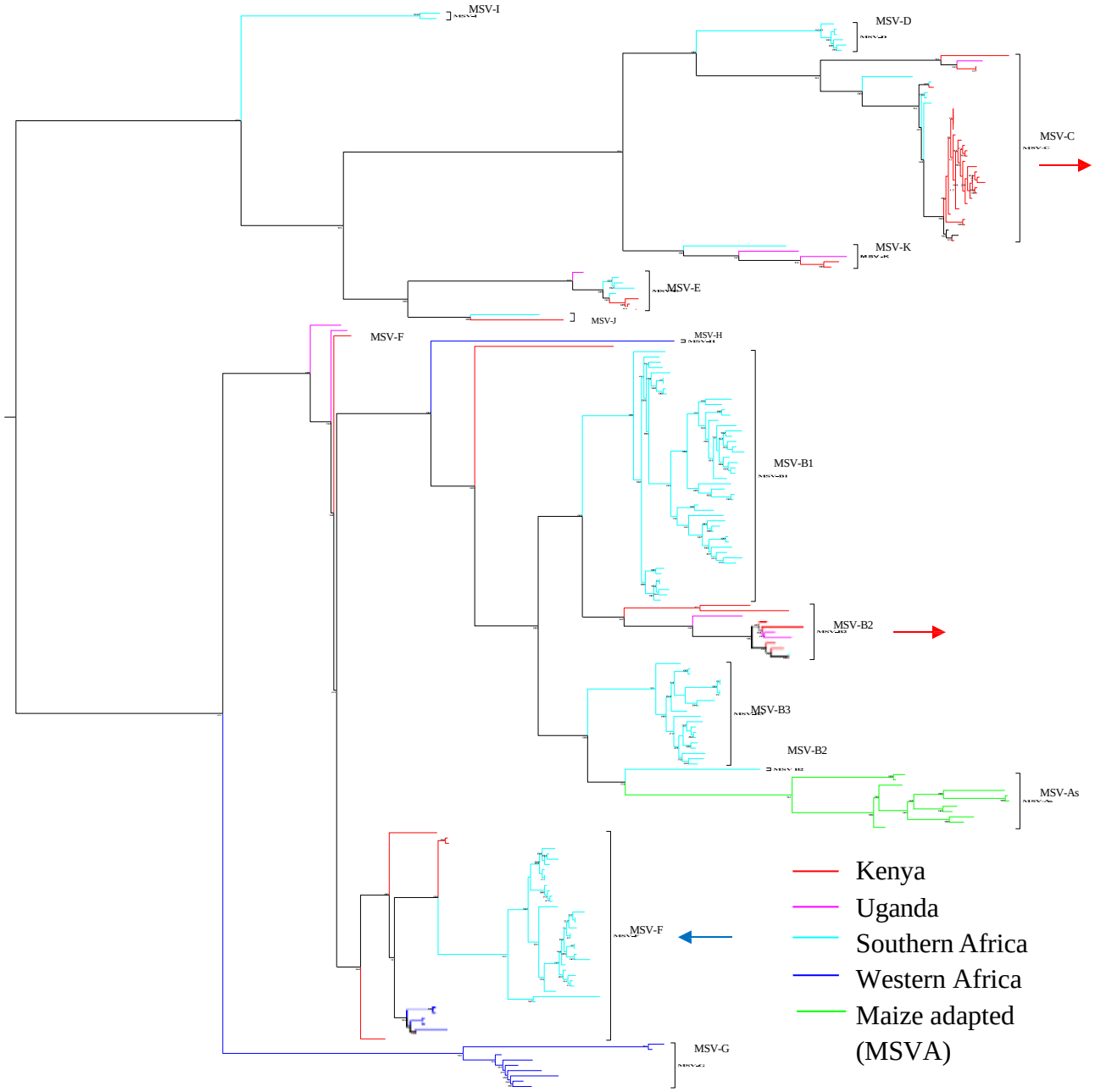

Fig. 4: Maximum Likelihood Phylogeny of 193 Full Genome Sequences of Grass Adapted MSVs and 11 MSV-A Sequences

Figure 4 included six MSV-B isolates, 26 MSV-C isolates, three MSV-E isolates, six MSV-F isolates, one MSV-J isolate, and two MSV-K isolates. Whereas too few MSV-E, -J and K isolates were detected to determine the preferred grass hosts of these MSV-strains, MSV-B isolates were primarily found infecting species in the genus *Digitaria*: a finding similar to that of Varsani et al.

(2008). Also similar to the findings of Varsani et al. (2008), MSV-C isolates were mainly found infecting grasses in the genera *Setaria* and *Digitaria* species.

It is, however, noteworthy that some MSV-A_1 isolates were also detected in and cloned from East African grasses both in this study and in that of Varsani et al. (2008). In this study, they were isolated mainly from *Digitaria sanguinalis, D. didactila, D. ciliaries, Panicum laetum,* and a *saccharum* hybrid. It is likely that grasses such as these are the over-wintering hosts for MSV-A_1 from which the MSV-A strain re-emerges periodically to infect maize plants during the maize growing season.

Although there are presently too few sampled isolates for most of the grass adapted MSV strains to properly assess their movement dynamics, there is some evidence within the tree presented in Figure 4 of trans-continental movements of MSV-B, -C and -F variants. Based on the nesting of the southern African MSV-C and MSV-B2 sequences within clades of Kenyan sequences in the tree, it is evident that variants of these two lineages have likely moved from East Africa (coloured red and pink in the tree) into southern Africa (coloured cyan in the tree). Conversely the nesting of East African MSV-F variants within the southern African clades indicate that variants of this lineage have likely moved from southern Africa to East Africa. It is curious to note that just like movement of MSV-A from West Africa to East Africa consistently occurs in low frequencies, the grass adapted MSV-G strain which is equally common in West Africa is evidently either very rare or missing in East Africa.

Conclusion

While focusing on the phylogeographic relationship of $MSVA_1$ sequences sampled in Kenya, it was apparent that $MSVA_1$ populations showed greater degree of geographical clustering than those reported in Uganda. The study also identified MSV-B, -C, -E, -F, -J and -K strains as grass adapted strains circulating in Kenya. This study therefore recommends a study on topography and the wind patterns in Kenya to understand the clustering of the $MSVA_1$ haplotypes.

References

Ammar, E. D. (1983). *Virus diseases of sugarcane and maize in Egypt.* Paper presented at the International Maize Virus Disease Colloquium and Workshop, Wooster, Ohio (USA), 2-6 Aug 1982.

Boni, M. F., Posada, D., & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics, 176*, 1035-1047.

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., & Stanley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *Journal of general virology, 87*, 3053-3065.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research, 32*, 1792-1797.

Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics, 16*, 573-582.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology, 52*, 696-704.

Kraberger, S., Saumtally, S., Pande, D., Khoodoo, M. H., Dhayan, S., Dookun-Saumtally, A., . . . Lakay, F. M. (2017). Molecular diversity, geographic distribution and host range of monocot-infecting mastreviruses in Africa and surrounding islands. *Virus research, 238*, 171-178.

Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics, 9*, 299-306.

Martin, D., Posada, D., Crandall, K., & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses, 21*, 98-102.

Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics, 16*, 562-563.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics, 26*, 2462-2463.

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

Martin, D. P., Willment, J. A., Billharz, R., Velders, R., Odhiambo, B., Njuguna, J., . . . Rybicki, E. P. (2001). Sequence Diversity and Virulence in *Zea mays* of Maize Streak Virus Isolates. *Virology, 288*, 247-255.

Monjane, A. L., Harkins, G. W., Martin, D. P., Lemey, P., Lefeuvre, P., Shepherd, D. N., . . . Varsani, A. (2011). Reconstructing the history of maize streak virus strain a dispersal to reveal diversification hot spots and its origin in southern Africa. *J Virol, 85*, 9623-9636. doi: 10.1128/JVI.00640-11

Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., . . . Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). *Arch Virol, 158*, 1411-1424.

Owor, B. E., Martin, D. P., Shepherd, D. N., Edema, R., Monjane, A. L., Rybicki, E. P., . . . Varsani, A. (2007a). Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *Journal of general virology, 88*, 3154-3165.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., . . . Varsani, A. (2007b). Successful application of FTA Classic Card technology and use of bacteriophage ϕ29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. *Journal of virological methods, 140*, 100-105.

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

Pande, D. (2014) MSV Sequences. http://web.cbio.uct.ac.za/˜darren/SupplementaryFigure1.fas

Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences, 98*, 13757-13762.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One, 5*, e9490.

Rybicki, E. P., Dennis, S., Napier, G., & Hughes, F. L. (1998). Novel grass and wheat strains of maize streak virus detected by DNA amplification and sequencing. *Molecular Plant Pathology. On-Line http://www. bspp. org. uk/mppol/1998/0929rybicki*.

Schnippenkoetter, W. H., Martin, D. P., Hughes, F. L., Fyvie, M., Willment, J. A., James, D., . . . Rybicki, E. P. (2001). The relative infectivities and genomic characterisation of three distinct mastreviruses from South Africa. *Arch Virol, 146*, 1075-1088.

Volume 1, Number 1, 2022 *ISSN N0 2791-1926*

Shepherd, D. N., Martin, D. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Rybicki, E. P., & Varsani, A. (2008). A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *Journal of virological methods, 149*, 97-102.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *Journal of molecular evolution, 34*, 126-129.

Storey, H. H. (1936). Virus Diseases of East African Plants: IV. A Survey of the Viruses attacking Gramineae. *East African Agricultural Journal, 1*, 333-337.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research, 22*, 4673-4680.

Varsani, A., Monjane, A. L., Donaldson, L., Oluwafemi, S., Zinga, I., Komba, E. K., . . . Martin, D. P. (2009). Comparative analysis of Panicum streak virus and Maize streak virus diversity, recombination patterns and phylogeography. *J Virol, 6*, 194. doi: 10.1186/1743-422X-6-194 Varsani, A., Oluwafemi, S., Windram, O. P., Shepherd, D. N., Monjane, A. L., Owor, B. E., . . . Martin, D. P. (2008). Panicum streak virus diversity is similar to that observed for maize streak virus. *Arch Virol, 153*, 601-604.

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